Cochrane Library

Cochrane Database of Systematic Reviews

Combination of the non-invasive tests for the diagnosis of endometriosis (Review)

Nisenblat V, Prentice L, Bossuyt PMM, Farquhar C, Hull ML, Johnson N

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, Issue 7. Art. No.: CD012281.
DOI: 10.1002/14651858.CD012281.

TABLE OF CONTENTS

[Diagnostic Test Accuracy Review]

Combination of the non-invasive tests for the diagnosis of endometriosis

Vicki Nisenblat[1], Lucy Prentice[2], Patrick MM Bossuyt[3], Cindy Farquhar[4], M Louise Hull[1], Neil Johnson[1]

[1]Discipline of Obstetrics and Gynaecology, School of Medicine, Robinson Research Institute, The University of Adelaide, Adelaide, Australia. [2]Obstetrics and Gynaecology, Tauranga Hospital, Bay of Plenty DHB, Tauranga, New Zealand. [3]Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands. [4]Department of Obstetrics and Gynaecology, University of Auckland, Auckland, New Zealand

Contact: Vicki Nisenblat, Discipline of Obstetrics and Gynaecology, School of Medicine, Robinson Research Institute, The University of Adelaide, Level 6, Medical School North,, Frome Rd, Adelaide, SA, 5005, Australia. vnisenblat@gmail.com.

Editorial group: Cochrane Gynaecology and Fertility Group.
Publication status and date: New, published in Issue 7, 2016.

Citation: 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, Issue 7. Art. No.: CD012281. DOI: 10.1002/14651858.CD012281.

ABSTRACT

Background

About 10% of women of reproductive age suffer from endometriosis, a costly chronic disease causing pelvic pain and subfertility. Laparoscopy is the gold standard diagnostic test for endometriosis, but is expensive and carries surgical risks. Currently, there are no non-invasive tests available in clinical practice to accurately diagnose endometriosis. This review assessed the diagnostic accuracy of combinations of different non-invasive testing modalities for endometriosis and provided a summary of all the reviews in the non-invasive tests for endometriosis series.

Objectives

To estimate the diagnostic accuracy of any combination of non-invasive tests for the diagnosis of pelvic endometriosis (peritoneal and/or ovarian or deep infiltrating) compared to surgical diagnosis as a reference standard. The combined tests were evaluated as replacement tests for diagnostic surgery and triage tests to assist decision-making to undertake diagnostic surgery for endometriosis.

Search methods

We did not restrict the searches to particular study designs, language or publication dates. We searched CENTRAL to July 2015, MEDLINE and EMBASE to May 2015, as well as the following databases to April 2015: CINAHL, PsycINFO, Web of Science, LILACS, OAIster, TRIP, ClinicalTrials.gov, DARE and PubMed.

Selection criteria

We considered published, peer-reviewed, randomised controlled or cross-sectional studies of any size, including prospectively collected samples from any population of women of reproductive age suspected of having one or more of the following target conditions: ovarian, peritoneal or deep infiltrating endometriosis (DIE). We included studies comparing the diagnostic test accuracy of a combination of several testing modalities with the findings of surgical visualisation of endometriotic lesions.

Data collection and analysis

Three review authors independently collected and performed a quality assessment of the data from each study by using the QUADAS-2 tool. For each test, the data were classified as positive or negative for the surgical detection of endometriosis and sensitivity and specificity estimates were calculated. The bivariate model was planned to obtain pooled estimates of sensitivity and specificity whenever sufficient data were available. The predetermined criteria for a clinically useful test to replace diagnostic surgery were a sensitivity of 0.94 and a specificity of 0.79 to detect endometriosis. We set the criteria for triage tests at a sensitivity of 0.95 and above and a specificity of 0.50 and

above, which 'rules out' the diagnosis with high accuracy if there is a negative test result (SnOUT test), or a sensitivity of 0.50 and above and a specificity of 0.95 and above, which 'rules in' the diagnosis with high accuracy if there is a positive result (SpIN test).

Main results

Eleven eligible studies included 1339 participants. All the studies were of poor methodological quality. Seven studies evaluated pelvic endometriosis, one study considered DIE and/or ovarian endometrioma, two studies differentiated endometrioma from other ovarian cysts and one study addressed mapping DIE at specific anatomical sites. Fifteen different diagnostic combinations were assessed, including blood, urinary or endometrial biomarkers, transvaginal ultrasound (TVUS) and clinical history or examination. We did not pool estimates of sensitivity and specificity, as each study analysed independent combinations of the non-invasive tests.

Tests that met the criteria for a replacement test were: a combination of serum IL-6 (cut-off >15.4 pg/ml) and endometrial PGP 9.5 for pelvic endometriosis (sensitivity 1.00 (95% confidence interval (CI) 0.91 to 1.00), specificity 0.93 (95% CI, 0.80, 0.98) and the combination of vaginal examination and transvaginal ultrasound (TVUS) for rectal endometriosis (sensitivity 0.96 (95% CI 0.86 to 0.99), specificity 0.98 (95% CI 0.94 to 1.00)). Tests that met the criteria for SpIN triage tests for pelvic endometriosis were: 1. a multiplication of urine vitamin-D-binding protein (VDBP) and serum CA-125 (cut-off >2755) (sensitivity 0.74 (95% CI 0.60 to 0.84), specificity 0.97 (95% CI 0.86 to 1.00)) and 2. a combination of history (length of menses), serum CA-125 (cut-off >35 U/ml) and endometrial leukocytes (sensitivity 0.61 (95% CI 0.54 to 0.69), specificity 0.95 (95% CI 0.91 to 0.98)). For endometrioma, the following combinations qualified as SpIN test: 1. TVUS and either serum CA-125 (cut-off ≥25 U/ml) or CA 19.9 (cut-off ≥12 U/ml) (sensitivity 0.79 (95% CI 0.64 to 0.91), specificity 0.97 (95% CI 0.91 to 1.00)); 2. TVUS and serum CA 19.9 (cut-off ≥12 U/ml) (sensitivity 0.54 (95% CI 0.37 to 0.70), specificity 0.97 (95% CI 0.91 to 1.0)); 3-4. TVUS and serum CA-125 (cut-off ≥20 U/ml or cut-off ≥25 U/ml) (sensitivity 0.69 (95% CI 0.49 to 0.85), specificity 0.96 (95% CI 0.88 to 0.99)); 5. TVUS and serum CA-125 (cut-off ≥35 U/ml) (sensitivity 0.52 (95% CI 0.33 to 0.71), specificity 0.97 (95% CI 0.90 to 1.00)). A combination of vaginal examination and TVUS reached the threshold for a SpIN test for obliterated pouch of Douglas (sensitivity 0.87 (95% CI 0.69 to 0.96), specificity 0.98 (95% CI 0.95 to 1.00)), vaginal wall endometriosis (sensitivity 0.82 (95% CI 0.60 to 0.95), specificity 0.99 (95% CI 0.97 to 1.0)) and rectovaginal septum endometriosis (sensitivity 0.88 (95% CI 0.47 to 1.00), specificity 0.99 (95% CI 0.96 to 1.00)).

All the tests were evaluated in individual studies and displayed wide CIs. Due to the heterogeneity and high risk of bias of the included studies, the clinical utility of the studied combination diagnostic tests for endometriosis remains unclear.

Authors' conclusions

None of the biomarkers evaluated in this review could be evaluated in a meaningful way and there was insufficient or poor-quality evidence. Laparoscopy remains the gold standard for the diagnosis of endometriosis and using any non-invasive tests should only be undertaken in a research setting.

PLAIN LANGUAGE SUMMARY

Combination of different types of tests for the non-invasive diagnosis of endometriosis

Review Question

Can any combination of non-invasive tests be accurate enough to replace or reduce the need for surgery in the diagnosis of endometriosis?

Background

Women with endometriosis have endometrial tissue (the tissue that lines the womb and is shed during menstruation) growing outside the womb within the pelvic cavity. This tissue responds to reproductive hormones, causing painful periods, chronic lower abdominal pain and difficulty conceiving. Currently, the only reliable way of diagnosing endometriosis is to perform keyhole surgery and visualise the endometrial deposits inside the abdomen. Because surgery is risky and expensive, combinations of various tests have been evaluated for their ability to detect endometriosis non-invasively. An accurate test could lead to the diagnosis of endometriosis without the need for surgery or it could reduce the need for diagnostic surgery so only women who were most likely to have endometriosis would require it.

Study characteristics

The evidence included in this review is current to April 2015. We included 11 studies on combinations of several testing methods involving 1339 participants. All studies evaluated women of reproductive age who were undertaking diagnostic surgery to investigate symptoms of endometriosis or for other indications. Fifteen combinations of different blood, endometrial and urinary biomarkers were studied, incorporating ultrasound, clinical history and examination. Each combination of tests was assessed in small individual studies.

Key results and quality of evidence

Several studies identified the combined tests that might be of value in diagnosing endometriosis, but there are too few reports to be sure of their diagnostic benefit.

The reports were of low methodological quality, which is why these results cannot be considered reliable unless confirmed in large high-quality studies. Overall, there is not enough evidence to demonstrate benefit of any combined non-invasive test for use in clinical practice for the diagnosis of endometriosis over the current 'gold standard' of diagnostic laparoscopy.

Future research

More high-quality research studies are needed to accurately assess the diagnostic potential of any type of non-invasive tests or their combinations that were identified in only a few studies as possibly having value in the detection of endometriosis.

SUMMARY OF FINDINGS

Summary of findings 1. Summary of findings table

Review question	What is the diagnostic accuracy of the combined test of different testing modalities with or without clinical history or examination in detecting pelvic endometriosis [peritoneal endometriosis, endometrioma, DIE]?
Importance	A simple and reliable non-invasive test for endometriosis with the potential to either replace laparoscopy or to triage women in order to reduce surgery, would minimise surgical risk and reduce diagnostic delay
Patients	Women of reproductive age: 1) with suspected endometriosis, or 2) with persistent ovarian mass, or 3) undergoing infertility workup/gynaecological laparoscopy
Settings	Hospitals (public or private of any level): outpatient clinics (general gynaecology, reproductive medicine, pelvic pain) or research laboratories
Reference standard	Visualisation of endometriosis at surgery (laparoscopy or laparotomy), with or without histological confirmation
Study design	Cross sectional studies with a 'single-gate' design (n = 10) or a 'two-gate' design (n = 1); prospective enrolment; a single study could assess more than one test
Risk of bias and applicability concerns	Overall judgement: Poor quality of most of the studies (no study had a 'low risk' assessment in all four domains)
Patient selection bias	High risk: 1 study; Unclear risk: 5 studies; Low risk 5 studies
Index test interpretation bias	High risk: 9 studies; Unclear risk: 1 studies; Low risk 1 study
Reference standard interpretation bias	High risk: 0 studies; Unclear risk: 3 studies; Low risk 8 studies
Flow and timing selection bias	High risk: 3 studies; Unclear risk: 0 studies; Low risk 8 studies
Applicability concerns	Concerns regarding patient selection: high concern - 6 studies, unclear concern - 0 studies; low concern 5 studies; Concerns regarding index test: high concern - 0 studies, unclear concern - 1 study, low concern - 10 studies; Concerns regarding reference standard: high concern - 0 studies; unclear concern - 0 studies; low concern - 11 studies
Diagnostic criteria	Replacement test: sensitivity ≥ 94% and specificity ≥ 79% SnOUT triage test: sensitivity ≥ 95% and specificity ≥ 50% SpIN triage test: sensitivity ≥ 50% and specificity ≥ 95% Test with the diagnostic estimates within 5% of the set threshold were considered as approaching the criteria

Biomarker	N of studies; N of women	Outcomes				Diagnostic estimates [95% CI]	Implications
		True positives (endometriosis)	False positives (incorrectly classified as endometriosis)	False negatives (incorrectly classified as disease-free)	True negatives (disease-free)		
1. Tests for diagnosis of overall pelvic endometriosis							
1. IL-6 [serum] + PGP 9.5 [endometrium] for pelvic endometriosis, rASRM I-II --- cut-off IL-6 >15.4 pg/ml; PGP 9.5 - present; both tests positive	1; 78	38	3	0	37	Sens 1.00 [0.91, 1.00] Spec 0.93 [0.80, 0.98]	Meets criteria for a replacement and SnOUT triage test; approaches criteria for a SpIN triage test Insufficient evidence to draw meaningful conclusions
2. CA-125 [serum] + aromatase P450 [endometrium] for pelvic endometriosis, rASRM I-IV --- cut-off CA-125 >35 U/ml; aromatase - present; both tests positive	1; 58	33	7	3	15	Sens 0.92 [0.78, 0.98] Spec 0.68 [0.45, 0.86]	Approaches criteria for a SnOUT triage test; Insufficient evidence to draw meaningful conclusions
3. VDBP-Cr [urine] x CA-125 [serum] for pelvic endometriosis, rASRM I-IV --- cut-off > 2755; multiplication of both tests	1; 95	42	1	15	37	Sens 0.74 [0.60, 0.84] Spec 0.97 [0.86, 1.00]	Meets criteria for a SpIN triage test; Insufficient evidence to draw meaningful conclusions
4. NNE_Cr [urine] + CA-125 [serum] for pelvic endometriosis, rASRM III-IV --- cut-off > 27.23; sum of both tests	1; 59	30	3	9	17	Sens 0.77 [0.61, 0.89] Spec 0.85 [0.62, 0.97]	Insufficient evidence to draw meaningful conclusions

5. History + PV examination + TVUS for pelvic endometriosis, rASRM I-IV [focus on endometriosis with para-ovarian adhesions] --- Hx - dysmenorrhoea, dyspareunia; PV - presence of at least one of the following: pelvic tenderness, a fixed retroverted uterus, tender USL, deeply infiltrating nodules on USL or in POD; TVUS - fixed ovaries (ovaries did not move freely over the ipsilateral internal iliac vessels or pelvic sidewall or uterus with the gentle pressure); all tests positive	1; 106	34	27	3	42	Sens 0.92 [0.78, 0.98] Spec 0.61 [0.48, 0.72]	Approaches criteria for a SnOUT triage test; Insufficient evidence to draw meaningful conclusions
6. History + CA-125 [serum] + leukocytes [endometrium] for pelvic endometriosis, rASRM I-IV --- cut-off Hx - length of menses; CA-125 >35 U/ml; leukocytes - different cut-offs for each of the 8 leukocyte subsets; all tests positive	1; 368	106	10	67	185	Sens 0.61 [0.54, 0.69] Spec 0.95 [0.91, 0.98]	Meets criteria for a SpIN triage test; Insufficient evidence to draw meaningful conclusions
7. History + CA-125 [serum] for pelvic endometriosis, rASRM I-IV --- cut-off Hx - parity, past IUD, past endometriosis, alcohol intake, dyspareunia; CA-125 - not reported; both tests positive	1; 101	64	12	5	20	Sens 0.93 [0.84, 0.98] Spec 0.63 [0.44, 0.79]	Approaches criteria for a SnOUT triage test; Insufficient evidence to draw meaningful conclusions
2. Tests for diagnosis of DIE or ovarian endometriosis							
1. PV examination + CA-125 [serum] for DIE, endometrioma or severe adhesions --- cut-off PV - menstrual nodularities present; CA-125 ≥35 U/ml; both tests positive	1; 41	10	0	14	17	Sens 0.42 [0.22, 0.63] Spec 1.00 [0.80, 1.00]	Insufficient evidence to draw meaningful conclusions
2. PV examination OR CA-125 [serum] for DIE, endometrioma or severe adhesions	1; 41	21	3	3	14	Sens 0.88 [0.68, 0.97]	Insufficient evidence to draw

---------- cut-off PV - menstrual nodularities present; CA-125 ≥35 U/ml; either test positive						Spec 0.82 [0.57, 0.96]	meaningful conclusions
3. PV examination + CA125 [serum] for DIE ---------- cut-off PV - menstrual nodularities present; CA-125 ≥35 U/ml; both tests positive	1; 30	5	2	8	15	Sens 0.38 [0.14, 0.68] Spec 0.88 [0.64, 0.99]	Insufficient evidence to draw meaningful conclusions
4. PV examination OR CA-125 [serum] for DIE ---------- cut-off PV - menstrual nodularities present; CA-125 ≥35 U/ml; either test positive	1; 30	11	5	2	12	Sens 0.85 [0.55, 0.98] Spec 0.71 [0.44, 0.90]	Insufficient evidence to draw meaningful conclusions
5. PV examination + CA-125 [serum] for endometrioma ---------- cut-off PV - menstrual nodularities present; CA-125 ≥35 U/ml; both tests positive	1; 26	5	2	4	15	Sens 0.56 [0.21, 0.86] Spec 0.88 [0.64, 0.99]	Insufficient evidence to draw meaningful conclusions
6. PV examination OR CA-125 [serum] for endometrioma ---------- cut-off PV - menstrual nodularities present; CA-125 ≥35 U/ml; either test positive	1; 26	8	6	1	11	Sens 0.89 [0.51, 1.00] Spec 0.65 [0.38, 0.86]	Insufficient evidence to draw meaningful conclusions
3. Tests for differentiating ovarian endometriosis versus other benign ovarian cysts in women of reproductive age							
1. TVUS + CA-125 [serum] + CA-19.9 [serum] for endometrioma vs other ovarian cysts ---------- cut-off TVUS - presence of a round shaped homogeneous hypoechoic 'tissue' within the ovary; CA-125 ≥ 25 U/ml; CA-19.9 ≥12 U/ml; all tests positive	1; 118	19	1	20	78	Sens 0.49 [0.32, 0.65] Spec 0.99 [0.93, 1.00]	Approaches criteria for a SpIN triage test; Insufficient evidence to draw meaningful conclusions

2. TVUS + (CA-125 [serum] OR CA-19.9 [serum]) for endometrioma vs other ovarian cysts --- cut-off TVUS - presence of a round shaped homogeneous hypoechoic 'tissue' within the ovary; CA-125 ≥ 25 U/ml; CA-19.9 ≥ 12 U/ml; either blood test positive	1; 118	31	2	8	77	Sens 0.79 [0.64, 0.91] Spec 0.97 [0.91, 1.00]	Meets criteria for a SpIN triage test; Insufficient evidence to draw meaningful conclusions
3. TVUS + CA-19.9 [serum] for endometrioma vs other ovarian cysts --- cut-off TVUS - presence of a round shaped homogeneous hypoechoic 'tissue' within the ovary; CA-19.9 ≥ 12 U/ml; both tests positive	1; 118	21	2	18	77	Sens 0.54 [0.37, 0.70] Spec 0.97 [0.91, 1.00]	Meets criteria for a SpIN triage test; Insufficient evidence to draw meaningful conclusions
4. TVUS OR CA-19.9 [serum] for endometrioma vs other ovarian cysts --- cut-off TVUS - presence of a round shaped homogeneous hypoechoic 'tissue' within the ovary; CA-19.9 ≥ 12 U/ml; either test positive	1; 118	36	24	3	55	Sens 0.92 [0.79, 0.98] Spec 0.70 [0.58, 0.79]	Approaches criteria for a SnOUT triage test; Insufficient evidence to draw meaningful conclusions
5. TVUS + CA-125 [serum] for endometrioma vs other ovarian cysts --- cut-off TVUS - presence of a round shaped homogeneous hypoechoic 'tissue' within the ovary; CA-125 ≥ 20 U/ml; both tests positive	1; 101	20	3	9	69	Sens 0.69 [0.49, 0.85] Spec 0.96 [0.88, 0.99]	Meets criteria for a SpIN triage test; Insufficient evidence to draw meaningful conclusions
6. TVUS OR CA-125 [serum] for endometrioma vs other ovarian cysts --- cut-off TVUS - presence of a round shaped homogeneous hypoechoic 'tissue' within	1; 101	27	34	2	38	Sens 0.93 [0.77, 0.99] Spec 0.53 [0.41, 0.65]	Approaches criteria for a SnOUT triage test; Insufficient evidence to draw

the ovary; CA-125 ≥20 U/ml; either test positive							meaningful conclusions
7. TVUS + CA-125 [serum] for endometrioma vs other ovarian cysts --- cut-off TVUS - presence of a round shaped homogeneous hypoechoic 'tissue' within the ovary; CA-125 ≥ 25 U/ml; both tests positive	1; 101	20	3	9	69	Sens 0.69 [0.49, 0.85] Spec 0.96 [0.88, 0.99]	Meets criteria for a SpIN triage test; Insufficient evidence to draw meaningful conclusions
8. TVUS OR CA-125 [serum] for endometrioma vs other ovarian cysts --- cut-off TVUS - presence of a round shaped homogeneous hypoechoic 'tissue' within the ovary; CA-125 ≥ 25 U/ml; either test positive	1; 101	26	27	3	45	Sens 0.90 [0.73, 0.98] Spec 0.63 [0.50, 0.74]	Approaches criteria for a SnOUT triage test; Insufficient evidence to draw meaningful conclusions
9. TVUS + CA-125 [serum] for endometrioma vs other ovarian cysts --- cut-off TVUS - presence of a round shaped homogeneous hypoechoic 'tissue' within the ovary; CA-125 ≥35 U/ml; both tests positive	1; 101	15	2	14	70	Sens 0.52 [0.33, 0.71] Spec 0.97 [0.90, 1.00]	Meets criteria for a SpIN triage test; Insufficient evidence to draw meaningful conclusions
10. TVUS OR CA-125 [serum] for endometrioma vs other ovarian cysts --- cut-off TVUS - presence of a round shaped homogeneous hypoechoic 'tissue' within the ovary; CA-125 ≥35 U/ml; either test positive	1; 101	26	18	3	54	Sens 0.90 [0.73, 0.98] Spec 0.75 [0.63, 0.84]	Approaches criteria for a replacement and SnOUT triage test; Insufficient evidence to draw meaningful conclusions
4. Tests for mapping of DIE at specific anatomical locations							
1. PV examination + TVUS for POD obliteration	1; 200	26	3	4	167	Sens 0.87 [0.69, 0.96]	Meets criteria for a SpIN triage test;

--- PV - nodularity or stiffened or thickened area or a palpable cystic expansion in POD; TVUS - a. uterus, adnexa and rectosigmoid colon fixed to each other with disappearance of the peritoneal structure (complete POD obliteration); b. peritoneal limits partially identified with the presence or absence of suspended or lateralised fluid collection (incomplete POD obliteration); both tests positive						Spec 0.98 [0.95, 1.00]	Insufficient evidence to draw meaningful conclusions
2. PV examination + TVUS for vaginal endometriosis --- cut-off PV - nodularity or stiffened or thickened area or a palpable cystic expansion in vaginal wall; TVUS - thickening or the presence of a hypoechogenic cystic or non-cystic nodularity within the posterior vaginal wall	1; 200	18	1	4	177	Sens 0.82 [0.60, 0.95] Spec 0.99 [0.97, 1.00]	Meets criteria for a SpIN triage test; Insufficient evidence to draw meaningful conclusions
3. PV examination + TVUS for RVS endometriosis --- PV - nodularity or stiffened or thickened area or a palpable cystic expansion in RVS; TVUS - presence of a hypoechogenic nodularity or cystic mass within RVS (area between rectum and posterior vaginal wall from the level of introitus up to a level defined by the lower border of posterior lip of cervix); both tests positive	1; 200	7	2	1	190	Sens 0.88 [0.47, 1.00] Spec 0.99 [0.96, 1.00]	Meets criteria for a SpIN triage test; Insufficient evidence to draw meaningful conclusions
4. PV examination + TVUS for rectal endometriosis --- PV - nodularity or stiffened or thickened area or a palpable cystic expansion in rectosigmoid; TVUS - presence of a regular or irregular hypoechogenic mass distorting	1; 200	46	3	2	149	Sens 0.96 [0.86, 0.99] Spec 0.98 [0.94, 1.00]	Meets criteria for a SnOUT and SpIN triage test; Insufficient evidence to draw meaningful conclusions

	and replacing the normal appearance of the muscular layer of the rectal wall; both tests positive

(r)ASRM: (revised) American Society for Reproductive Medicine; **CA-125**: cancer antigen; **DIE**: deep infiltrating endometriosis; **IL**: interleukin; **IUD**: intrauterine device; **POD**: pouch of Douglas; **PV**: per vaginam; **TVUS**: transvaginal ultrasound; **USL**: uterosacral ligament; **VDBPCr**: vitamin-D-binding protein level corrected for creatinine.

BACKGROUND

Target condition being diagnosed

Endometriosis

Endometriosis is defined as an inflammatory condition characterised by endometrial-like tissue at sites outside of the uterus (Johnson 2013). Endometriotic lesions can occur at different locations, including the pelvic peritoneum and the ovary, or penetrate pelvic structures below the surface of peritoneum, as deeply infiltrating endometriosis. Each of these types of endometriosis are thought to represent a separate clinical entity, but also can coexist in the same woman. Rarely, endometriotic implants can be found at more distant sites, including lung, liver, pancreas and operative scars, with consequent variations in presenting symptoms.

Endometriosis afflicts 10% of women of reproductive age causing dysmenorrhoea (painful periods), dyspareunia (painful intercourse), chronic pelvic pain and infertility (Vigano 2004). The clinical presentation can vary from asymptomatic and unexplained infertility to severe dysmenorrhoea and chronic pain. These symptoms can occur with bowel or urinary symptoms, an abnormal pelvic examination or the presence of a pelvic mass, however no symptom is specific to endometriosis. The prevalence of endometriosis in a symptomatic population is reported as 35% to 50% (Giudice 2004).

Women with endometriosis are also at increased risk of developing several cancers (Somigliana 2006) and autoimmune disorders (Sinaii 2002). The presence of disease is associated with changes in the immune response, vascularisation, neural function, the peritoneal environment and the eutopic endometrium, suggesting that endometriosis is a systemic, rather than localised, condition (Giudice 2004). Endometriosis has a profound effect on psychological and social well-being and imposes a substantial economic burden on society. Women with endometriosis incur significant direct medical costs from diagnostic and therapeutic surgeries, hospital admissions and fertility treatments, however these costs are superceded by the indirect costs of endometriosis including absenteeism from work and loss of productivity (Gao 2006; Simoens 2012). In the USA, the financial burden of endometriosis is estimated at US $12,419 per woman (Simoens 2012).

Although the pathogenesis of endometriosis has not been fully elucidated, it is commonly thought that endometriosis occurs when endometrial tissue contained within the menstrual fluid flows retrogradely through the fallopian tubes and implants at an ectopic site within the pelvic cavity (Sampson 1927). However, this theory does not explain the fact that although retrograde menstruation is seen in up to 90% of women, only 10% of women develop endometriosis. There is evidence that a variety of environmental, immunological and hormonal factors are associated with endometriosis (Vigano 2004), and genetic loci that confer a risk of endometriosis have been identified (Nyholt 2012). The relative contribution of these and other causal factors remains to be elucidated.

Although it is impossible to time the onset of disease, on average, women have a six- to 12-year history of symptoms before obtaining a surgical diagnosis of endometriosis, indicative of considerable diagnostic delay (Matsuzaki 2006). Untreated endometriosis is associated with reduced quality of life and contributes to outcomes such as depression, inability to work, sexual dysfunction and missed opportunity for motherhood (Gao 2006).

Treatment of endometriosis

There is no cure for endometriosis. Treatment options include expectant management, pharmacological (hormonal) therapy and surgery (Johnson 2013). Treatment is individualised, taking into consideration the therapeutic goal (pain relief or conception), and the location of the disease. Current pharmacological therapies such as the combined oral contraceptive pill, progestogens, weak androgens and gonadotropin-releasing hormone (GnRH) agonists and antagonists act to reduce the effect of oestrogen on endometrial tissues and suppress menstruation. These drugs can ameliorate the symptoms of dysmenorrhoea and chronic pelvic pain, but are associated with side effects such as breast discomfort, irritability, androgenic symptoms and bone loss. Surgical excision of endometriotic lesions can reduce pain symptoms, however is associated with high recurrence rates of 40% to 50% at five years post-surgery (Guo 2009). Early treatment of endometriosis improves pain levels and physical and psychological functioning. Furthermore, improvements in menstrual management (the use of the Mirena coil and the continuous use of the combined contraceptive pill) and fertility preservation (oocyte vitrification) raise the possibility of suppressing the progression of endometriosis and prospectively managing subfertility in endometriosis sufferers. The potential success of these preventative strategies is dependent on an accurate and early diagnosis. A major impediment to earlier and more efficacious treatment of this disease is diagnostic delay due to the invasive nature of standard diagnostic tests (Dmowski 1997).

Clinical history and pelvic examination can raise the possibility of a diagnosis of endometriosis, but the heterogeneity in clinical presentation, the high prevalence of asymptomatic endometriosis (2% to 50%), and the poor association between presenting symptoms and severity of the disease contribute to the difficulty in obtaining a reliable diagnosis of endometriosis based solely on presenting symptoms (Ballard 2008; Fauconnier 2005; Spaczynski 2003). Although an abnormal pelvic examination correlates with the presence of endometriosis on laparoscopy in 70% to 90% of cases (Ling 1999), there is a wide differential diagnosis for most positive physical findings. Furthermore, a normal clinical examination does not exclude endometriosis, as laparoscopically-proven disease has been diagnosed in more than 50% women with a clinically normal pelvic examination (Eskenazi 2001). A variety of tests utilising pelvic imaging, blood markers, eutopic endometrium characteristics, urinary markers or peritoneal fluid components have been suggested as diagnostic measures for endometriosis. Although large numbers of the reported markers distinguish women with and without endometriosis in small pilot studies, many do not show convincing potential as a diagnostic test when they are evaluated in larger studies by different research groups. The diagnostic value of these tests has not previously been fully systematically evaluated and summarised using Cochrane methods. Currently, there is no simple non-invasive test for the diagnosis of endometriosis that is routinely implemented in clinical practice.

Surgical diagnostic procedures for endometriosis include laparoscopy (minimal access, or keyhole surgery) or laparotomy

(open surgery via an abdominal incision). In the last several decades, laparoscopy has become an increasingly common procedure and has largely replaced traditional open surgery in women suspected of having endometriosis (Yeung 2009). Laparoscopy has significant advantages over laparotomy creating fewer complications and shorter recovery times. Furthermore, a magnified view at laparoscopy allows better visualisation of the peritoneal cavity. Despite continuing controversy in the literature with regard to the superiority of one surgical modality over another in treating pelvic pathology, laparoscopy is the preferred technique to evaluate the pelvis and abdomen and to treat benign conditions such as ovarian endometriomas (Medeiros 2009). Surgery is currently also the only accepted way to determine the extent and severity of endometriosis. Several classification systems have been suggested for endometriosis (Adamson 2008; Batt 2003; Chapron 2003a; Martin 2006), but most researchers and clinicians use the revised American Society for Reproductive Medicine (rASRM) classification, which is internationally accepted as a respected tool for the objective assessment of the disease (American Society for Reproductive Medicine 1997). The rASRM classification system considers appearance, size and depth of peritoneal or ovarian implants and adhesions visualised during laparoscopy (Table 1) and allows uniform documentation of the extent of disease. Unfortunately this classification system has little value in clinical practice due to the lack of correlation between laparoscopic staging, the severity of symptoms and response to treatment (Chapron 2003b; Guzick 1997; Vercellini 1996). A recent endeavour to attain consensus around the optimal classification for endometriosis has been undertaken by the World Endometriosis Society (Johnson 2015).

The European Society of Human Reproduction and Embryology (ESHRE) Special Interest Group for Endometriosis stated in their guidelines for the diagnosis and treatment of endometriosis that for women presenting with symptoms suggestive of endometriosis, a definitive diagnosis of most forms of endometriosis requires visual inspection of the pelvis at laparoscopy as the 'gold standard' investigation (Kennedy 2005). Currently the visual or histological identification of endometriotic tissue in the pelvic cavity during surgery is not just the best available but the only diagnostic test for endometriosis in clinical practice.

The disadvantages of laparoscopic surgery include, but are not limited to, the high cost, the need for general anaesthesia and the potential for adhesion formation post procedure. Laparoscopy has been associated with a 2% risk of injury to pelvic organs, a 0.001% risk of damaging a major blood vessel and a mortality rate of 0.0001% (Chapron 2003c). Even though the major complications of laparoscopy are rare, it is difficult to determine the exact incidence of complications, and delayed recognition adds to surgical morbidity and mortality. Only one third of women who undertake a laparoscopic procedure will receive a diagnosis of endometriosis; therefore many disease-free women are unnecessarily exposed to surgical risk (Frishman 2006).

The validity of laparoscopy as a reference test for endometriosis has been assessed as being highly dependent on the skills of the surgeon. The diagnostic accuracy of laparoscopic visualisation has been compared with histological confirmation in a sole systematic review and it was estimated as having a sensitivity of 0.94 and specificity of 0.79 (Wykes 2004). Subsequent studies suggested that incorporation of histological verification in the diagnosis of endometriosis may improve diagnostic accuracy (Almeida Filho 2008; Marchino 2005; Stegmann 2008), but these papers have not been systematically reviewed. The clinical significance of histological verification remains debatable, and a diagnosis based on visual findings can be considered reliable with an accurate inspection of the abdominal cavity by properly trained and experienced surgeons (Redwine 2003). Furthermore, excised potential endometriotic tissues are rarely serially sectioned in clinical practice and small lesions can be missed by pathologists in mild disease. Thus sampling inconsistencies are also likely to influence the accuracy of histological reporting.

Summary

A diagnostic test in place of surgery would reduce associated surgical risks, increase diagnostic accessibility and improve treatment outcomes. The need for an accurate and non-invasive diagnostic test for endometriosis continues to encourage extensive research in the field and was endorsed at the international consensus workshop at the 10th World Congress of Endometriosis in 2008 (Rogers 2009). Although multiple markers and imaging techniques have been explored as diagnostic tests for endometriosis, none of them have been implemented routinely in clinical practice and many have not been subject to systematic review.

Index test(s)

This review assesses combinations of tests, including blood, urine and endometrial biomarkers and imaging modalities that have been proposed as non-invasive tests for the diagnosis of endometriosis (Table 2). This review is part of the review series on non-invasive diagnostic tests for endometriosis. The other reviews from this series are: '*Blood biomarkers for the non-invasive diagnosis of endometriosis*' (Nisenblat 2016a), '*Endometrial biomarkers for the non-invasive diagnosis of endometriosis*' (Gupta 2016), '*Urinary biomarkers for the non-invasive diagnosis of endometriosis*' (Liu 2015) and '*Imaging modalities for the non-invasive diagnosis of endometriosis*' (Nisenblat 2016b).

The definition of 'non-invasive' varies between medical dictionaries but refers to a procedure that does not involve penetration of skin or physical entrance to the body (McGraw-Hill Dictionary of Medicine 2006; The Gale Encyclopedia of Medicine 2008). Although intracavity imaging and tests involving venipuncture or endometrial sampling are invasive by this definition, when compared to diagnostic surgery for endometriosis, these tests are generally considered to be 'non-invasive' or 'minimally invasive'. For the purpose of these reviews, we will define all tests that do not involve anaesthesia and surgery as non-invasive.

The potential advantages of using imaging modalities, blood biomarkers, endometrial biomarkers, urinary biomarkers, clinical parameters that include examination findings and clinical history, or a combination of them to diagnose endometriosis, include their less invasive nature, lower cost and increased availability when compared to surgery. These tests are more acceptable to women, and usually provide a rapid result. However, the testing is dependant on the reliability of laboratory techniques and quality control protocols for the biomarker assays, on the skills of the operators performing imaging tests or examination and on women's access to appropriate radiology services.

The cellular and molecular processes that have been identified to characterise ectopic endometrium and peritoneal fluid in human and animal models (D'Hooghe 2001; Hull 2008; Kao 2003) have inspired the use of markers of these pathophysiological processes present in blood, urine and endometrium samples as a single test or a combination of several biomarkers. Of these tests, urinary biomarker discovery is a new and rapidly expanding field with most studies published in the last five years. Several large systematic reviews of all proposed biomarkers for endometriosis identified multiple putative biomarkers, but none of these biomarkers could be recommended for use in clinical practice (May 2010; May 2011), which was supported by a more recent narrative review (Fassbender 2015). The biomarker research in endometriosis tends to shift towards diagnostic panels which include one or several testing modalities such as blood, endometrial or imaging tests. Systematic reviews on imaging in endometriosis (Guerriero 2015; Hudelist 2011a;Medeiros 2015; Moore 2002) and narrative reviews on the topic primarily addressed diagnostic performance of imaging methods and not as a part of a diagnostic panel. In line with general consensus, clinical parameters (history and examination) have low reliability in the diagnosis of endometriosis, however they may improve the diagnostic performance of other non-invasive tests when incorporated in a diagnostic model. So far, combinations of non-invasive tests have only been assessed in a limited number of small studies, which vary in the type of methodology and tests used and type of endometriosis evaluated. There is a current need to evaluate the diagnotic test accuracy of the combination of different testing modalities and diagnostic algorithms for endometriosis using Cochrane methods.

Clinical pathway

Women presenting with symptoms of endometriosis (dysmenorrhoea, dyspareunia, chronic pelvic pain or difficulty conceiving) generally are investigated with a pelvic ultrasound scan to exclude other pathologies, which is in line with international guidelines (Dunselman 2014; SOGC 2010; ACOG 2010). There are no other standard investigative tests, and although evidence suggests that magnetic resonance imaging (MRI) is superior to ultrasound, it is used conservatively because of its cost. If women seek pain management rather than conception, physicians generally initiate empirical treatment with progestogens or the combined oral contraceptive pill. Diagnostic laparoscopy is considered if empirical treatment fails or if women decline or do not tolerate empirical treatment. In women who have difficulty conceiving, laparoscopy can be undertaken before fertility treatment (particularly if severe pelvic pain or endometrioma are present) or after failed assisted reproductive technology (ART) treatments. Endometriosis can be also diagnosed during fertility investigations in women who have minimal or no pain symptomatology.

On average there is a delay of between six to 12 years from onset of symptoms to definitive diagnosis at surgery (Dunselman 2014). Early referral to a gynaecologist with the capability to perform diagnostic surgery is expected to reduce time to diagnosis. Collectively, young women, women in remote and rural locations and women of lower socioeconomic status have reduced access to surgery, and are less likely to obtain a prompt diagnosis of endometriosis.

Prior test(s)

Most women presenting with symptoms suggestive of endometriosis have a full history and examination and a routine gynaecological ultrasound before a decision is made to have diagnostic surgery. However, there is no consensus on whether or not ultrasound or any other test should be routinely used as part of a standardised approach.

Role of index test(s)

A new diagnostic test can fulfil one of three roles.

1. Replacement: replacing an existing test by having more accuracy, or a similar accuracy with other advantages.
2. Triage: used as an initial step in a diagnostic pathway to identify the group of women who need further testing with an existing test. Although ideally a triage test has a high sensitivity and specificity, it may have a lower sensitivity but higher specificity than the current test or vice versa. The triage test does not aim to improve the diagnostic accuracy of the existing test but rather to reduce the number of individuals having an unnecessary diagnostic test.
3. Add-on: used in addition to existing testing to improve diagnostic performance (Bossuyt 2008).

Ideally, a diagnostic test is expected to correctly identify all women with a disease and to exclude all women without that disease, in other words it should have a sensitivity and specificity of 1.00. A high sensitivity indicates that there are a low number of women who have a negative test and do have the disease (i.e. a low number of false-negative results). High specificity corresponds to a low number of women who have a positive test but do not have the disease (i.e. low false-positive results). In practice, however, it is extremely rare to find a test with equally high sensitivity and specificity. An acceptable replacement test would need to have a similar or higher sensitivity and specificity than the current gold standard of laparoscopy. The only systematic review that determines the accuracy of laparoscopy in diagnosing endometriosis reported a sensitivity of 0.94, and a specificity of 0.79 (Wykes 2004) and we have taken this as a cut-off for a replacement test.

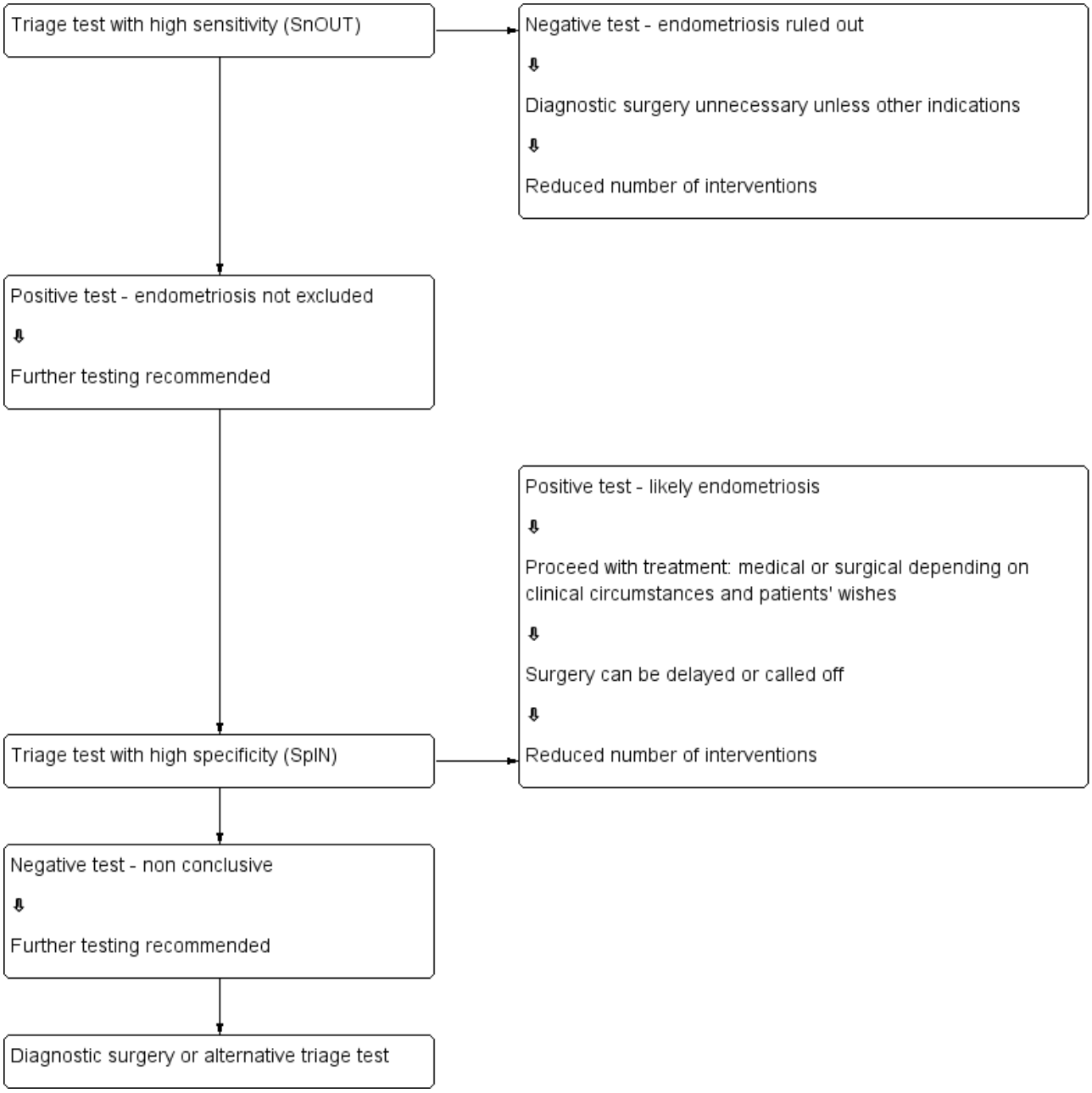
The purpose of triage tests can vary depending on the clinical context and a woman's priorities. One reasonable approach is to exclude the diagnosis to avoid further unnecessary and expensive diagnostic investigation. High-sensitivity tests have few false negative results and act to rule conditions out (SnOUT). A negative result from a test with high sensitivity will exclude the disease with high certainty independent of the specificity. As women without disease would be assured of having a negative test, unnecessary invasive interventions can be avoided. However, a positive result has less diagnostic value particularly when the specificity is low. We predetermined that a clinically useful SnOUT triage test should have a sensitivity of 0.95 or more and a specificity of 0.50 and above. We set the sensitivity cut-off for a SnOUT triage test at 0.95 and above, assuming that a 0.05 false negative rate is statistically and clinically acceptable. We set the specificity cut-off at 0.50 and above, to avoid diagnostic uncertainty in more than 50% of the population with a positive result.

An alternative approach would be to avoid a missed diagnosis. High-specificity tests have few false positive results and act to

rule conditions "in" (SpIN). A positive result for a highly specific triage test indicates a high likelihood of having endometriosis. This information could be used to prioritise these women for surgical treatment. A positive SpIN test could also provide a clinical rationale to start targeted disease-specific medical management in a woman without a surgical diagnosis, under the assumption that disease is present. Surgical management could then be reserved for cases when conservative treatment fails. This is particularly relevant in some populations where the therapeutic benefits of surgery for endometriosis have to be carefully balanced with the disadvantages (e.g. young women, women with medical conditions or pain-free women with a history of infertility). In this scenario we considered a sensitivity of 0.50 and above and a specificity of 0.95 and higher as suitable cut-offs for a SpIN triage test.

We evaluated combinations of tests for their potential to replace surgery (replacement test) or to improve the selection of women for surgery (triage test to rule out (SnOUT) or rule in (SpIN) the disease). Both types of triage test are clinically useful, minimising the number of unnecessary interventions. Sequential implementation of SnOUT and SpIN tests can also optimise a diagnostic algorithm (Figure 1). We did not assess any test as an add-on test, as we sought tests that reduce the need for surgery and not tests that improve the accuracy of the currently available surgical diagnosis.

Figure 1. Sequential approach to non-invasive testing of endometriosis



Alternative test(s)

There are no alternative tests for the diagnosis of endometriosis that are available in routine clinical practice.

Rationale

Many women with endometriosis suffer long-standing pelvic pain and infertility prior to a diagnosis. Surgery is the only current method of diagnosing endometriosis, but it is associated with high costs and surgical risks. Simple and reliable non-invasive tests for endometriosis, with the potential to either replace laparoscopy or to triage women in order to reduce surgery, would minimise surgical risk and reduce diagnostic delay. Physicians could then detect endometriosis at less advanced stages and institute earlier interventions. Early diagnosis would provide the opportunity for a preventive approach for this debilitating disease, potentially reducing healthcare-related costs and favouring more cost-effective and efficient treatments. Furthermore, identifying the tests that do not pertain to endometriotic disease would help clinicians and researchers focus on clinically relevant biomarker detection.

OBJECTIVES

Primary objectives

To estimate the diagnostic accuracy of any combination of non-invasive tests for the diagnosis of pelvic endometriosis (peritoneal and/ or ovarian or deep infiltrating) compared to surgical diagnosis as a reference standard. The combined tests were evaluated as replacement tests for diagnostic surgery as well as triage tests which would assist decision-making to undertake diagnostic surgery for endometriosis.

Secondary objectives

To investigate the influence of heterogeneity on the diagnostic accuracy of combined non-invasive test for endometriosis. Potiential sources of heterogeneity include:

1. characteristics of the study population: age (adolescents versus later reproductive years); clinical presentation (subfertility, pelvic pain, ovarian mass, asymptomatic women); stage of disease (rASRM classification system); geographic location of study;
2. histological confirmation in conjunction with laparoscopic visualisation compared to laparoscopic visualisation alone;
3. changes in technology over time: year of publication; modifications applied to conventional laboratory techniques;
4. methodological quality: differences in the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies-2) evaluation (Table 3), including a) low versus unclear or high risk; b) consecutive versus non-consecutive enrolment; c) blinding of surgeons to the results of index tests;
5. study design ('single-gate design' versus 'two-gate design' studies).

METHODS

Criteria for considering studies for this review

Types of studies

Published peer-reviewed studies that compared the results of a combination of several testing modalities with the results obtained from a surgical diagnosis of endometriosis.

Studies were included if they included the following study designs.

1. Randomised controlled trials (RCTs).
2. Observational studies with the following designs.
   a. Single-gate design (studies with a single set of inclusion criteria defined by clinical presentation). All participants had clinically suspected endometriosis.
   b. Two-gate design (studies where participants are sampled from distinct populations with respect to clinical presentation). The same study includes participants with a clinical suspicion of having the target condition (e.g. women with pelvic pain) and also participants in whom the target condition is not suspected (e.g. women admitted for tubal ligation). Two-gate studies were eligible only where all cases and controls belonged to the same population with respect to the reference standard (i.e. all the participants were scheduled for laparoscopy) (Rutjes 2005).
3. For studies on biological samples - performed on prospectively collected samples, irrespective of the actual time of the test assay. The timing of sample collection relative to surgery is important because the surgical excision of endometriotic lesions could influence biomarker expression and hence bias the results. Therefore, we only included studies where the biological sample was collected before the surgical procedure, i.e. prospectively collected. We considered to be eligible the studies performed on tissue bank samples collected from prospectively recruited, well-defined populations, which prevented the omission of valuable data from adequately designed studies. The time interval between sample collection and laboratory testing may influence test outcomes, which could be dependent on sample storage conditions and the stability of each individual biomarker during storage and freeze-thawing. This information was not readily available for most molecules, and we did not address it in this review, but we will consider it in future updates if more evidence emerges.
4. For studies on clinical or imaging examination - performed on prospectively recruited women with the index test being completed prior to the reference standard.

We did not impose limits on eligibility related to the healthcare settings where the study took place, the language of publication, the number of participants in the included studies or the number of studies that evaluated each index test.

The following studies were excluded.

1. Narrative or systematic reviews.
2. Studies of retrospective design where the sample collection, clinical or imaging examination were performed after execution of reference test.
3. Studies of retrospective design where the participants were selected from retrospective review of the case notes/ archived

samples and information on recruitment methods or study population was not available.
4. Case reports or case series.
5. Studies reported only in abstract form or in conference proceedings where the full text was not available. We applied this limitation after facing substantial difficulty in obtaining the information from the abstracts, which precluded a reliable assessment of eligibility and methodological quality.

Participants

Study participants included women of reproductive age (puberty to menopause) with suspected endometriosis based on clinical symptoms or pelvic examination, who undertook both the index test and reference standard.

The participants were selected from populations of women undergoing abdominal surgery for the following indications: 1) clinically suspected endometriosis (pelvic pain, infertility, abnormal pelvic examination, or a combination of the above); 2) ovarian mass, regardless of symptoms; 3) a mixed group, which consists of women with suspected endometriosis/ovarian mass or women with other benign gynaecological conditions (e.g. surgical sterilisation, fibroid uterus, etc). Asymptomatic women who had an incidental finding of endometriosis at surgery performed for another indication were also included.

Studies that included participants of postmenopausal age were eligible when the data for the reproductive age group was available in isolation. We excluded studies with participants that clearly would not undergo the index test in the relevant clinical situation or would not benefit from the test (e.g. women with ectopic pregnancies or acute pelvic inflammatory disease). We also excluded publications that only analysed participants with a positive index test or reference standard and did not provide data for the whole cohort.

Index tests

We assessed any combination of non-invasive tests for endometriosis comprising of more than one test modality. This included the combinations of blood, endometrial, urine and imaging tests with or without clinical parameters, such as pre-defined examination findings, specific symptoms or characteristics (e.g. length of menstrual cycle). The assessed index tests are presented in Table 2.

The panel of biomarkers from the same single category (e.g. several blood biomarkers or combination of imaging methods) was assessed in the relevant review on the topic and are presented separately in other reviews from this series. The studies that solely assessed specific technical aspects, qualitative description of lesion appearance or interobserver variability of the index tests without reporting the data on diagnostic performance were excluded from the review. When the evaluated biomarker(s) showed differential expression between the groups of women with and without endometriosis, the publication was considered only if the data were reported with sufficient detail for the construction of 2 x 2 contingency tables. However, when the contingency tables were not available because the expression level of index test did not significantly differ between the groups and the inclusion criteria were otherwise met, we made a critical appraisal and presented the study in the descriptive part of the review. Thus, we evaluated the adequately designed studies that identified biomarkers without diagnostic value, as they provide information that is likely to focus future research on other more clinically useful biomarkers.This methodology also identified biomarkers that were associated with endometriosis in some but not other studies. We did not include evaluations of screening or predictive accuracy tests in this review.

We considered the diagnostic performance of an index test to be high when the test reached the criteria for a replacement test (sensitivity of equal or greater than 0.94 with specificity of equal or greater than 0.79) or triage test (sensitivity of equal or greater than 0.95 with specificity of equal or greater than 0.50 or vice versa) or approached these criteria (diagnostic estimates within 0.05 of the set thresholds). We considered all other diagnostic estimates to be low.

Target conditions

Pelvic endometriosis, defined as endometrial tissue located in the pelvic cavity: involving any of the pelvic organs, peritoneum and pouch of Douglas (POD).

Three types of pelvic endometriosis were assessed.

1. Peritoneal endometriosis, defined as endometrial deposits detected on the peritoneum covering pelvic organs, pelvic side walls or POD.
2. Ovarian endometriosis (endometrioma), defined as an ovarian cyst lined by endometrial tissue, appearing as an ovarian mass of varying size.
3. Deep infiltrating endometriosis (DIE), defined as subperitoneal infiltration of endometrial implants, i.e. when the endometriotic implants penetrate the retroperitoneal space at a distance of 5 mm or more (Koninckx 1991). DIE may be present in multiple locations, involving either the anterior or posterior pelvic compartments, or both.

We did not include certain rare types of endometriosis such as extrapelvic, bladder and ureteric endometriosis because the majority were reported in case reports or case series, and laparoscopy or laparotomy are not reliable reference standards for these conditions.

We excluded the studies where diagnosis of endometriosis was not the primary outcome (e.g. malignant versus benign masses or normal versus abnormal pelvis) and the separate data for endometriosis were not available.

We also excluded the studies where the findings of the index test formed the basis of selection for the reference standard, because this was likely to distort an assessment of the diagnostic value of the index test.

We did include studies that recruited selected populations of women with endometriosis (i.e. those with specific rASRM stages), because there is a poor correlation between the rASRM classification and infertility or pain symptoms. Exclusion of these studies could result in a loss of potentially important diagnostic information from otherwise eligible publications. Where possible, the impact of these studies was addressed in the assessments of heterogeneity. When a study analysed a large population with a wide spectrum of endometriosis and additionally reported a subgroup analysis of the different stages of disease severity, we only considered estimates for the entire population. This is because

a subgroup analysis would not directly address the review question regarding the clinical utility of the biomarker in disease detection.

Reference standards

The reference standard was visualisation of endometriosis at surgery (laparoscopy or laparotomy) with or without histological confirmation, as this is currently the best available test for endometriosis. Information regarding the inter- and intra-observer correlation of the reference standard was reviewed if reported.

We only included studies in which the reference test was performed within 12 months of the sample collection or imaging test, on the assumption that the disease status could change within a period of one year or longer, either naturally or as a result of treatment. We excluded studies in which the participants did not undergo the reference standard or where the findings of the index test formed the basis of selection for undertaking the reference standard, as this was likely to distort an assessment of the diagnostic value of the index test.

Summary of inclusion/exclusion criteria

Inclusion criteria

1. Types of studies
   a. Published and peer-reviewed
   b. RCTs
   c. Observational designs, including:
      i. single-gate design (single set of inclusion criteria defined by clinical presentation): all the participants had clinically suspected endometriosis;
      ii. two-gate design (two sets of inclusion criteria with respect to clinical presentation and one set of inclusion criteria with respect to reference standard): the participants with or without a clinical suspicion of endometriosis scheduled for abdominal surgery.
   d. Performed on prospectively collected samples, including the tissue bank samples collected from a prospectively recruited well-defined population; for clinical/imaging testing - performed on prospectively recruited participants when index test performed before reference standard
   e. Published in any language
   f. Performed in any healthcare setting
   g. Any sample size
2. Participants
   a. Women of reproductive age
   b. Clinically suspected endometriosis, this also included:
      i. women who underwent abdominal surgery for other benign gynaecological conditions and had a surgical assessment for presence/absence of endometriosis;
      ii. asymptomatic women who have an incidental finding of endometriosis at surgery performed for another indication.
   c. Undertook both the index test and reference standard
3. Index tests
1. a. Combined non-invasive tests for endometriosis comprising of several testing modalities, including the combinations of blood, endometrial, urine and imaging tests with or without clinical parameters
   b. Data reported in sufficient detail for the construction of 2 x 2 tables for the tests that showed differential expression between the groups
   c. Tests where a 2 x 2 table could not be constructed because the results did not differ between women with and without endometriosis, but all other inclusion criteria were met
2. Target condition
   a. Pelvic endometriosis
      i. peritoneal endometriosis;
      ii. ovarian endometrioma;
      iii. DIE;
      iv. combinations of the above.
3. Reference standard
   a. Surgical visualisation of lesions for the diagnosis of endometriosis (laparoscopy or laparotomy) with or without histological verification
   b. Performed within 12 months of the endometrial sample collection

Exclusion criteria

1. Types of studies
   a. Narrative or systematic reviews
   b. Retrospective design where biological samples were collected or clinical/ imaging index test was performed after execution of reference test
   c. Prospectively collected samples that were selected from the archived material, but information on the study population or the selection process was unclear
   d. Case reports or case series
   e. Conference proceeding
2. Participants
   a. Included cohort was not representative of the target population that would benefit from the test (e.g. women with known genital tract malignancy, ectopic pregnancies or acute pelvic inflammatory disease)
   b. Study included participants of postmenopausal age and the data for the reproductive age group were not available in isolation
   c. Analysis only included participants with positive index test or positive reference standard
3. Index tests
   a. Biomarkers presented as a single test or a panel of several markers from the same category (e.g. only blood biomarkers)
   b. Study presented only specific technical aspects of an index test or focused on the biological events, rather than diagnostic performance of the test
   c. Study assessed screening or predictive test accuracy
4. Target condition
   a. Endometriosis was not the primary outcome of the trial (e.g. malignant versus benign masses or normal versus abnormal pelvis)
   b. Atypical, rare sites of endometriosis
5. Reference standard
   a. Reference standard performed only in a subset of the study/ control group
   b. Findings of the index test formed the basis of selection for the reference standard

c. Rather than specified in inclusion criteria

Search methods for identification of studies

We developed the search strategy in collaboration with the Trials Search Co-ordinator of the Gynaecology and Fertility Review Group, following recommendations of the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (de Vet 2008). We did not limit the searches to particular types of study design or impose language or publication date restrictions. The search strategy used a combination of both free text words and index terms. We initially created the search for one broad review looking at all diagnostic markers for endometriosis, but due to complexity, the review team split the originally planned review into five separate reviews. We designed two separate search strategies: one for all the biomarkers-based tests, and another for the imaging tests; both strategies were utilised in this review. We searched CENTRAL to July 2015 and performed all other searches from database inception to April 2015. We present the search strategies for each database and the number of hits per search in Appendix 1; Appendix 2; Appendix 3; Appendix 4; Appendix 5; Appendix 6; Appendix 7; Appendix 8; Appendix 9; Appendix 10. The summary of the results is presented in Results of the search.

Electronic searches

We searched the following databases to identify the published articles that assessed the diagnostic value of non-invasive tests for endometriosis.

a. CENTRAL (2015, July).
b. MEDLINE (inception to May 2015).
c. EMBASE (inception to May 2015).
d. CINAHL (inception to April 2015).
e. PsycINFO (inception to April 2015).
f. Web of Science (inception to April 2015).
g. LILACS (inception to April 2015).
h. OAIster (inception to April 2015).
i. TRIP (inception to April 2015).
j. Databases of the trial registers:
    i. ClinicalTrials.gov (inception to April 2015);
    ii. World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) (inception to April 2015).
k. Databases to identify reviews and guidelines as sources of references to potentially relevant studies.
    i. MEDION (inception to January 2014, the last available date);
    ii. DARE (inception to April 2015);
    iii. PubMed, a 'Systematic Review' search under the 'Clinical Queries' link (inception to April 2015).
l. Searches for papers recently published and not yet indexed in the major databases:
    i. PubMed (simple search for the six months to April 2015).

Searching other resources

We handsearched the reference list of all relevant publications (retrieved full texts of the key articles and identified reviews).

We abandoned an initial attempt to locate the grey literature (unpublished studies and conference proceedings), as we faced substantial difficulty in obtaining full-text publications or further details of studies reported in an abstract form.

Data collection and analysis

Selection of studies

Two authors of this review (LP, VN) and four other authors or contributors of the other reviews from this series (Devashana Gupta, Emily Liu, Rabia Shaikh and Deepika Arora) scanned the titles of studies identified by our search to remove any clearly irrelevant articles. The titles and abstracts of the remaining studies were reviewed to select potentially relevant publications. The relevant articles were then divided into four categories of endometriosis biomarkers: serum, endometrial, urinary and combined tests (imaging had already been completed in a separate search). Three of the combined biomarker review authors (LP, NJ, VN) independently reviewed each of the full-text versions of the articles selected by title and abstract and assessed them for eligibility for inclusion, based on the criteria listed above under Criteria for considering studies for this review. A single failed eligibility criterion was sufficient for a study to be excluded from the review.

The review authors who assessed the relevance of the studies and eligibility for inclusion were not blind to the information about each article, including the publishing journal, the names of authors, the institution and the results. Any disagreements were resolved by discussion.

When papers updated previous publications and were performed on the same study population at different recruitment points, the most complete data set that superseded previous publications was used to avoid double counting participants or studies. Missing data were retrieved by directly contacting authors to clarify study eligibility. When potentially relevant studies were found in languages other than English, a translation was undertaken. For excluded studies, the reasons for exclusion and details of which criteria were not met were documented. The characteristics of included and excluded studies are presented under Characteristics of included studies and Characteristics of excluded studies, respectively.

Data extraction and management

Data were independently extracted from eligible studies by three review authors (LP, NJ, VN) and any disagreement was resolved by consensus. If required, we contacted study investigators to resolve any questions regarding the data.

To collect details from included studies, a data extraction form was specifically designed for this review and pilot-tested on three studies of diagnostic accuracy tests for endometriosis. The following information was recorded for each study.

1. General information and study design: first author, year of publication, country, language, setting, objectives, inclusion/exclusion criteria, type of enrolment.
2. Characteristics of the study participants: age, symptoms/history/previous tests, type of target condition and its prevalence in the study population, number of participants enrolled and available for analysis, reasons for withdrawal.
3. Features of the index test and reference standard: type, diagnostic criteria, number and experience of the operators,

blinding of the operators to other tests or clinical data, interobserver variability, time interval between index test and reference standard.

4. The reported number of true positives (TP), false negatives (FN), true negatives (TN) and false positives (FP) was used to construct a two-by-two (2 x 2) table for each index test. If these values were not reported, we attempted to reconstruct the 2 x 2 tables from the summary estimates presented in the article.

Data were extracted into Review Manager® (RevMan) software, which was used to display graphically the quality assessment, the diagnostic estimates data, and the descriptive analyses.

Assessment of methodological quality

To assess the quality of each included study, we used QUADAS-2, a modified version of the QUADAS tool for systematic reviews of diagnostic accuracy studies (Whiting 2011).

The review-specific QUADAS-2 tool and explanatory document are presented in Table 3. Each paper was judged as having a 'low', 'high' or 'unclear' risk for each of four domains and concerns about applicability were assessed in three domains. We considered studies as having low methodological quality when they were at high or unclear risk of bias or when we had a high concern regarding applicability at least in one domain. The assessment of each included study was performed independently by three review authors (LP, NJ, VN) and disagreements were settled by consensus. Two review authors (LP, NJ) independently piloted the topic-specific tool to rate four of the included studies with a high level of agreement. Modifications specific to the combined biomarkers review were made to the signalling questions of the original QUADAS-2 tool and were as following.

Domain 1

We rephrased an original signalling question, 'Was a case-control design avoided?' as 'Was a two-gate design avoided?'. The diagnostic studies are cross-sectional in nature, aiming to compare the result of an index test with the result of the reference standard in the same group of participants. Study investigators measure the parameters at a single point in time and classify the groups by the outcome of the reference standard, albeit they perform the analysis retrospectively. Therefore, unlike epidemiological studies, the terminology 'cohort' and 'case-control' is less informative for diagnostic test trials, so we substituted them for 'single-gate' and 'two-gate' designs. We included this question because a two-gate design has more potential to introduce selection bias.

Domain 2

For the biomarker studies

2.1. We introduced an additional signalling question, 'Was the phase of the menstrual cycle considered in interpreting the index test?' to assess bias in the interpretation of the test results. Some biochemical markers are sensitive to fluctuation in steroid sex hormone levels across a menstrual cycle, which could result in the differential expression of endometriosis biomarkers at different cycle phases.

For the studies on clinical/imaging tests

2.2 We introduced an additional signalling question 'Was the index test performed by a single operator?' to assess interobserver variation bias.

2.3 We introduced an additional signalling question 'Were the same clinical data available when the index test results were interpreted as that which would be available when the test is used in practice?' to assess a bias in clinical applicability.

2.4 We rephrased an original signalling question, 'If a threshold was used, was it pre-specified?' as 'Did the study provide a clear pre-specified definition of what was considered to be a positive index test result' because this question was more applicable to imaging modalities.

We undertook the assessment of methodological quality for each domain, but we did not calculate a summary score to estimate the overall quality of studies (Whiting 2005).

Statistical analysis and data synthesis

The estimates of sensitivity and specificity were generated in forest plots and plotted in the receiver operating characteristic (ROC) space for each index test using Review Manager 5 software (RevMan 2014). The diagnostic performance of each test was investigated and inter-study variation in the performance of each index test was visually explored in relation to woman characteristics, study design, and study quality factors. Two or more tests evaluated in the same cohort were included as separate data sets, since the unit of analysis was the test result, not the woman.

For studies that reported subgroup analyses per phase of the menstrual cycle, we presented the data in a clinically relevant way. For instance, we presented pooled estimates when there was no statistically significant difference in biomarker expression between cycle phases. Alternatively, where putative biomarkers demonstrated cycle-dependent expression or were noted to be modulated by ovarian hormones, we reported the test performance either at several time points across the menstrual cycle or in the phase that demonstrated the most distinct difference between groups.

We planned to perform the bivariate logit normal random-effects model for all meta-analyses with four studies or more and a fixed-effect meta-analysis of sensitivity and specificity for smaller groups of studies (two or three) in the absence of substantial heterogeneity. The meta-analyses were planned to be performed using SAS NLMIXED software (Cary, NC: SAS Institute Inc) in order to provide plots of the estimated summary points of sensitivity and specificity and confidence regions. In this review a meta-analysis was not performed due to the paucity of data for each combination of non-invasive tests.

The comparative accuracy of index tests was assessed in two ways. In direct, fully-paired comparisons where all the study participants received more than one index test as well as the reference standard, the estimates were plotted in RevMan. If a meta-analysis was possible, test-level covariates in the bivariate logit normal model were used to identify statistically significant differences. Otherwise, the available comparative data were reported in a narrative way and illustrated using forest and ROC plots.

When test performance was judged against the predetermined diagnostic criteria, the point estimates of sensitivity and specificity

were considered as the most informative presentation of test performance. We acknowledge that tests with point estimates that did not reach the predetermined criteria but confidence intervals (CIs) which contained values above the threshold, could have diagnostic value. Furthermore, tests with point estimates that reached the criteria but CIs which contained values below the threshold, could have an overestimated diagnostic value. If the range of the CIs rather than the point estimates of the data are used, the predetermined cut-off becomes meaningless. Therefore, we did not consider CIs in qualifying the test performance, but utilised this information in interpreting the reliability of the obtained data.

Dealing with missing data

Missing data were defined as any information on the study population, index tests or reference standard that was not available in the publication which was required to determine the eligibility of the study for inclusion, the methodological quality or to construct the results table. If missing data were identified, we contacted the authors in an attempt to obtain this information. If missing data prevented a clear judgment regarding applicability for inclusion or the construction of accurate 2 x 2 tables and the data were not available from the primary investigators (for example, we were unable to locate the contact details of the authors or there was no reply from the authors or the authors replied that the requested information was unavailable), we excluded the study from the review.

Investigations of heterogeneity

We planned to assess heterogeneity by visually examining the forest plots of sensitivities and specificities and the ROC plots for each index test. The potential sources of heterogeneity are stated in the Secondary objectives. For diagnostic tests with more than 10 eligible studies, we planned to formally explore heterogeneity by using study-level covariates. We were unable to assess sources of heterogeneity in this review because there was only one study for each test. We also planned to assess the sensitivity of results to the inclusion and exclusion of outlying studies in all analyses, but refrained from doing so, again because of the small number of studies for most analyses.

Sensitivity analyses

We planned to conduct sensitivity analyses to assess the impact of the methodological quality of included studies on the results of any meta-analyses if sufficient data were available. Low-quality studies were defined by the identification of a high risk of bias for one or more QUADAS-2 domains. We also planned to use the 'leave-one-out' procedure to assess the impact of each study on the meta-analysis results (leading study effect). In this review we were unable to undertake sensitivity analyses due to the paucity of studies evaluating each biomarker.

Assessment of reporting bias

A comprehensive search of multiple sources for eligible studies, a search of trial registers and no language restrictions minimised the risk of reporting bias. However, publication bias generally arises when studies have a higher chance of being published if their results are positive. Therefore unpublished and published study databases and conference proceedings were initially searched and evaluated. During the process of qualifying the studies for inclusion in this review, we faced substantial difficulty in obtaining full-text publications or further details of studies published in an abstract form. This precluded a reliable assessment of eligibility and methodological quality and it was decided not to include these publication sources in this review.

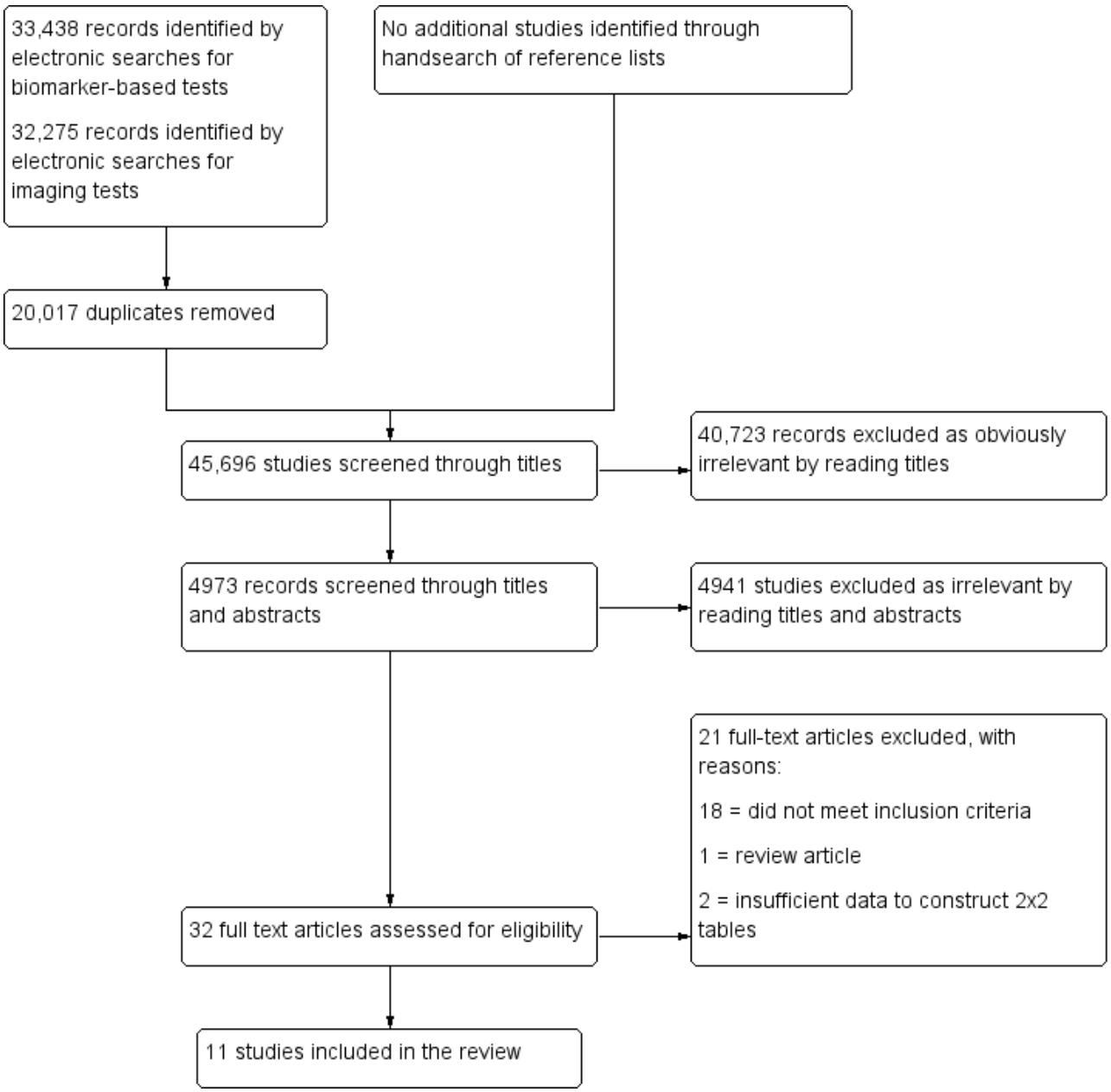
RESULTS

Results of the search

The literature search identified 33,438 references for the biomarker-based tests in the following databases: CENTRAL (n = 226), MEDLINE (n = 10,328), EMBASE (n = 10,313), CINAHL (n = 1131), PsycINFO (n = 174), Web of Science (n = 7425), LILACS (n = 420), OAIster (n = 446), Trip (n = 1648), trial registers for ongoing and registered trials (n = 523), MEDION (n = 2), DARE (n = 99), PubMed, a 'systematic review' search (n = 418) and simple search PubMed (n = 267). For the imaging tests, the search identified 32,275 references as following: CENTRAL (n = 445), MEDLINE (n = 7391), EMBASE (n = 12,161), CINAHL (n = 668), PsycINFO (n = 174), Web of Science (n = 7425), LILACS (n = 420), OAIster (n = 446), TRIP (n = 1648), trial registers for ongoing and registered trials (n = 523), MEDION (n = 190), DARE (n = 99), PubMed, a 'systematic review' search (n = 418) and simple search PubMed (n = 267). These databases were searched from inception to 20 April - 31 July 2015.

The flow of the selection process is presented in Figure 2. Titles were screened to exclude duplicates (n = 20,017) and clearly irrelevant studies (n = 40,723). A further 4941 references were eliminated after the abstracts were reviewed because either they did not address the research question or they clearly did not meet the inclusion criteria. The full texts of the remaining 32 references were retrieved and assessed for eligibility. Data from five studies required additional clarification from the authors and two non-English publications were translated. Ultimately, 11 studies were eligible and provided data for the review and 21 studies were excluded.

Figure 2. Flow of the studies identified in literature search for systematic review on combination of non-invasive tests for diagnosis of endometriosis.



Basic features of the included studies

The list and details of the included studies are presented in Characteristics of included studies. The 11 eligible studies included 1339 participants, with a median of 101 women per study (range 55 to 368). Of these studies, four were conducted in Europe, four in Asia, one in Australia, one in North America and one in the Middle East. Ten studies were performed at University Hospitals, two of which were tertiary endometriosis centres and one study was performed at a biotechnology firm. Three studies were published in 1996, three studies were published between 2003 and 2009 and the remaining five studies were published between 2012 and 2014. All the included studies evaluated women of reproductive age. There were no randomised controlled trials and all the studies were observational, mainly of cross-sectional design. Ten studies were 'single-gate', where both cases and controls were sampled from the same population and one study was of a 'two-gate design', including a wider group of participants who were undergoing surgery for various indications. Laparoscopy was used for diagnosis in all studies, laparotomy was co-utilised in four studies and seven studies used histopathology to confirm the surgical diagnosis. Seven studies evaluated any pelvic endometriosis, of which one study included only participants with minimal-mild endometriosis (rASRM stage I-II), one study included only participants with moderate-severe endometriosis (rASRM stage III-IV) and one study concentrated on endometriosis with peri-ovarian adhesions. Two other studies addressed only ovarian endometriosis, one study focused on a combination of ovarian endometriosis and deep infiltrating endometriosis (DIE), and one study addressed mapping

DIE at specific anatomical sites. The reported prevalence of endometriosis varied from 29% to 69%. Four studies received financial support, of which one study reported commercial funding and most authors of that publication worked in the biotechnology industry. Four other groups of authors declared no conflict of interest and no information was available from the remaining studies.

Basic features of the excluded studies

The list and descriptions of the excluded studies are presented in Characteristics of excluded studies. Based on a full-text assessment, 21 publications were excluded, of which 15 studies evaluated several testing modalities but did not present diagnostic estimates for the combined test. A further two studies reported statistically significant differences in biomarker levels between the study and control groups, but contained insufficient diagnostic accuracy information for the construction of 2 x 2 contingency tables. One study was of retrospective design where the participants were recruited after the surgical procedure and enrolled postmenopausal women. In one excluded paper the target condition was outside the inclusion criteria and normal versus abnormal pelvis comparison was made without any independent data for endometriosis. One study incorporated imaging evaluation into the combined test but reported only 'lesion-level' analysis and one study was excluded because it was a review article.

Methodological quality of included studies

The quality of the included studies is illustrated in the QUADAS-2 results summary (Figure 3 and Figure 4). Overall, the studies were of poor methodological quality and all studies had an unclear or high risk of bias in at least one domain.

Figure 3. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study

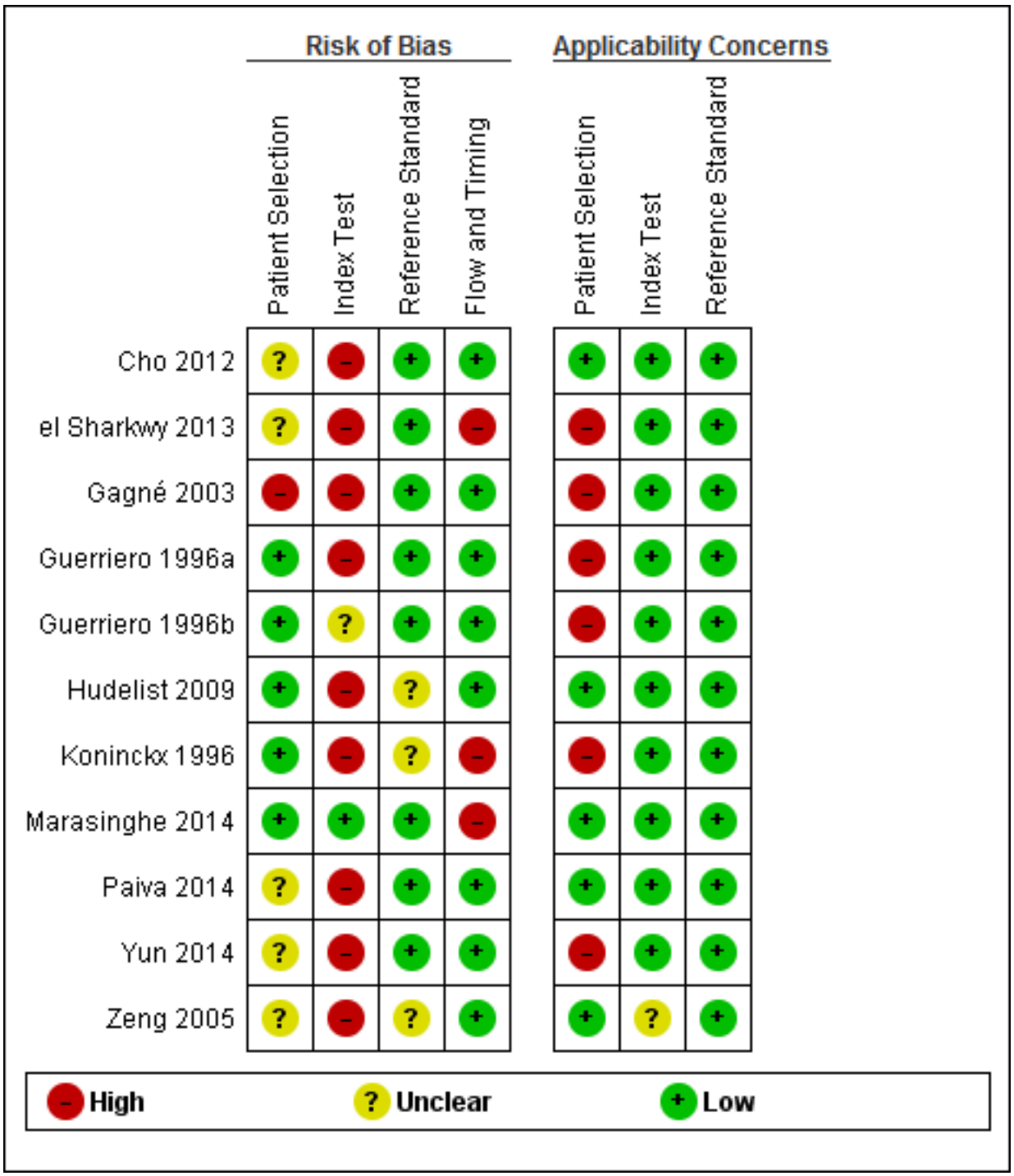
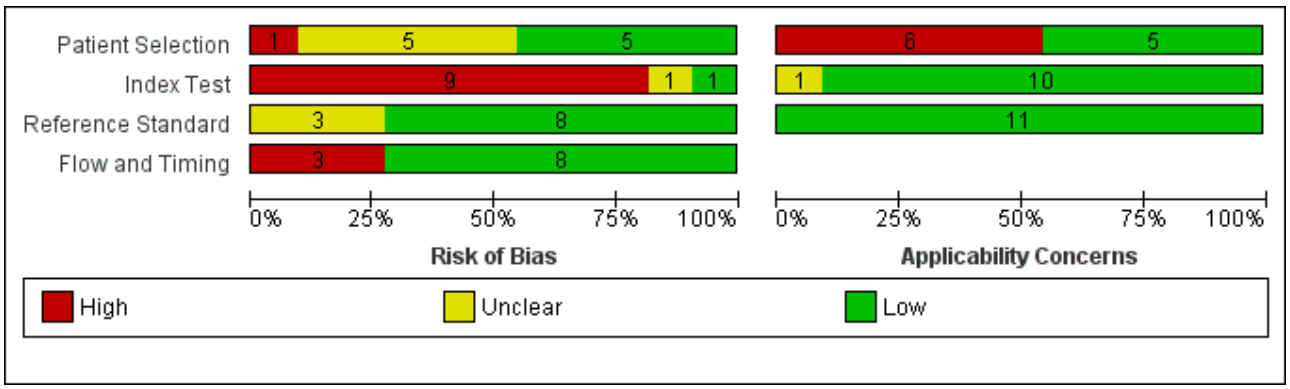


Figure 4. Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies



Five studies presented a low risk of patient selection bias (Guerriero 1996a; Guerriero 1996b; Hudelist 2009; Koninckx 1996; Marasinghe 2014), five studies demonstrated an unclear risk and one study was assessed at high risk for this domain. Non-consecutive or non-random selection of participants, utilisation of a two-gate design for participant selection, the absence of a clear definition of inclusion/exclusion criteria and using a highly selected group of women were the main reasons for a high risk assessment of bias.

One study demonstrated a low risk of index test interpretation bias (Marasinghe 2014), one study demonstrated an unclear risk and nine studies carried a high risk. A lack of clear pre-specified criteria for a positive diagnosis and index test operators not being blind to the results of reference standard were the main reasons for a high risk assessment. The skill level of a test operator and the interobserver variability, both of which directly affect performance of the tests, were rarely reported.

Eight studies were at low risk of bias in the 'reference standard' domain (Cho 2012; el Sharkwy 2013; Gagné 2003; Guerriero 1996a; Guerriero 1996b; Marasinghe 2014; Paiva 2014; Yun 2014), three studies were classified as unclear risk and no studies demonstrated a high risk. An unclear risk of bias was assigned if there was not enough information to determine how likely the reference standard was to have correctly classified the target condition. Specifically, surgical procedures were not well-described, the criteria for a positive reference standard were not stated, it was unclear if histology was utilised to confirm surgical diagnosis, or there was no information regarding the experience of the surgeons or the pathologists involved.

Eight studies presented a low risk of bias in the 'flow and timing' domain (Cho 2012; Gagné 2003; Guerriero 1996a; Guerriero 1996b; Hudelist 2009; Paiva 2014; Yun 2014; Zeng 2005), no studies demonstrated an unclear risk and three studies carried a high risk. In every study all participants received the same reference standard. The time interval between the index test and the reference standard was placed as 12 months or less and the most commonly reported time interval was immediately before surgery. A high risk of bias was assigned if there were unexplained withdrawals that exceeded 5% of the enrolled population or if the reason for withdrawal could introduce selection bias regarding the samples analysed.

Five studies presented a low concern for patient selection applicability, no studies demonstrated an unclear concern and six were of high concern. A high concern in patient selection applicability was assigned if the study utilised two-gate selection for cases and controls or if only a limited spectrum of disease was evaluated. In our view, any sampling deviation from a representative group of the entire clinically relevant population could skew the estimates of diagnostic accuracy in any direction.

In 10 studies there was a low concern of index test applicability, whereas in one study the concern was unclear and none of the studies presented a high concern. An unclear concern was assigned when the study did not present sufficient information regarding the conduct of the tests, such as the laboratory methods or reagents used or the level of expertise of the test operators.

All 11 studies were of low concern for applicability in regards to the reference standard and none of the studies had high or unclear concern. All the included studies implemented pelvic surgery (laparoscopy or laparotomy) as a reference standard, which could be relied upon to match the review question.

Findings

A total of 15 diagnostic combinations of several testing modalities were evaluated in the 11 included studies (Summary of findings 1). Of these, seven were assessed for their value in detecting pelvic endometriosis, two tests were appraised in a context of DIE or endometrioma and 10 tests were evaluated for their accuracy to differentiate endometrioma from other benign ovarian cysts. One additional test looked at specific anatomical sites of DIE and therefore may be considered for a preoperative mapping rather for a primary diagnosis of the disease. The ways the tests were combined in a diagnostic panel varied between the studies and included the following: 1. all the tests of the panel are positive considering a specific cut-off for each constituent; 2. either of the tests included in a diagnostic panel is positive; 3. sum or multiplication of values of all the tests comprising diagnostic panel utilising distinct cut-off value; 4. multivariate logistic regression model.

1. Tests for the diagnosis of any pelvic endometriosis

1) IL-6 (> 15.4 pg/ml) [serum] + PGP 9.5 [endometrium]

The diagnostic performance of serum IL-6 in combination with endometrial PGP 9.5 was evaluated in one study with a total

of 78 women (el Sharkwy 2013). The test was performed in the follicular phase of the menstrual cycle and was evaluated for only minimal-mild endometriosis, rASRM I-II. The definition of positive test was a cut-off value > 15.4 pg/ml for IL-6 and positive immunohistochemistry (IHC) staining for PGP 9.5 in the functional layer of endometrium. When both components of the test were positive, the sensitivity was 1.00 (95% CI 0.91 to 1.00), and the specificity 0.93 (95% CI 0.80 to 0.98) (Figure 5; Figure 6). The point estimates met the criteria for a replacement and SnOUT triage test and approached the criteria for a SpIN triage test. The diagnostic estimates of the combined test were higher than for each individual tests assessed in this study: for serum IL-6 with a cut-off above 15.4 pg/ml the sensitivity and specificity were 0.89 (95% CI 0.75 to 0.97) and 0.82 (95% CI 0.67, 0.93), respectively; for PGP 9.5 the sensitivity and specificity were 0.92 (95% CI 0.79 to 0.98) and 0.80 (95% CI 0.64 to 0.91), respectively. The CIs were broad for each of the included tests, which was particularly prominent for individual tests. Further testing in larger studies including participants with a wider spectrum of endometriosis is needed to confirm the role of the above test in detecting endometriosis.

Figure 5. Forest plot of the combined tests for detection of pelvic endometriosis. Plot shows the estimates of sensitivity and specificity (squares) with 95% CI (black line) specific for each evaluation (each evaluation was derived from a single study), country in which the study was conducted and severity of the disease assessed by each study, reported as rASRM stage. FN: false negative; FP: false positive; TN: true negative; TP: true positive.

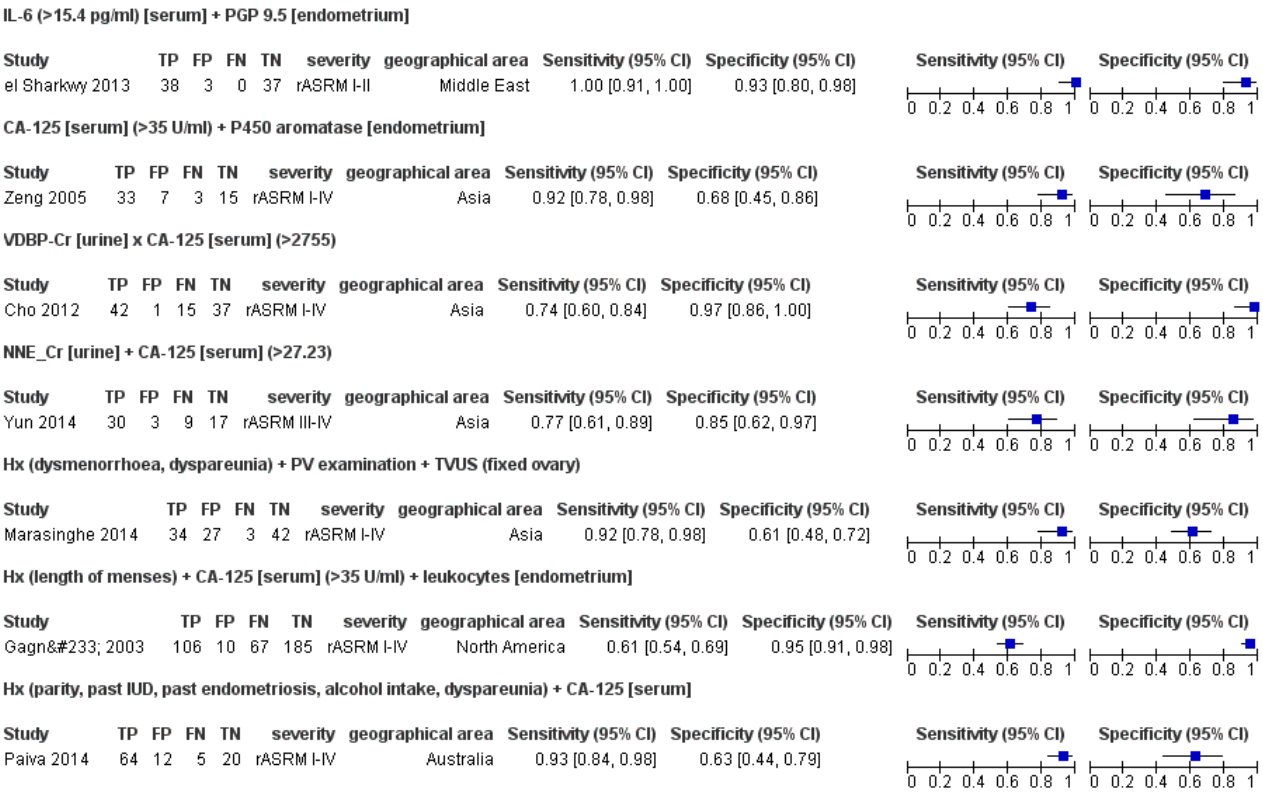
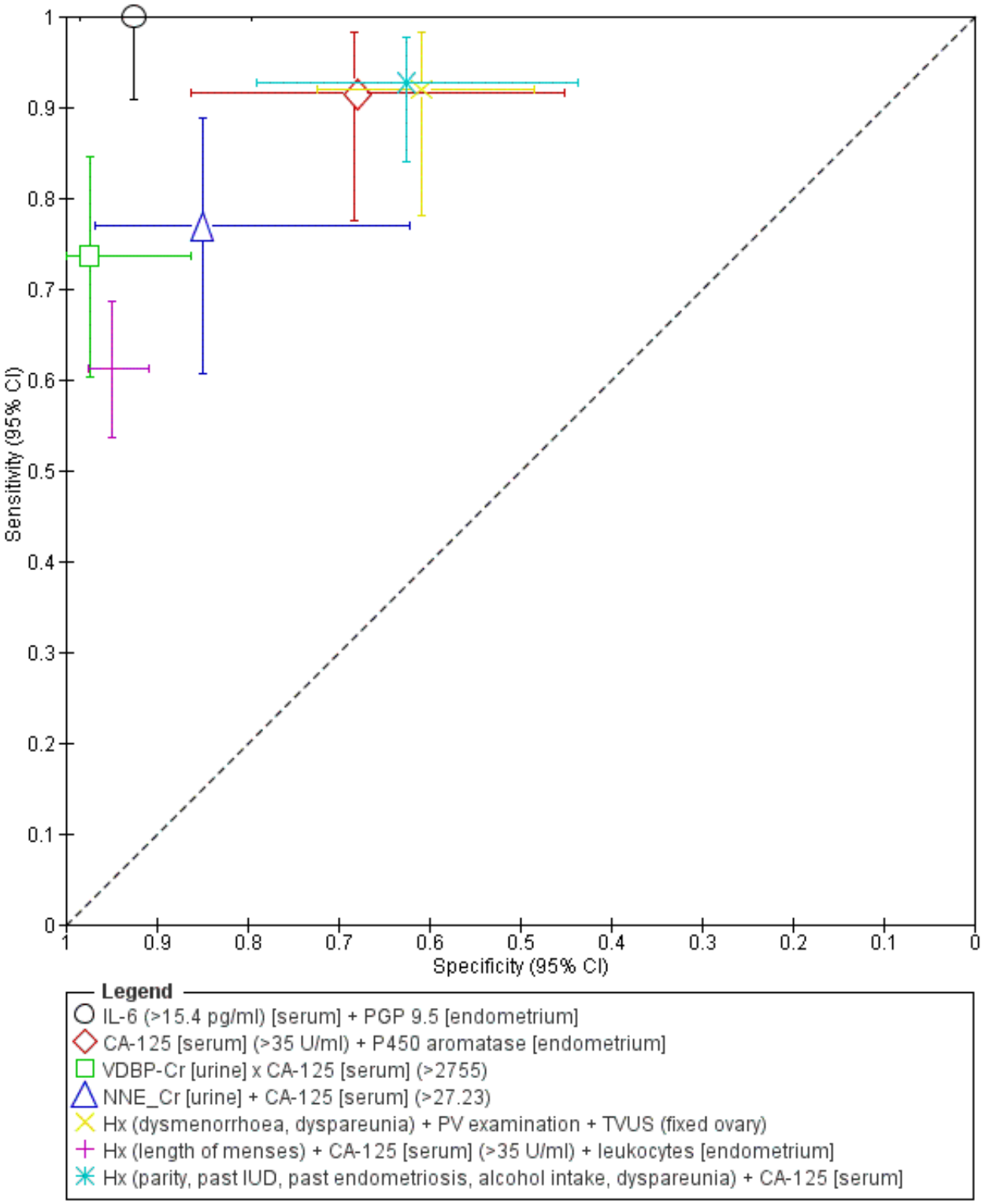


Figure 6. Study specific estimates of the diagnostic accuracy of the combined tests for detection of pelvic endometriosis plotted in ROC space. Each point represents the pair of sensitivity and specificity from each evaluation (each evaluation was derived from a single study). The size of each point is proportional to the sample size the shape designates different tests. The bars correspond to 95% CIs of each individual evaluation.



2) CA-125 [serum] (> 35 U/ml) + aromatase P450 [endometrium].

One study comprising 58 women evaluated the role of serum CA-125 combined with endometrial aromatase P450 in diagnosing endometriosis (Zeng 2005). The test was performed in the follicular or luteal phases of the menstrual cycle, but the effect of the cycle phase on the test performance was not assessed. The study addressed pelvic endometriosis, rASRM I-IV. The test was positive when the CA-125 level was above 35 U/ml and endometrial IHC was positive for aromatase. Considering both positive components, the test had a sensitivity of 0.92 (95% CI 0.78 to 0.98) and a specificity of 0.68 (95% CI 0.45 to 0.86) (Figure 5; Figure 6), approaching the criteria for a SnOUT triage test. Direct comparison between the combination and each individual test assessed in this study revealed that the combined test had higher sensitivity but lower specificity than each individual test: CA-125 sensitivity 0.44 (95% CI 0.28 to 0.62), specificity 0.82 (95% CI 0.60 to 0.95); aromatase P450 sensitivity 0.83 (95% CI 0.67 to 0.94), specificity 0.86 (95% CI 0.65 to 0.97) with wide CIs for each evaluation. This result requires further validation in large well-defined populations, accounting for a menstrual cycle phase of testing.

3) VDBP-Cr [urine] x CA-125 [serum] (> 2755)

The diagnostic performance of the combination of urinary VDBP (vitamin-D-binding protein) and serum CA-125 was evaluated in one study, which included 95 women (Cho 2012). The test was performed in the follicular or luteal cycle phase. Even though the study included endometriosis of varying severity (rASRM I-IV), more than 90% of women with endometriosis had moderate - severe disease (52/57). Urinary VDBP levels were significantly higher in endometriosis only in luteal cycle phase, however the performance of the combined test was not stratified by a cycle phase. The test was considered positive when a multiplication of urinary VDBP level corrected for creatinine (VDBP-Cr) by serum CA-125 level was above 2755. The test demonstrated a sensitivity of 0.74 (95% CI 0.60 to 0.84) and a specificity of 0.97 (95% CI 0.86 to 1.00) (Figure 5; Figure 6) and met the criteria for a SpIN triage test. In direct comparison this combination had higher diagnostic estimates than VDBP only (sensitivity 0.58 (95% CI 0.44 to 0.71), specificity 0.55 (95% CI 0.38 to 0.71). Both individual urinary VDBP and combined VDBP + CA-125 test manifested wide CIs; separate data for CA-125 only were not available from this study. Further evaluation of VDBP - CA-125 combination across the spectrum of endometriosis particularly in the luteal phase may help to clarify the diagnostic role of this tests in endometriosis.

4) NNE-Cr [urine] + CA-125 [serum] (> 27.23)

The role of combining urinary NNE (enolase I) and serum CA-125 was assessed in one study with a total of 59 participants (Yun 2014). The test was performed in the follicular or luteal cycle phase and assessed only moderate - severe endometriosis, rASRM III-IV. Urinary NNE expression was not influenced by cycle phase and was significantly greater (P = 0.026) in women with endometriosis only after correction for creatine excretion. A positive test was defined as an arithmetical sum of the urinary NNE-Cr and serum CA-125 level above 27.23. This test exhibited a sensitivity of 0.77 (95% CI 0.61 to 0.89) and a specificity 0.85 (95% CI 0.62 to 0.97) (Figure 5; Figure 6). Although the diagnostic estimates of the combination were superior to those of NNE only (sensitivity 0.56 (95% CI 0.40 to 0.72), specificity 0.70 (95% CI 0.46 to 0.88); the diagnostic estimated for CA-125 only were not available), the criteria for either replacement or triage test were not met. Considering a single study, there is not enough information on diagnostic utility of NNE + CA-125 combination in detecting endometriosis.

5) History (dysmenorrhoea and dyspareunia) + PV examination + TVUS

One study comprising 106 participants evaluated the combination of history, gynaecological examination and transvaginal ultrasound (TVUS) for detecting pelvic endometriosis, rASRM I-IV (Marasinghe 2014). The authors did not specify the menstrual cycle phase of the testing. The test was considered as positive when: 1. the clinical history was positive for dysmenorrhoea and dyspareunia. Pain severity was assessed using a visual analogue scale ranging one to 10 with a score of one considered as 'no pain'; 2. bimanual pelvic vaginal examination (PV) was used to detect the presence of pelvic tenderness, a fixed retroverted uterus, tender uterosacral ligaments and deeply infiltrating nodules on the uterosacral ligaments or in the cul-de-sac; 3. TVUS demonstrated 'fixed ovaries', defined when the ovaries did not move freely over the ipsilateral internal iliac vessels or pelvic sidewall or uterus with the gentle pressure; any of these findings resulted in a 'positive' test result. Sonographic criterion suggested that the authors focused on endometriosis with per-ovarian adhesions. The combination of positive history, examination findings and TVUS findings had a sensitivity of 0.92 (95% CI 0.78 to 0.98) and a specificity of 0.61 (95% CI 0.48 to 0.72) (Figure 5; Figure 6). The diagnostic estimates approached the criteria for a SnOUT triage test, although contained wide CIs. The reported sensitivity and specificity for each component of the test in this study were 0.46 and 0.77 for dyspareunia, 0.76 and 0.70 for dysmenorrhoea, 0.73 and 0.88 for positive vaginal examination, and 0.78 and 0.94 for fixed ovaries on TVUS.

6) History (length of menses) + CA-125 [serum] (> 35 U/ml) + leukocytes [endometrium]

One study with a total of 368 participants assessed the performance of history, serum CA-125 and endometrial leukocyte subsets in diagnosing endometriosis (Gagné 2003). The test was performed in the luteal cycle phase and was utilised for detecting a wide spectrum of pelvic endometriosis, rASRM I-IV. The diagnostic test for endometriosis included the following: 1. clinical history (length of menses) 2. serum CA-125 with a cut-off value above 12.8 U/ml; 3. endometrial leukocytes (CD3+, CD16+, CD3-HLADR-, CD3-CD45RA-, CD3+CD16-, CD3+CD56-, CD56-CD16+, CD16b+) with specific cut-off for each leukocyte subset. The test parameters were selected by univariate analysis and then included in the predictive model by utilising a multiple logistic regression with subsequent bootstrap method validation. The model adjusted for gravidity and histologic dating (early, mid or late luteal phase) demonstrated a sensitivity of 0.61 (95% CI 0.54 to 0.69) and a specificity of 0.95 (95% CI 0.91 to 0.98) (Figure 5; Figure 6). In this study, the combined test performed better than CA-125 only (sensitivity 0.20 (95% CI 0.15 to 0.27), specificity 0.92 (95% CI 0.87 to 0.95)) and met the criteria for a SpIN triage test. The diagnostic estimates for endometrial leukocytes only were not available.

7) History (parity, past use of IUD, past endometriosis, alcohol intake, dyspareunia) + CA-125 [serum]

The combination of the clinical and demographical parameters with serum CA-125 for discriminating women with and without endometriosis was evaluated in one study comprising 101 participants (Paiva 2014). The test was performed at different

phases of menstrual cycle (menstrual, follicular or luteal) and assessed the full spectrum of endometriosis, rASRM I-IV. The level of CA-125 did not vary across the menstrual cycle. Fourteen parameters identified by univariate analysis were used with logistic regression to produce the diagnostic model, which included 1. clinical data: parity, ever had an intrauterine device (IUD), history of endometriosis, alcohol intake, dyspareunia); 2. serum CA-125, cut-off value not specified. The diagnostic model demonstrated a sensitivity of 0.93 (95% CI 0.84 to 0.98) and a specificity of 0.63 (95% CI 0.44 to 0.79) (Figure 5; Figure 6). The test approached the criteria for a SnOUT triage test, although exhibited wide CIs for both sensitivity and specificity. The information on diagnostic performance of individual components of the test was not available.

2. Tests for diagnosis of DIE or ovarian endometriosis

1) PV examination (menstrual nodularities) + CA-125 [serum] (> 35 IU/L)

2) PV examination (menstrual nodularities) OR CA-125 [serum] (> 35 IU/L)

One study comprising 55 participants, evaluated the role of gynaecological examination in adjunct with serum CA-125 for detecting one of the following target conditions: 1. DIE or ovarian endometrioma or severe pelvic adhesions; 2. only DIE; 3. only ovarian endometrioma (Koninckx 1996). The test included 1. bimanual PV examination during menstruation, which was scored as positive when an induration or painful nodularities was felt; 2. serum CA-125 measured in mid-follicular cycle phase with a cut-off value above 35 IU/L. Two variations of the test were assessed for each target condition: 1. both components of the test were positive or 2. either of the two tests was positive.

For detecting DIE, endometrioma or severe adhesions the test exhibited a sensitivity of 0.42 (95% CI 0.22 to 0.63) with a specificity of 1.00 (95% CI 0.80 to 1.00) when both examination and serum CA-125 were positive. The test achieved a sensitivity of 0.88 (95% CI 0.68 to 0.97) with a specificity of 0.82 (95% CI 0.57 to 0.96) when either component of the combined test was positive (Figure 7; Figure 8). For DIE only, the test demonstrated a sensitivity of 0.38 (95% CI 0.14 to 0.68) with a specificity of 0.88 (95% CI 0.64 to 0.99) when both examination and serum CA-125 were positive, and a sensitivity of 0.85 (95% CI 0.55 to 0.98) with a specificity of 0.71 (95% CI 0.44 to 0.90) when either component of the test was positive (Figure 7; Figure 8). For endometrioma, the test had a sensitivity of 0.56 (95% CI 0.21 to 0.86) with a specificity of 0.88 (95% CI 0.64 to 0.99) when both examination and serum CA-125 were positive, and a sensitivity of 0.89 (95% CI 0.51 to 1.00) with a specificity of 0.65 (95% CI 0.38 to 0.86) when either component of the composite test was considered (Figure 7; Figure 8). None of these tests met the criteria for either a replacement or of the triage tests and all evaluations were featured by wide CIs. The reported diagnostic parameters for CA-125 on its own were sensitivity 0.5 with specificity 0.88 for DIE, endometrioma or severe adhesions; sensitivity 0.47 with specificity 0.81 for DIE and sensitivity 0.67 with specificity 0.81 for endometrioma.

Figure 7. Forest plot of the combined tests for detection of DIE or ovarian endometriosis. Plot shows the estimates of sensitivity and specificity (squares) with 95% CI (black line) specific for each evaluation (each evaluation was derived from a single study), country in which the study was conducted and type of target condition assessed by each study. FN: false negative; FP: false positive; TN: true negative; TP: true positive.

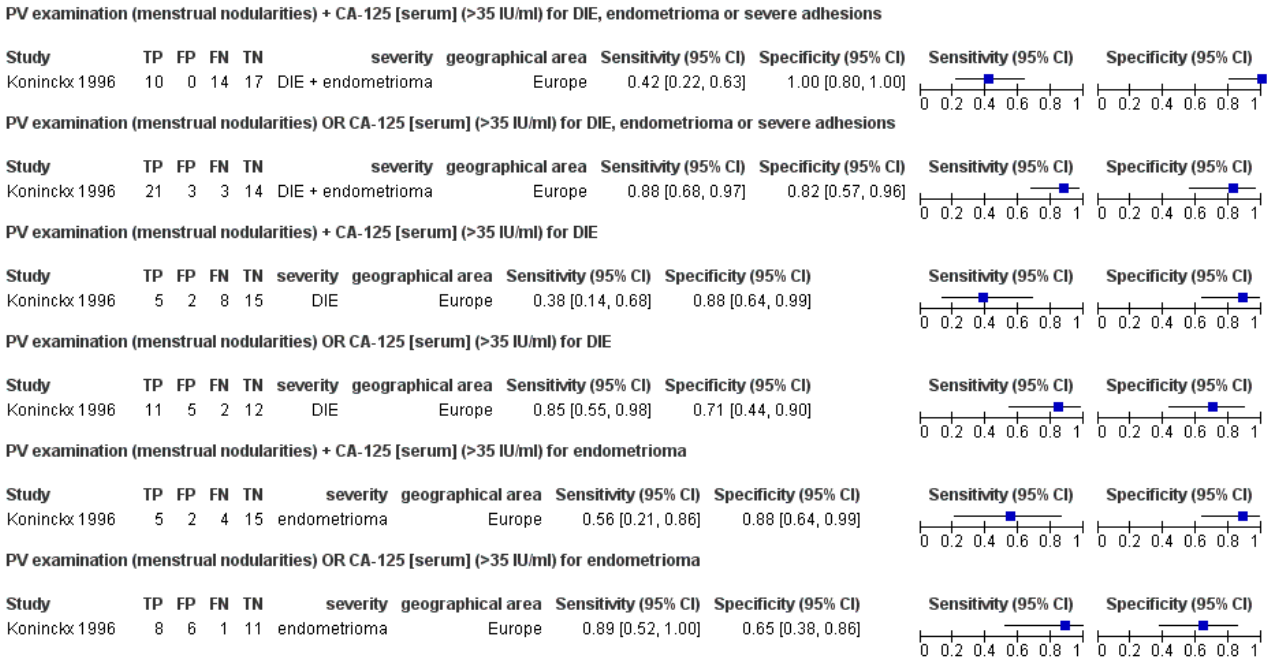
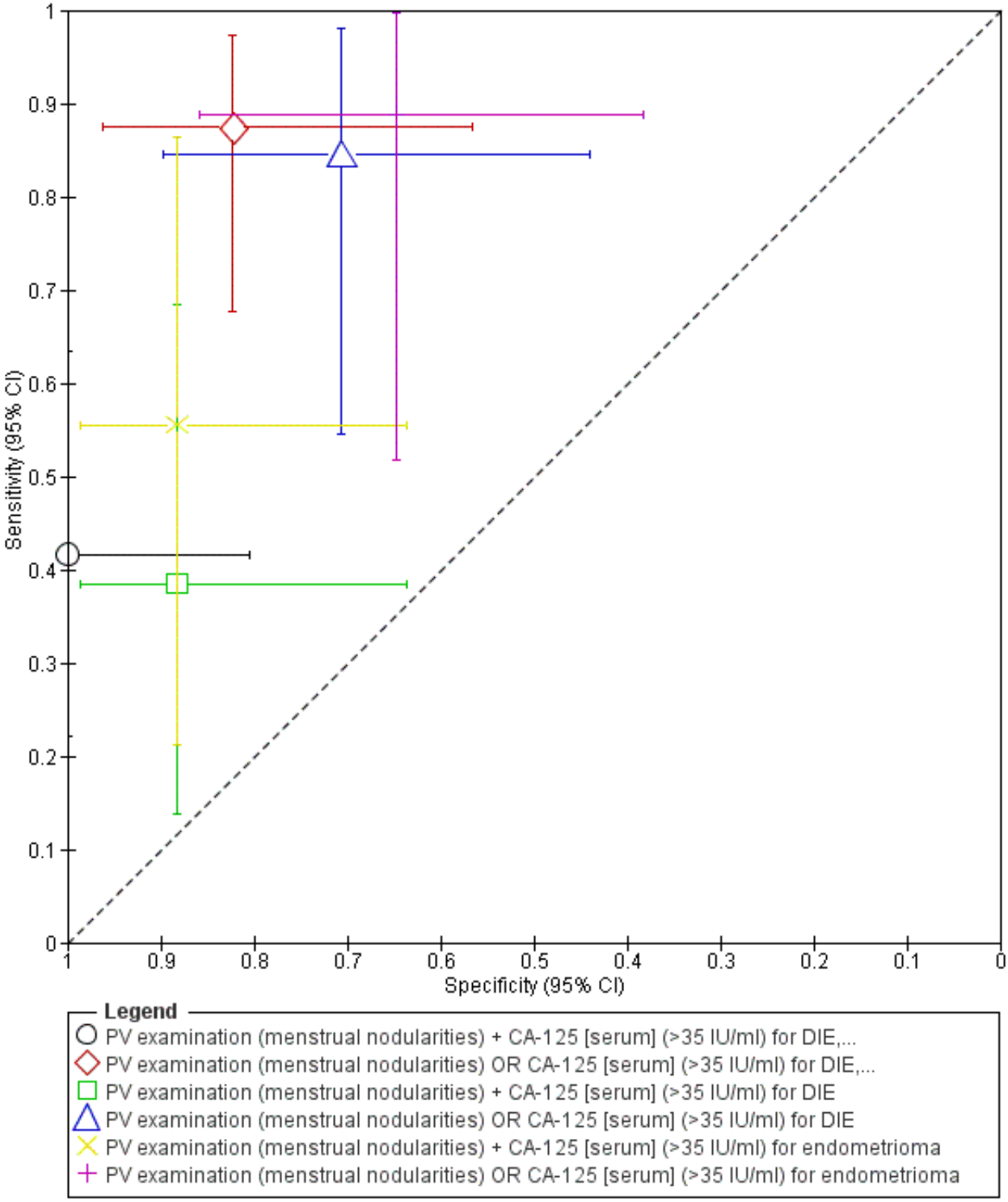


Figure 8. Study specific estimates of the diagnostic accuracy of the combined tests for detection of DIE or ovarian endometriosis plotted in ROC space. Each point represents the pair of sensitivity and specificity from each evaluation (each evaluation was derived from a single study). The size of each point is proportional to the sample size the shape designates different tests. The bars correspond to 95% CIs of each individual evaluation.



3. Tests for differentiating ovarian endometriosis versus other benign ovarian cysts in women of reproductive age

1) TVUS + CA-125 [serum] (≥ 25 U/ml) + CA-19.9 [serum] (≥12 U/ml)

2) TVUS + (CA-125 [serum] (≥ 25 U/ml) OR CA-19.9 [serum] (≥12 U/ml))

3) TVUS + CA-19.9 [serum] (≥ 12 U/ml)

4) TVUS OR CA-19.9 [serum] (≥ 12 U/ml)

One study comprising 118 participants evaluated a composite test of TVUS and serum tumour markers CA-125 and/ or CA-19.9 for discriminating ovarian endometrioma from other benign cysts in women of reproductive age (Guerriero 1996a). All the participants were tested in the follicular phase of the menstrual cycle. Positive TVUS test (presence of endometrioma) was described as a presence of a round shaped homogeneous hypoechoic 'tissue' within the ovary with clear demarcation from the parenchyma and without papillary proliferations. A cut-off for a positive serum biomarker was above 25 U/ml for CA-125 and above 12 U/ml for CA-19.9.

The test demonstrated a sensitivity of 0.49 (95% CI 0.32 to 0.65) and a specificity of 0.99 (95% CI 0.93 to 1.00) when all three components of the test were positive (Figure 9; Figure 10), approaching the criteria for a SpIN triage test. The test had a higher sensitivity (0.79 (95% CI 0.64 to 0.91)) and slightly lower specificity (0.97 (95% CI 0.91 to 1.00)) when either positive blood test was considered in adjunct with TVUS (Figure 9; Figure 10), meeting the criteria for a SpIN triage test.

Figure 9. Forest plot of the combined tests (TVUS and/or CA-125 and/or CA-19.9) for differentiation of ovarian endometriosis vs. other benign ovarian cysts. Plot shows the estimates of sensitivity and specificity (squares) with 95% CI (black line) specific for each evaluation (each evaluation was derived from a single study Guerriero 1996a), country in which the study was conducted and target condition assessed by each study. FN: false negative; FP: false positive; TN: true negative; TP: true positive.

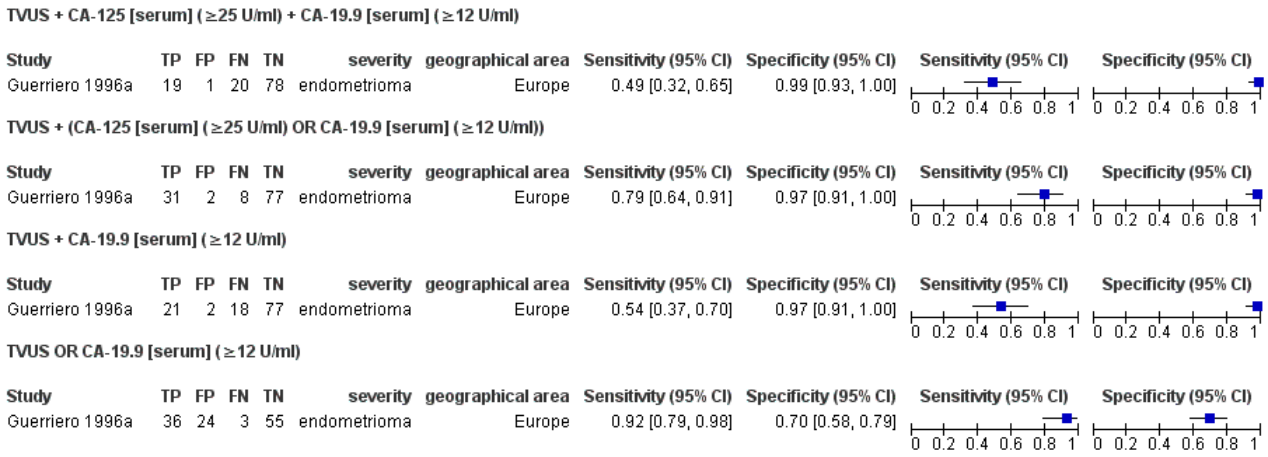
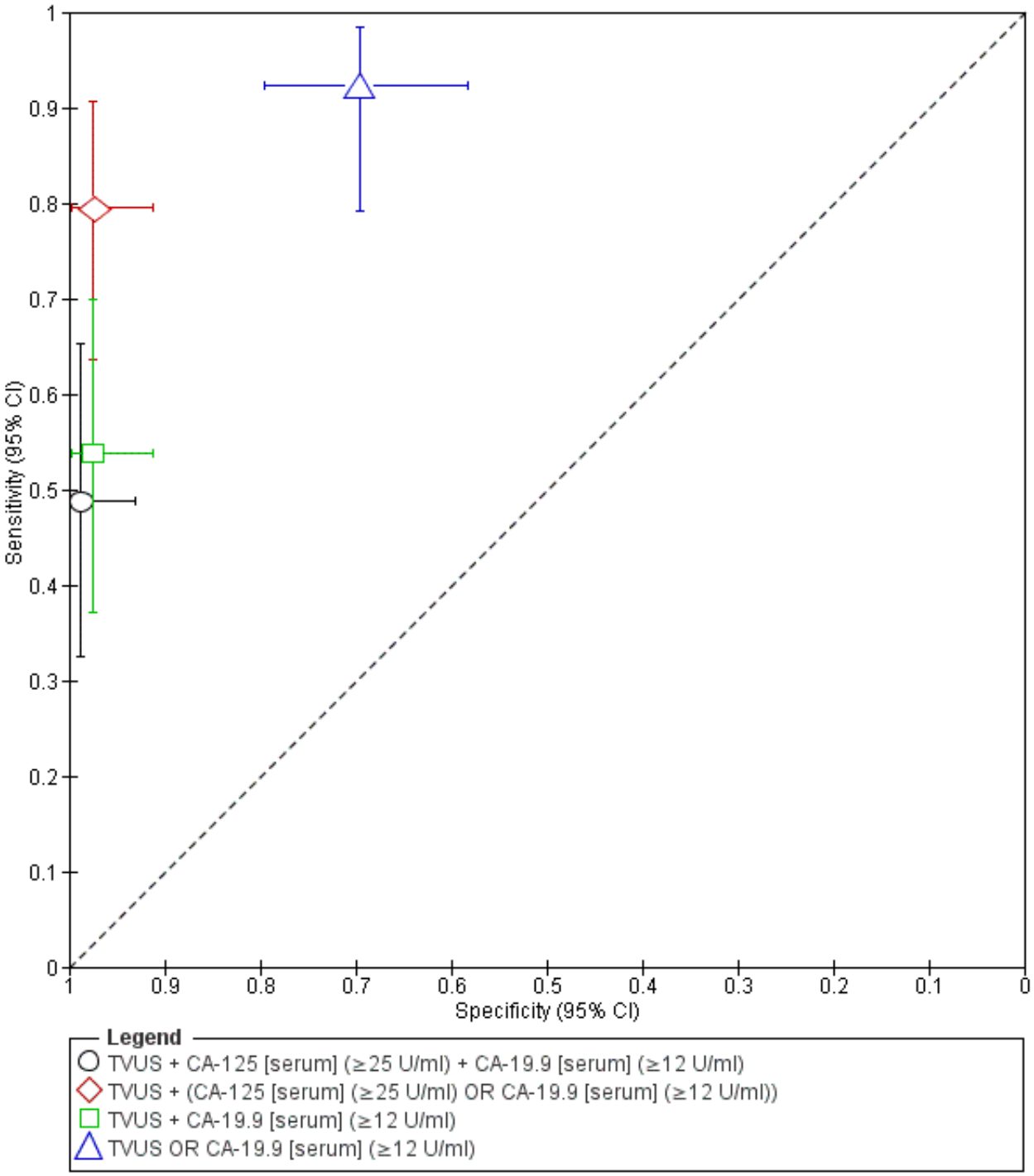


Figure 10. Study specific estimates of the diagnostic accuracy of the combined tests (TVUS and/or CA-125 and/or CA-19.9) for differentiation of ovarian endometriosis vs. other benign ovarian cysts plotted in ROC space. Each point represents the pair of sensitivity and specificity from each evaluation (each evaluation was derived from a single study Guerriero 1996a). The size of each point is proportional to the sample size the shape designates different tests. The bars correspond to 95% CIs of each individual evaluation.



When only TVUS and CA-19.9 were considered, a sensitivity was 0.54 (95% CI 0.37 to 0.70) and a specificity was 0.97 (95% CI 0.91 to 1.00) for both positive components (Figure 9; Figure 10), which could qualify as a SpIN triage test. When either TVUS or CA-19.9 was positive, the test demonstrated a sensitivity of 0.92 (95% CI 0.79 to 0.98) and a specificity of 0.70 (95% CI 0.58 to 0.78) (Figure 9; Figure

10), approaching the criteria for a SnOUT triage test. Considering the data reported by a single study and wide overlapping CIs, we suggest caution in interpretation of the presented findings.

In a head-to-head direct comparison, the test based on a combination of ether positive blood biomarker with TVUS and a combination of CA-19.9 with TVUS performed better than the test including both positive CA-19.9 and CA-125. The study-specific diagnostic estimates for TVUS only were sensitivity 0.85 (95% CI 0.69 to 0.94) and specificity 0.97 (95% CI 0.91 to 1.00). This suggested that addition of biomarkers to ultrasound examination did not improve diagnostic performance of the test in this study, resulting in lower sensitivity and only marginally higher or similar specificity for most combinations. This was particularly noticeable for the combinations with CA-125. Further, no blood biomarker combinations from this study without TVUS met the criteria of either replacement or triage test (Nisenblat 2016a).

5) TVUS + CA-125 [serum] (≥ 20 U/ml)

6) TVUS OR CA-125 [serum] (≥ 20 U/ml)

7) TVUS + CA-125 [serum] (≥ 25 U/ml)

8) TVUS OR CA-125 [serum] (≥ 25 U/ml)

9) TVUS + CA-125 [serum] (≥ 35 U/ml)

10) TVUS OR CA-125 [serum] (≥ 35 U/ml)

One study with a total of 101 women evaluated the combination of TVUS and serum CA-125 in diagnosing ovarian endometrioma when compared with other benign ovarian cysts (Guerriero 1996b). The study was performed by the same group that evaluated the above mentioned combined testing for endometrioma (TVUS or CA-125 or CA-19.9), utilising similar sonographic criteria and similar testing time (follicular cycle phase). Two variations of the test included 1. both TVUS and CA-125 positive and 2. either test is positive.Three different cut-off thresholds for CA-125 were used for each pair of the combined test (≥ 20 U/ml; ≥ 25 U/ml; ≥ 35 U/ml).

For TVUS and CA-125 with a cut-off value ≥ 20 U/ml, the diagnostic estimates met the criteria for a SpIN triage test, when both tests were positive (sensitivity 0.69 (95% CI 0.49 to 0.85), specificity of 0.96 (95% CI 0.88 to 0.99)) and approached the criteria for a SnOUT triage test when either positive test was considered (sensitivity 0.93 (95% CI 0.77 to 0.99), specificity of 0.53 (95% CI 0.41 to 0.65)) (Figure 11; Figure 12).

Figure 11. Forest plot of the combined tests (TVUS and/or CA-125 at varying thresholds) for differentiation of ovarian endometriosis vs. other benign ovarian cysts. Plot shows the estimates of sensitivity and specificity (squares) with 95% CI (black line) specific for each evaluation (each evaluation was derived from a single study Guerriero 1996b), country in which the study was conducted and target condition assessed by each study. FN: false negative; FP: false positive; TN: true negative; TP: true positive.

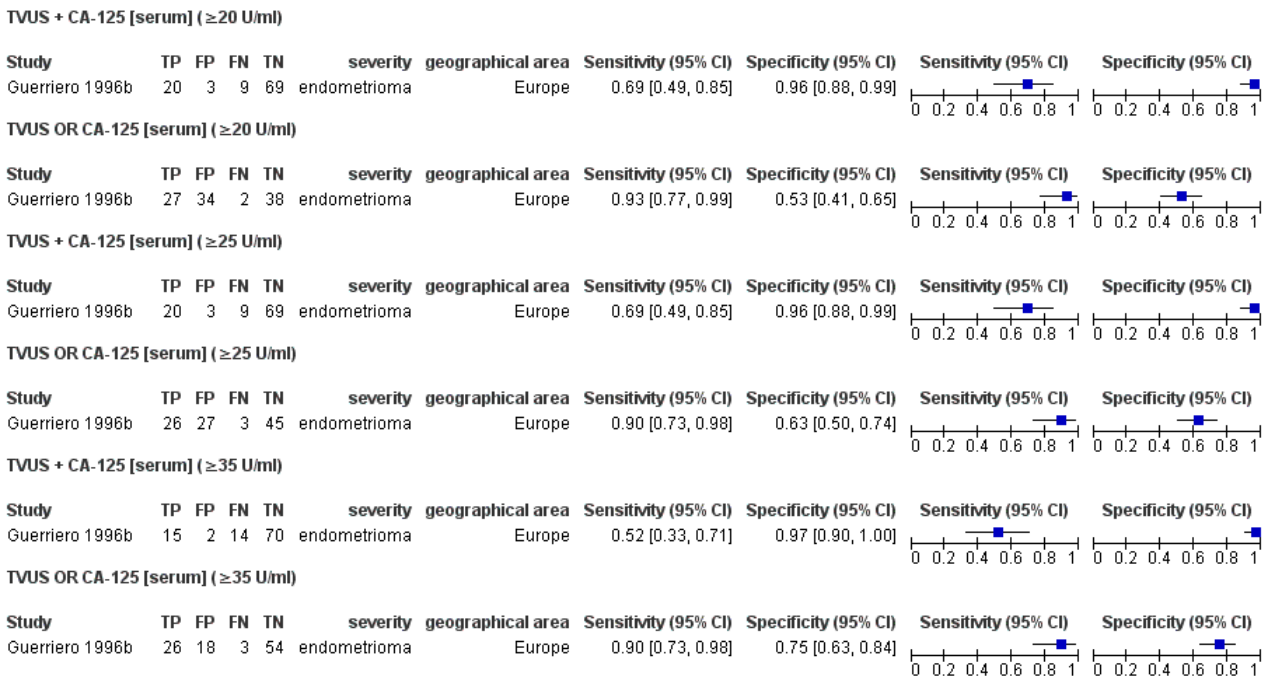
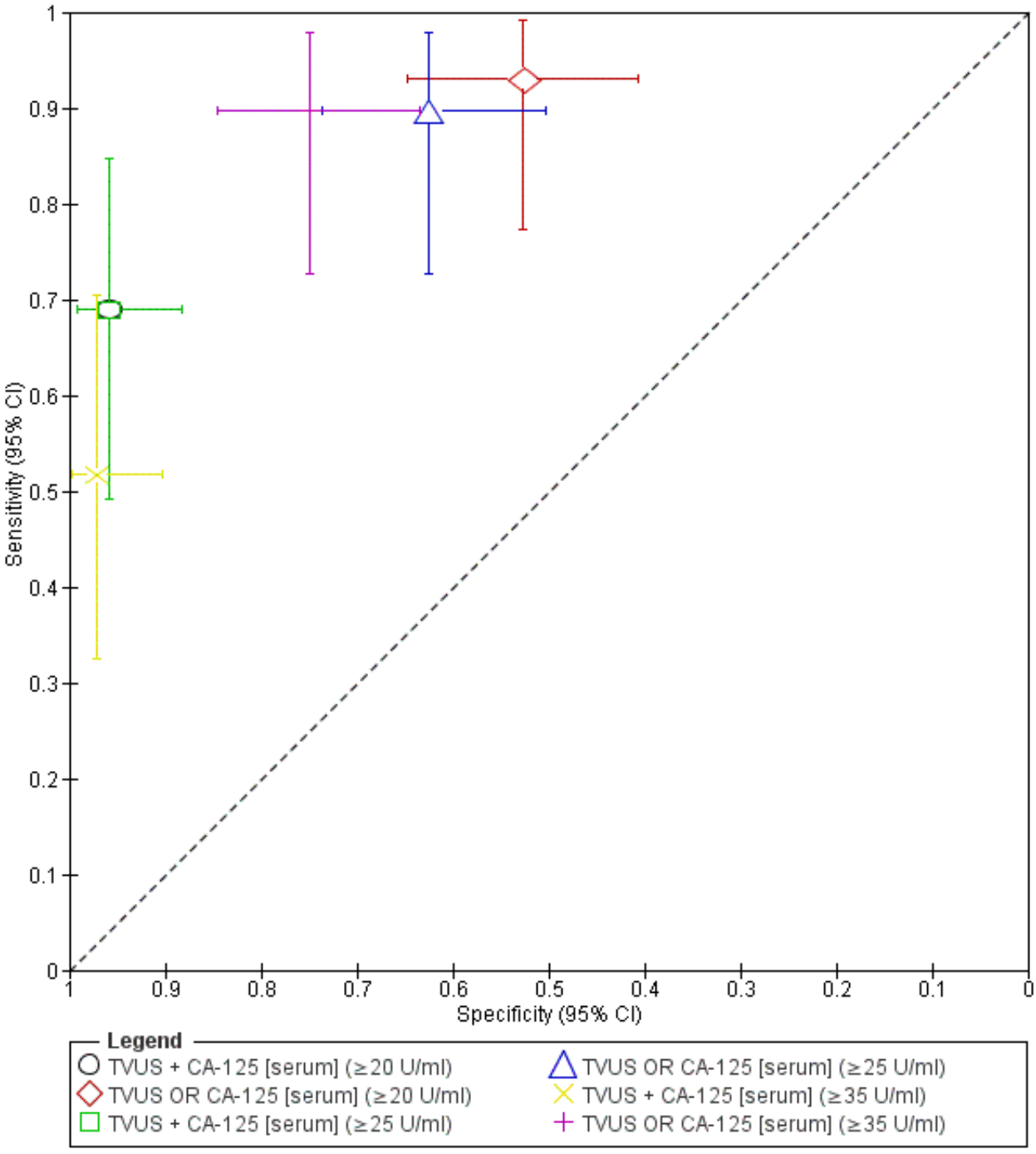


Figure 12. Study specific estimates of the diagnostic accuracy of the combined tests (TVUS and/or CA-125 at varying thresholds) for differentiation of ovarian endometriosis vs. other benign ovarian cysts plotted in ROC space. Each point represents the pair of sensitivity and specificity from each evaluation (each evaluation was derived from a single study Guerriero 1996b). The size of each point is proportional to the sample size the shape designates different tests. The bars correspond to 95% CIs of each individual evaluation.



For TVUS and CA-125 with a cut-off value ≥ 25 U/ml, the diagnostic estimates met the criteria for a SpIN triage test when both tests were positive (sensitivity 0.69 (95% CI 0.49 to 0.85), specificity of 0.96 (95% CI 0.88 to 0.99)) and approached the criteria for a SnOUT triage test when either positive test was considered (sensitivity 0.90

(95% CI 0.73 to 0.98), specificity of 0.63 (95% CI 0.50 to 0.74)) (Figure 11; Figure 12).

For CA-125 with a cut-off value ≥ 35 U/ml, the diagnostic estimates met the criteria for a SpIN triage test when both tests were positive (sensitivity 0.52 (95% CI 0.33 to 0.71), specificity of 0.97 (95% CI 0.90 to1.00)) and approached the criteria for a replacement and SnOUT triage test when either positive test was considered (sensitivity 0.90 (95% CI 0.73 to 0.98), specificity of 0.75 (95% CI 0.63 to 0.84)) (Figure 11; Figure 12).

In direct comparison, TVUS and CA-125 at a cut-off value ≥ 25 U/ml and ≥ 35 U/ml performed better than at a cut-off value ≥20 U/ml only when either positive test was considered, but did not improve diagnostic estimates of the combination when both tests were positive. In this study TVUS alone (sensitivity 0.83 (95% CI 0.64 to 0.94), specificity 0.93 (95% CI 0.85 to 0.98)), was more sensitive than the combination TVUS + CA-125 and more specific than the combination TVUS OR CA-125. None of the blood biomarkers without TVUS could qualify as useful diagnostic test (Nisenblat 2016a).

4. Tests for mapping of DIE at specific anatomical locations

1) PV examination + TVUS

One study comprising 200 participants evaluated the combination of gynaecological examination and TVUS for detecting DIE at specific anatomical locations: 1. pouch of Douglas (POD) obliteration; 2. vaginal wall; 3. rectovaginal septum (RVS); 4. rectum. The cycle phase of the testing was not reported (Hudelist 2009). The gynaecological bimanual examination was considered positive when nodularity or stiffened or thickened area or a palpable cystic expansion were detected at the evaluated anatomical sites. TVUS criteria were defined for each anatomical location, specifically 1. uterus, adnexa and rectosigmoid colon fixed to each other with disappearance of the peritoneal structure (complete POD obliteration); peritoneal limits partially identified with the presence or absence of suspended or lateralised fluid collection (incomplete POD obliteration); 2. thickening or the presence of a hypoechogenic cystic or non-cystic nodularity within the posterior vaginal wall (vaginal DIE); 3. presence of a hypoechogenic nodularity or cystic mass within RVS - area between rectum and posterior vaginal wall from the level of introitus up to a level defined by the lower border of posterior lip of cervix (RVS DIE); 4. presence of a regular or irregular hypoechogenic mass distorting and replacing the normal appearance of the muscular layer of the rectal wall (rectal DIE). Considering that only specific sites of endometriosis were assessed, these tests can not be considered diagnostic but can be utilised for preoperative mapping of the disease and more careful planning of endometriosis surgery. Therefore, the tests for preoperative mapping of the disease were only evaluated as triage tests to inform decisions to undertake surgery for endometriosis.

The combination of PV examination and TVUS demonstrated the following diagnostic estimates:

1. for POD obliteration: a sensitivity of 0.87 (95% CI 0.69 to 0.96) and a specificity of 0.98 (95% CI 0.95 to 1.00), meeting the criteria for a SpIN triage test;
2. for vaginal wall DIE: a sensitivity of 0.82 (95% CI 0.60 to 0.95) and a specificity of 0.99 (95% CI 0.97 to 1.00), meeting the criteria for a SpIN triage test;
3. for RVS endometriosis: a sensitivity of 0.88 (95% CI 0.47 to 1.00) and a specificity of 0.99 (95% CI 0.96 to 1.00), meeting the criteria for a SpIN triage test; sensitivity demonstrated wide CIs;
4. for rectal endometriosis: a sensitivity of 0.96 (95% CI 0.86 to 0.99) and a specificity of 0.98 (95% CI 0.94 to 1.00), meeting the criteria for a replacement and both a SnOUT and SpIN triage tests.

Separate diagnostic estimates for TVUS only were not reported in this study.

Investigations of heterogeneity and sensitivity analyses

The potential sources of heterogeneity are outlined in Secondary objectives. There was only one study evaluating each test, therefore investigations of heterogeneity and sensitivity analyses were not possible in this review.

DISCUSSION

Summary of main results

Summary of main results presented in this review

Fifteen combinations of several non-invasive methods for the diagnosis of endometriosis were evaluated in 11 included studies published between 1996 and 2014 and comprising 1339 participants. The composite tests have been assessed in small individual studies, providing insufficient data to perform a meta-analysis. None of the included studies were of high methodological quality. There were too few studies to perform a meaningful evaluation for any of the combination tests. Although some tests were sensitive and specific enough to qualify as a replacement or triage test for detecting endometriosis, each was explored in only one study and warrant further validation.

Combinations of several testing methods that met the criteria for a replacement test.

1. IL-6 >15.4 pg/ml [serum] + PGP 9.5 [endometrium] - for pelvic endometriosis
2. PV examination + TVUS - for rectal endometriosis

Combinations of several testing methods that met the criteria for a SpIN triage test.

1. VDBP-Cr [urine] x CA-125 [serum] >2755 - for pelvic endometriosis
2. History (length of menses) + CA-125 [serum] >35 U/ml + leukocytes [endometrium] - for pelvic endometriosis
3. TVUS + (CA-125 [serum] ≥25 U/ml OR CA-19.9 [serum] ≥12 U/ml) - for ovarian endometrioma
4. TVUS + CA-19.9 [serum] ≥12 U/ml - for ovarian endometrioma
5. TVUS + CA-125 [serum] ≥20 U/ml - for ovarian endometrioma
6. TVUS + CA-125 [serum] ≥25 U/ml - for ovarian endometrioma
7. TVUS + CA-125 [serum] ≥35 U/ml - for ovarian endometrioma
8. PV examination + TVUS - for the following anatomic locations:
   a. POD obliteration;
   b. vaginal wall;
   c. RVS.

In all the included studies, combinations of the biomarkers had higher diagnostic estimates than those reported for each individual component of the combined test. However, addition of CA-125

had only a small contribution to the diagnostic performance of the test. We also observed that combinations of biomarkers with TVUS for detecting ovarian endometrioma had lower sensitivity and largely comparable specificity than that presented for TVUS alone in the review on imaging tests (Nisenblat 2016b). Considering the results of meta-analysis from the imaging tests review, addition of vaginal examination to TVUS improved diagnostic performance of ultrasound in detecting vaginal and rectal endometriosis, but did not seem to be superior to TVUS alone for detecting of POD obliteration and RVS endometriosis.

The findings of this 'combination of the tests' review need to be interpreted with caution. Considering both the level of heterogeneity and the high/unclear risk of bias of included studies, the results do not appear sufficiently reliable to inform clinical practice.

Strengths and weaknesses of the review

This review is part of a comprehensive review series of minimally invasive biomarkers for the diagnosis of endometriosis, including 1) imaging tests (Nisenblat 2016b), 2) urinary biomarkers (Liu 2015), 3) blood biomarkers (Nisenblat 2016a), 4) endometrial biomarkers (Gupta 2016), and 5) combination of several testing modalities, presented in this review. The main strength of the review series is its attempt to systematically review the vast number of widely heterogeneous studies in the literature while applying similar methods of study selection, data extraction and quality appraisal. It therefore should allow the most accurate picture of diagnostic test accuracy of non-invasive tests for endometriosis. The review series provide systematisation of the current evidence on all the non-invasive tests for endometriosis, identifying pitfalls in all the imaging and biomarker research areas and provides practical suggestions on further directions for high-quality diagnostic research in the field of endometriosis.

The following are the main strong points of this review.

1. A very thorough search of the current literature was undertaken and included studies written in languages other than English.
2. Data extraction by three independent reviewers and use of a modified QUADAS-2 tool to perform quality assessments.
3. Stringent selection criteria ensured that eligible studies utilised prospectively collected samples for the biomarker-based tests and prospectively enrolled and tested women for the imaging-based tests.
4. Inclusion of only clinically relevant population limited to women of reproductive age, which minimised the risk of bias in interpreting the reference standard and index test.
5. The authors of the studies were approached in attempt to obtain any missing information required to assess eligibility and critically appraise the studies.
6. The combinations including examination and imaging tests also provided information on detecting of specific anatomical sites of DIE, which aids in preoperative mapping of the disease.
7. In this review only one study (9%) was of a 'two-gate design' (a poorer quality design feature than ‘single-gate design’). Therefore, the majority of the included studies comprised a clinically relevant population that would have undergone tests in practice and were at low risk of misinterpretation of the test results secondary to bias in selecting an adequate control group.

These strengths permit the conclusion that this is the most robust review on the topic currently available to inform improvements in the care of women being considered for diagnosis of endometriosis.

The main limitation of the review is that there was a single study for each evaluated index test and no meta-analysis was possible. The studies varied with respect to the included populations, severity of endometriosis, menstrual cycle phase at testing, or radiological protocols for index tests. Sources of heterogeneity were unable to be explored for any test due to a single study in all evaluations. All the included studies were of high or unclear risk of bias, which contributed to the low quality of evidence presented in this review.

Additional weaknesses of this review include the following.

1. Most of the biomarker studies determined the diagnostic cut-off thresholds using a ROC analysis without any subsequent validation in an independent cohort.
2. The variation in the selection of the case and control groups with inclusion of participants that may not reflect a clinically representative population. The reported prevalence of endometriosis in this review (up to 69%) was generally higher than the previously reported prevalence for endometriosis (6% to 10% in the general female population and 35% to 50% in symptomatic women) (Giudice 2004). This may reflect a high risk of patient selection bias in tertiary referral centres, where most of the studies were conducted. Selection bias appeared to be reduced, but not eliminated by consecutively enrolling participants, however the information on method of enrolment was missing in most of the included studies. In this review, 44% of the studies included women with a limited spectrum or specific type of endometriosis. These studies were included to avoid omission of potentially valuable diagnostic information, but could skew the diagnostic estimates in either direction and subsequently interfere with the interpretation of the index test results. It was not possible to evaluate population and disease spectrum effects because of the paucity of suitable data.
3. Inappropriate assignation to the endometriosis and control groups could not be excluded in many studies and is another weakness of the review. Surgical misdiagnosis is a potential cause of bias as the number and experience of the surgical team, the surgical diagnostic criteria and the surgical methods were poorly described in most of the included studies. We now have a standardised technique for performing laparoscopy and we recommend that any future studies use this standardised method of undertaking laparoscopy (Becker 2014). Additionally, we did not confine the studies included in this review to those that reported histological confirmation of endometriotic lesions. In this review, 36% of the included studies relied on surgical diagnosis without histological confirmation. Although a recent ESHRE guideline stated that evidence is lacking to support laparoscopy without histology to confirm endometriosis (Dunselman 2014), the clinical significance of histological verification remains debatable. Diagnosis by surgical visualisation only, remains a common clinical practice and can be considered reliable when an accurate inspection of the abdominal cavity is performed by experienced surgeons. We chose to include the studies that used surgery alone to diagnose endometriosis as we did not wish to lose this potentially valuable information, however this could impact the accuracy of assignment to the case and control groups.

4. Furthermore, excluding unpublished data could potentially eliminate valuable information on the tests that were not altered by endometriosis. The decision to exclude unpublished studies was made due to difficulties in reliable assessment of eligibility and methodological quality of the studies reported only in abstract form. However, this contributed to high risk of publication bias.
5. The optimal methodology of systematic reviews of diagnostic test accuracy is still emerging. This includes assessment and assignment of methodological quality, approach to data analysis and interpretation of the results. There are no well-established criteria for replacement or triage diagnostic tests, therefore we chose criteria that were both realistic and clinically applicable to assist in the interpretation of the complex results. For a replacement test, we considered the threshold reported by the only systematic review on accuracy of the reference standard (laparoscopy) in detecting endometriosis (Wykes 2004) to be the most objective. The meta-analysis was published in 2004 and included four eligible studies comprising 433 women. We acknowledge the limitations associated with emphasising a single review, particularly if it does not present the latest and possibly more accurate data that reflect advances in surgical expertise and technology. A further systematic analysis of more recently published studies to determine the accuracy of laparoscopy was beyond the scope of this review. The criteria for triage tests utilised the common concepts of SnOUT and SpIN in medical statistics and the cut-offs were set at levels we considered to be clinically relevant (see Role of index test(s)). We encourage the readers to apply independent interpretations of the presented diagnostic estimates with using thresholds that may be more applicable to specific populations and clinical circumstances.

Applicability of findings to the review question

QUADAS-2 assigned a low rank to clinical applicability with respect to patient selection in 55% of the studies (6/11), summarised as a high concern. This occurred when the set of women in the study was broader than that seen in clinical practice or when the spectrum of the target condition was limited and the findings may not be applicable to the review question and to clinical practice. Applicability of the index test and reference standard was judged to be satisfactory using the QUADAS-2 tool for all studies. However, the majority of included studies were conducted in academic institutions with a high level of expertise in laboratory techniques or in gynaecological imaging and the index test outcome measures may not be able to be reproduced in all institutions or extrapolated to general gynaecological practice.

Some potentially relevant well-designed studies were excluded as they did not directly address the review question. For example, we did exclude studies that compared endometrioma with other ovarian masses as they either did not meet our inclusion criteria for reproductive age or assessed the numbers of cysts rather than the number of women. Therefore the review question on non-invasive diagnosis of ovarian endometriosis could not be fully addressed. Some forms of endometriosis, such as bladder, ureteric or those involving the extra-pelvic sites (e.g. umbilicus, hernia sacs, abdominal wall, lung, kidney, etc.) were also excluded from the review as they are informed predominantly by case reports or small case series and diagnostic laparoscopy is not an applicable reference test for these conditions. Although these target conditions are rare, from a clinical perspective the diagnostic options for these forms of endometriosis remains unclear.

AUTHORS' CONCLUSIONS

Implications for practice

Although several combinations of tests reached the threshold of diagnostic accuracy to be considered as a replacement test for diagnostic laparoscopy or a triage to improve selection for surgery, these results hinged on only one study in each case, so would need to be confirmed prior to widespread implementation.

One of the combination tests that qualified for a replacement test for detecting endometriosis included endometrial PGP 9.5. It must be noted that PGP 9.5 has not yet reached the criteria for routine use as a low-invasive diagnostic test in clinical practice, as demonstrated in the endometrial biomarkers review (Gupta 2016). Its utility is dependent on its consistency of detection in the endometrium and its consistency of accuracy in diagnosing endometriosis. More work on establishing the best way of endometrial sampling and universal laboratory methods is needed. Besides, in-office sampling of the endometrium may not be applicable to the group of adolescent girls, for whom early diagnosis and abstaining from the diagnostic surgery are particularly important.

Serum CA-125 showed disappointing results and appeared to have no value in diagnosing endometriosis as a single test (Nisenblat 2016a). This is consistent with international guidelines which do not recommend CA-125 testing in women with suspected endometriosis (ACOG 2010; Dunselman 2014; SOGC 2010). CA-125 was incorporated in the diagnostic panels that showed high diagnostic performance, however its value as a part of a combined panel has to be established.

Combination of transvaginal ultrasound (TVUS) with blood biomarkers (CA-125 or CA 19.9) could establish the diagnosis of ovarian endometrioma with high certainty, whereas negative test could not confirm that participants are disease-free. Scrutiny of the diagnostic test accuracy statistics reveals that, in fact, addition of any of these biomarkers does not substantially add to the accuracy of diagnosing endometrioma provided by TVUS alone, as demonstrated in the imaging review from this series (Nisenblat 2016b). Combination of TVUS with vaginal examination was accurate enough to detect endometriosis in the pouch of Douglas (POD), vaginal wall and rectovaginal septum (RVS), but a normal examination could not exclude endometriosis. This is consistent with international guidelines which recommend TVUS as a first line investigation in conjunction with a history and pelvic examination in women with suspected endometriosis, but does not recommend its use as a replacement test for diagnostic surgery (ACOG 2010; Dunselman 2014; SOGC 2010). Considering the findings of the imaging tests review from this series, several imaging methods displayed high accuracy in detecting pelvic, ovarian or deep infiltrating endometriosis (DIE), demonstrating estimates superior to those for imaging and biomarkers combinations (Nisenblat 2016b). These tests included TVUS with bowel preparation (TVUS-BP) and rectal water contrast (RWC-TVS) and MRI, but none of these tests was included in any of the combined test panels.

Rectal endometriosis was the only site that could be accurately detected by using TVUS and pelvic examination. This is particularly

important for detecting rectosigmoid endometriosis as presurgical bowel preparation and surgeries that combine the expertise of gynaecologists and colorectal surgeons (or involve gynaecological surgeons with the expertise to undertake bowel surgery) can be planned preoperatively when rectosigmoid lesions are relatively reliably detected.

Therefore, the evidence on combinations of the tests to be used in clinical practice as a replacement test to supplant laparoscopic diagnosis or a triage test to reduce the requirement for laparoscopic surgical diagnosis remains insufficient. Although diagnostic potential was demonstrated for a number of tests, the level of heterogeneity, wide confidence intervals and high/unclear risk of bias in most studies included in this review series undermines the reliability of the presented results and hence, these data cannot be used confidently to inform clinical practice.

If the findings of large high-quality studies confirm that any of these tests are suitable replacement tests, this would be strong grounds to consider these as an alternative low invasive diagnostic approach instead of the current gold standard laparoscopic surgical diagnosis. An accurate 'negative' non-invasive test is expected to reduce the need for diagnostic surgery in 50% to 70% of women with chronic pelvic pain or infertility (Giudice 2004), although it is likely that some women with a negative test would still require surgery to explore other pathologies. An accurate 'positive' non-invasive test for endometriosis is likely to increase the need for surgery in women with mild symptoms or subfertility (D'Hooghe 2006). The ability to diagnose endometriosis in an outpatient setting would see the diagnosis being made sooner, with fewer follow-up visits, earlier institution of effective treatment and likely a cost-saving benefit for the woman and the health service (both in direct medical costs and in time off work). Other potential advantages of a non-invasive test over a surgical diagnosis include reduced discomfort, shorter recovery times and a reduction in the rare but serious complications of anaesthesia and surgery.

If the findings of the included studies suggesting any of these tests as a triage test can be replicated in other settings, this would be strong grounds to consider these tests to improve selection for more invasive surgical diagnosis. The triage process can be further improved by utilising a sequential approach with both SpIN and SnOUT types of tests (Figure 1). This would reduce the complications and costs of surgery and is expected to cut down long surgical wait lists making surgery more accessible for the women who are likely to benefit from it.

Although guidelines from multiple authorities suggest medical management as a first-line treatment for pelvic pain, most women would prefer having a definite diagnosis before commencing potentially long-term therapy. Also, the indications for fertility management are less clear in 'undiagnosed' women with suspected endometriosis. If therapeutic surgery is considered, reliable detection of ovarian endometriomas potentially enables surgeons to assess ovarian reserve and counsel women about fertility preservation before operating on ovarian tissue and risking a reduction in their future fertility. Reliably detecting DIE/posterior DIE could add weight to a decision to prioritise surgery and could improve preoperative informed consent. Until an accurate non-invasive diagnostic test is developed and tested in large clinical populations, it is impossible to predict accurately its impact on surgical uptake and the number of women that would benefit from performing the test.

It is important to emphasise that in the absence of well-established criteria for an adequate diagnostic test, the diagnostic criteria for replacement and triage tests were determined by the authors of this review series in a way that we believe will aid the interpretation for clinically active readers. However, we encourage readers to apply different criteria according to each clinical population and situation.

Implications for research

Currently randomised controlled trials of treatment require women with and without endometriosis to have had diagnostic surgery for accurate group allocation. For ethical reasons, therapeutic surgery is usually performed at the same time potentially biasing treatment trial outcomes. Thus our current inability to diagnose and assess the progression of endometriosis in a non-invasive way is a significant limitation in the advancement of clinical research in endometriosis.

It does appear that combinations of diagnostic tests hold promise for the future, based purely on the number of studies whose diagnostic test accuracy results reached or approached the threshold for a replacement or a triage test, compared to the paucity of studies in which single tests approached these thresholds.

The QUADAS quality assessment of the included studies identified several weaknesses in study design that can impede an objective evaluation of the findings. We recommend that future researchers undertaking studies for endometriosis diagnostic test accuracy consider: 1) including large cohorts after predefining the sample size via a power calculation (Liu 2005); 2) focusing on a 'single-gate' design that only includes a clinically relevant population (Rutjes 2005); 3) utilising a diagnostic accuracy study design that adheres to the recommendations of the Standards for Reporting of Diagnostic Accuracy (STARD) initiative (Bossuyt 2003); 4) incorporating the QUADAS checklist into the study design (Whiting 2011); 5) formally assessing inter-and intra-observer variability of the laboratory methods and imaging tests; 6) establishing universally acceptable laboratory methodologies (Rahimoglu 2014), radiological protocols and diagnostic criteria for a positive test; 7) utilising universally acceptable methods of performing laparoscopy (Becker 2014) as the reference standard test; 8) implementing validation techniques to assess how the results of a statistical analysis will generalise to an independent data set; 9) undertaking direct comparisons of promising tests in conjunction with a cost effectiveness analysis, 10) applying testing to different clinical phenotypes (Vitonis 2014) rather than to women classified according to rASRM staging; and 11) assessing the long-term outcomes and lifetime healthcare costs of women who have participated in diagnostic test accuracy trials of specific diagnostic tests.

Evaluation of the strongest candidates for possible replacement and triage non-invasive diagnostic tests should continue. Specific opportunities for further research identified by this review include the following.

1. Assessing the diagnostic potential of the tests identified in this and other reviews from the series as promising replacement or triage test for endometriosis in larger, high-quality studies.
   a. Developing a simplified and improved detection technique for endometrial neuronal immuno-histochemical

biomarkers, such as PGP9.5 utilising digitally enhanced assessment of differences in immunohistochemistry appearances; an innovative clinical research to improve techniques of endometrial biopsy, including utilisation of endocervical analgesia, further improvements of the biopsy cannulas and possibly development of endometrial brushes for superficial sampling of the endometrium (although this might not be suited for markers that are only expressed within the stroma).

b. Further development of the composite tests with incorporation of different testing modalities.

c. Attempt to develop a diagnostic algorithm for diagnosis and accurate topographical mapping of endometriosis by utilising several imaging or combined testing methods.

d. Establishing of universal radiological diagnostic criteria and study protocols.

2. Incorporation of clinical history or pelvic examination in a diagnostic model.
3. Exploring the value of sequential testing, implementing SnOUT and SpIN triage tests in diagnosing endometriosis, including clinical parameters in the decision tree algorithm in conjunction with the cost-effectiveness of such testing.
4. Direct comparison of several promising tests in well-designed diagnostic accuracy studies.
5. Evaluation of the whole spectrum of disease across all phases of the menstrual cycle, aiming to identify the most appropriate target population and the best time of testing.
6. Attempting testing in the populations that differ by clinical phenotype rather than by rASRM staging in view of poor correlation of this classification with clinical presentations and treatment outcomes.
7. To add separate evaluations of the biomarker and imaging tests for differentiating ovarian endometrioma from other ovarian masses, including malignant and borderline tumours in women of reproductive age.
8. Assessing the long-term outcomes and lifetime healthcare costs in diagnostic test accuracy trials that have evaluated specific diagnostic blood tests.

ACKNOWLEDGEMENTS

We would like to thank Associate Professor Petra Macaskill for her valuable comments and substantial contribution to the development of the statistical methods for the review. Sincere thanks to the late Professor Ali Akoum and Professor Ian Fraser for an intellectual input and help with drafting of the protocol. We are grateful to Marian Showell, the Information Specialist of the Cochrane Gynaecology and Fertility Group, for her help in designing and conducting the literature search and in locating the full texts of the relevant studies. We also thank the authors and contributors of the review series Devashana Gupta, Emily Liu, Rabia Shaikh and Deepika Arora, for their dedicated assistance in studies' selection process. Finally, we thank all contacted authors who contributed information to this review.

REFERENCES

References to studies included in this review

Cho 2012 *{published data only}*

Cho S, Choi YS, Yim SY, Yang HI, Jeon YE, Lee KE, et al. Urinary vitamin D-binding protein is elevated in patients with endometriosis. *Human Reproduction* 2012;**27**(2):515-22.

el Sharkwy 2013 *{published data only}*

el Sharkwy IAE, Evers JLH, Dunselman GAJ, Vanderlinden PJQ. Combination of non-invasive and semi-invasive tests for diagnosis of minimal to mild endometriosis. *Archives of Gynecology & Obstetrics* 2013;**288**(4):793-7.

Gagné 2003 *{published data only}*

Gagné D, Rivard M, Pagé M, Lépine M, Platon C, Shazand K, et al. Development of a non surgical diagnostic tool for endometriosis based on the detection of endometrial leukocyte subsets and serum CA-125 levels. *Fertility & Sterility* 2003;**80**(4):876-85.

Guerriero 1996a *{published data only}*

Guerriero S, Ajossa S, Paoletti AM, Mais V, Angiolucci M, Melis GB. Tumour marker and transvaginal ultrasonography in the diagnosis of endometrioma. *Obstetrics & Gynaecology* 1996;**88**(3):403-7.

Guerriero 1996b *{published data only}*

Guerriero S, Mais V, Ajossa S, Paoletti AM, Angiolucci M, Melis GB. Transvaginal ultrasonography combined with CA-125 plasma levels in the diagnosis of endometrioma. *Fertility & Sterility* 1996;**65**(2):293-8.

Hudelist 2009 *{published data only}*

Hudelist G, Oberwinkler KH, Singer CF, Tuttlies F, Rauter G, Ritter O, et al. Combination of transvaginal sonography and clinical examination for preoperative diagnosis of pelvic endometriosis. *Human Reproduction* 2009;**24**(5):1018-24.

Koninckx 1996 *{published data only}*

Koninckx P, Meuleman C, Oosterlynck D, Cornillie F. Diagnosis of deep endometriosis by clinical examination during menstruation and plasma CA-125 concentration. *Fertility & Sterility* 1996;**65**(2):280-7.

Marasinghe 2014 *{published data only}*

Marasinghe J, Senanayake H, Saravanabhava N, Arambepola C, Condous G, Greenwood P. History, pelvic examination findings and mobility of ovaries as a sonographic marker to detect pelvic adhesions with fixed ovaries. *Journal of Obstetrics and Gynaecology Research* 2014;**40**(3):785-90.

Paiva 2014 *{published data only}*

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* 2014;**6**(3):135-43.

Yun 2014 *{published data only}*

Yun BH, Lee YS, Chon SJ, Jung YS, Yim SY, Kim HY, et al. Evaluation of elevated urinary enolase I levels in patients with endometriosis. *Biomarkers* 2014;**19**(1):16-21.

Zeng 2005 *{published data only}*

Zeng F, Xue M, Zevallos HBV, Lai D, Arthur J, Ng C, et al. Diagnostic value of the detection of aromatase cytochrome P450 and CA125 for endometriosis [芳香化酶细胞色素P450及CA125 联合检测对子宫内膜异位症的诊断价值]. *Journal of Central South University (Medical Sciences)* 2005;**30**(6):682-5.

References to studies excluded from this review

Abrao 2007 *{published data only}*

Abrao MS, Goncalves MODC, 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. *Human Reproduction* 2007;**22**(12):3092-7.

Adamyan 1993 *{published data only}*

Adamyan L, Fanchenko N, Alexeyeva M, Andreyeva Y, Novikov Y, Jahan I. Hormonal and immunologic methods in the diagnosis and treatment of patients with benign ovarian tumors and endometriotic cysts. *International Journal of Fertility* 1993;**38**(2):92-8.

Alcazar 2011 *{published data only}*

Alcazar J, Guerriero S, Minguez J, Ajossa S, Paoletti A, Ruiz-Zambrana A, et al. Adding cancer antigen 125 screening to gray scale sonography for predicting specific diagnosis of benignadnexal masses in premenopausal women: is it worthwhile?. *Journal of Ultrasound in Medicine* 2011;**30**(10):1381-6.

Badawy 1984 *{published data only}*

Badawy SZ, Cuenca V, Stitzel A, Jacobs RD, Tomar RH. Autoimmune phenomena in infertile patients with endometriosis. *Obstetrics & Gynecology* 1984;**63**(3):271-5.

Bazot 2009 *{published data only}*

Bazot M, Lafont C, Rouzier R, Roseau G, Thomassin-Naggara I, Daraii E. Diagnostic accuracy of physical examination, transvaginal sonography, rectal endoscopic sonography, and magnetic resonance imaging to diagnose deep infiltrating endometriosis. *Fertility & Sterility* 2009;**92**(6):1825-33.

Borboletto 1995 *{published data only}*

Borboletto C, Goncalves W, Giusa M, de Freitas V, Baracat E, de Lima G. Transvaginal ultrasound, color Doppler velocimetry and CA -125 in the diagnosis and follow-up of pelvic endometriosis [Ultra-sonografia transvaginal, doplervelocimetria colorida e dosagem de CA-125 no diagnóstico e no seguimento da endometriose pélvica]. *Femina* 1995;**23**(8):723-6.

Cho 2007 *{published data only}*

Cho S, Oh Y, Nam A, Kim H, Park J, Kim J, et al. Evaluation of serum and urinary angiogenic factors in patients with endometriosis. *American Journal of Reproductive Immunology* 2007;**58**(6):497-504.

da Silva 2014 *{published data only}*

da Silva C, Belo A, Andrade S, Campos P, Ferreira M, da Silva-Filho A, et al. Identification of local angiogenic and inflammatory markers in the menstrual blood of women with endometriosis. *Biomedicine and Pharmacotherapy* 2014;**68**(7):899-904.

Dias 2012 *{published data only}*

Dias J, Podgaec S, de Oliveira R, Marin M, Baracat E, Abrao M. Patients with endometriosis of the rectosigmoid have a higher percentage of natural killer cells in peripheralblood. *Journal of Minimally Invasive Gynecology* 2012;**19**(3):317-24.

Eskenazi 2001 *{published data only}*

Eskenazi B, Warner M, Bonsignore L, Olive D, Samuels S, Vercellini P. Validation study of nonsurgical diagnosis of endometriosis. *Fertility & Sterility* 2001;**76**(5):929-35.

Fedele 1988 *{published data only}*

Fedele L, Vercellini P, Arcaini L, Grazia da Dalt M, Candiani G. CA 125 in serum, peritoneal fluid, active lesions, and endometrium of patients with endometriosis. *American Journal of Obstetrics & Gynecology* 1988;**158**(1):166-70.

Guerriero 1997 *{published data only}*

Guerriero S, Mallarini G, Ajossa S, Risalvato A, Satta R, Mais V, et al. Transvaginal ultrasound and computed tomography combined with clinical parameters and CA-125 determinations in the differential diagnosis of persistent ovarian cysts in premenopausal women. *Ultrasound in Obstetrics & Gynecology* 1997;**9**(5):339-43.

Hudelist 2011b *{published data only}*

Hudelist G, Ballard K, English J, Wright J, Banerjee S, Mastoroudes H, et al. Transvaginal sonography vs. clinical examination in the preoperative diagnosis of deep infiltrating endometriosis. *Ultrasound in Obstetrics & Gynecology* 2011;**37**(4):480-7.

Kocbek 2014 *{published data only}*

Kocbek V, Bersinger N, Brglez V, Mueller M, Petan T, Rizner T. Phospholipase A2 group IIA is elevated in endometriomas but not in peritoneal fluid and serum of ovarian endometriosis patients. *Gynecological Endocrinology* 2015;**31**(3):214-8.

Kuessel 2014 *{published data only}*

Kuessel L, Jaeger-Lansky A, Pateisky P, Rossberg N, Schulz A, Schmitz A, et al. Cytokeratin-19 as a biomarker in urine and in serum for the diagnosis ofendometriosis – a prospective study. *Gynecological Endocrinology* 2014;**30**(1):38-41.

Lee 2014 *{published data only}*

Lee Y, Tan CW, Venkatratnam A, Tan CS, Liang C, Loh S, et al. Dysregulated sphingolipid metabolism in endometriosis. *Journal of Clinical Endocrinology & Metabolism* 2014;**99**(10):1913-21.

Nezhat 1994 *{published data only}*

Nezhat C, Santolaya J, Nezhat F, Nezhat C. Comparison of transvaginal sonography and bimanual pelvic examination in patients with laparoscopicallyconfirmed endometriosis. *Journal of the American Association of Gynecologic Laparoscopists* 1994;**1**(2):127-30.

Szubert 2014 *{published data only}*

Szubert M, Suzin J, Duechler M, Szuławska A, Czyż M, Kowalczyk-Amico K. Evaluation of selected angiogenic and inflammatory markers in endometriosis before and after danazol treatment. *Reproduction, Fertility and Development* 2014;**26**(3):414-20.

Weerakiet 2000 *{published data only}*

Weerakiet S, Wongkularb A, Rochanawutanon M, Rojanasakul A. Transvaginal ultrasonography combined with pelvic examination in the diagnosis of ovarian endometrioma. *Journal of the Medical Association of Thailand* 2000;**83**(5):523-8.

Wolfler 2005 *{published data only}*

Wolfler M, Nagele F, Kolbus A, Seidl S, Schneider B, Huber J. A predictive model for endometriosis. *Human Reproduction* 2005;**20**(6):1702-8.

Yong 2013 *{published data only}*

Yong P, Sutton C, Suen M, Williams C. Endovaginal ultrasound-assisted pain mapping in endometriosis and chronic pelvic pain. *Journal of Obstetrics & Gynaecology* 2013;**33**:715-9.

Additional references

ACOG 2010

The American College of Obstetricians and Gynecologists. Practice Bulletin No. 114: Management of Endometriosis. *Obstetrics and Gynecology* 2010;**116**(1):223-36.

Adamson 2008

Adamson GD, Pasta DJ. Endometriosis Fertility Index (EFI): the new validated endometriosis staging system. *Fertility and Sterility* 2010;**94**(5):1609-15.

Almeida Filho 2008

Almeida Filho DP, Oliveira LJ, Amaral VF. Accuracy of laparoscopy for assessing patients with endometriosis. *Sao Paulo Medical Journal* 2008;**126**:305-8.

American Society for Reproductive Medicine 1997

American Society for Reproductive Medicine. Revised American Society for Reproductive Medicine classification of endometriosis: 1996. *Fertility and Sterility* 1997;**67**(5):817-21.

Ballard 2008

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. *British Journal of Obstetrics and Gynaecology* 2008;**115**(11):1382-91.

Batt 2003

Batt R, Mitwally MF. Endometriosis from thelarche to midteens: pathogenesis and prognosis, prevention and pedagogy. *Journal of Pediatric and Adolescent Gynecology* 2003;**16**:333-47.

Becker 2014

Becker CM, Laufer MR, Stratton P, Hummelshoj l, Missmer SA, Zondervan KT, et al. World Endometriosis Research Foundation Endometriosis Phenome and biobanking harmonization project: I. Surgical phenotype data collection in endometriosis research. *Fertility and Sterility* 2014;**102**(5):1213-22.

Bossuyt 2003

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003;**326**(7379):41-4.

Bossuyt 2008

Bossuyt PM, Leeflang MM. Chapter 6: Developing Criteria for Including Studies. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.4 [updated September 2008]. The Cochrane Collaboration, 2008.

Chapron 2003a

Chapron C, Fauconnier A, Vieira M, Barakat H, Dousset B, Pansini V, et al. Anatomical distribution of deeply infiltrating endometriosis: surgical implications and proposition for a classification. *Human Reproduction* 2003;**18**(1):157-61.

Chapron 2003b

Chapron C, Fauconnier A, Dubuisson JB, Barakat H, Vieira M, Bréart G. Deep infiltrating endometriosis: relation between severity of dysmenorrhea and extent of disease. *Human Reproduction* 2003;**18**:760-6.

Chapron 2003c

Chapron C, Cravello L, Chopin N, Kreiker G, Blanc B, Dubuisson JB. Complications during set-up procedures for laparoscopy in gynecology: open laparoscopy does not reduce the risk of major complications. *Acta Obstetrica Gynecologica Scandinavica* 2003;**82**:1125-9.

D'Hooghe 2001

D'Hooghe TM, Bambra CS, Xiao L, Peixe K, Hill JA. Effect of menstruation and intrapelvic injection of endometrium on inflammatory parameters of peritoneal fluid in the baboon (Papio anubis and Papio cynocephalus). *American Journal of Obstetrics and Gynecology* 2001;**184**(5):917-25.

D'Hooghe 2006

D'Hooghe TM, Mihalyi AM, Simsa P, Kyama CK, Peeraer K, De Loecker P, et al. Why we need a non-invasive diagnostic test for minimal to mild endometriosis with a high sensitivity. *Gynecologic and Obstetric Investigation* 2006;**62**(3):136-8.

de Vet 2008

de Vet HCW, Eisinga A, Riphagen II, Aertgeerts B, Pewsner D. Chapter 7: Searching for studies. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.4 [updated September 2008]. The Cochrane Collaboration, 2008.

Dmowski 1997

Dmowski WP, Lesniewicz R, Rana N, Pepping P, Noursalehi M. Changing trends in the diagnosis of endometriosis: a comparative study of women with pelvic endometriosis presenting with chronic pelvic pain or infertility. *Fertility and Sterility* 1997;**67**:238-43.

Dunselman 2014

Dunselman GA, Vermeulen N, Becker C, Calhaz-Jorge C, D'Hooghe T, De Bie B, et al. ESHRE guideline: management of women with endometriosis. *Human Reproduction* 2014;**29**(3):400-12.

Fassbender 2015

Fassbender A, Burney RO, Dorien FO, D'Hooghe T, Giudice L. Update on biomarkers for the detection of endometriosis. *BioMed Research International* 2015;**Epub**:1-14.

Fauconnier 2005

Fauconnier A, Chapron C. Endometriosis and pelvic pain: epidemiological evidence of the relationship and implications. *Human Reproduction Update* 2005;**11**:595-606.

Frishman 2006

Frishman GN, Salak JR. Conservative surgical management of endometriosis in women with pelvic pain. *Journal of Minimally Invasive Gynecology* 2006;**13**:546-58.

Gao 2006

Gao X, Yeh YC, Outley J, Simon J, Botteman M, Spalding J. Health-related quality of life burden of women with endometriosis: a literature review. *Current Medical Research and Opinion* 2006;**22**:1787-97.

Giudice 2004

Giudice LC, Kao LC. Endometriosis. *Lancet* 2004;**364**:1789-99.

Guerriero 2015

Guerriero S, Ajossa S, Orozco R, Perniciano M, Jurado M, Melis GB, et al. Diagnostic accuracy of transvaginal ultrasound for diagnosis of deep endometriosis in the recto-sigmoid: a meta-analysis. *Ultrasound in Obstetrics and Gynecology* 2015;**Epub ahead of print**:1.

Guo 2009

Guo SW. Recurrence of endometriosis and its control. *Human Reproduction Update* 2009;**15**(4):441-61.

Gupta 2016

Gupta D, Hull ML, Fraser I, Miller L, Bossuyt PMM, Johnson N, et al. Endometrial biomarkers for the non-invasive diagnosis of endometriosis. *Cochrane Database of Systematic Reviews* 2016, Issue 4. [DOI: 10.1002/14651858.CD012165]

Guzick 1997

Guzick DS, Silliman NP, Adamson GD, Buttram VC Jr, Canis M, Malinak LR, et al. Prediction of pregnancy in infertile women based on the American Society for Reproductive Medicines revised classification of endometriosis. *Fertility and Sterility* 1997;**67**(5):822-9.

Hudelist 2011a

Hudelist G, English J, Thomas AE, Tinelli A, Singer CF, Keckstein J. Diagnostic accuracy of transvaginal ultrasound for non-invasive diagnosis of bowel endometriosis: systematic review and meta-analysis. *Ultrasound in Obstetrics and Gynecology* 2011;**37**(3):257-63.

Hull 2008

Hull ML, Escareno CR, Godsland JM, Doig JR, Johnson CM, Phillips SC, et al. Endometrial-peritoneal interactions during endometriotic lesion establishment. *American Journal of Pathology* 2008;**173**(3):700-15.

Johnson 2013

Johnson NP, Hummelshoj L. Consensus on current management of endometriosis. *Human Reproduction* 2013;**28**(6):1552-68.

Johnson 2015

Johnson NP, Hummelshoj L, Adamson GD, Kecstein J, Taylor H, Abrao MS, et al. Consensus on the classification of endometriosis. *Human Reproduction* 2016 in press;**in preparation/press**:1.

Kao 2003

Kao LC, Germeyer A, Tulac S, Lobo S, Yang JP, Taylor RN, et al. Expression profiling of endometrium from women with endometriosis reveals candidate genes for disease-based implantation failure and infertility. *Endocrinology* 2003;**144**:2870-81.

Kennedy 2005

Kennedy S, Bergqvist A, Chapron C, D'Hooghe T, Dunselman G, Greb R, et al. ESHRE guideline for the diagnosis and treatment of endometriosis. *Human Reproduction* 2005;**20**:2698-704.

Koninckx 1991

Koninckx PR, Meuleman C, Demeyere S, Lesaffre E, Cornillie FJ. Suggestive evidence that pelvic endometriosis is a progressive disease, whereas deeply infiltrating endometriosis is associated with pelvic pain. *Fertility and Sterility* 1991;**55**(4):759-65.

Ling 1999

Ling F. Randomized controlled trial of depot leuprolide in patients with chronic pelvic pain and clinically suspected endometriosis. *Obstetrics and Gynecology* 1999;**93**:51-8.

Liu 2005

Liu A, Schisterman EF, Mazumdar M, Hu J. Power and sample size calculation of comparative diagnostic accuracy studies with multiple correlated test results. *Biometrical Journal* 2005;**47**(2):140-50.

Liu 2015

Liu E, Nisenblat V, Farquhar C, Fraser I, Bossuyt PMM, Johnson N, et al. Urinary biomarkers for the non-invasive diagnosis of endometriosis. *Cochrane Database of Systematic Reviews* 2015, Issue 12. [DOI: 10.1002/14651858.CD012019]

Marchino 2005

Marchino GL, Gennarelli G, Enria R, Bongioanni F, Lipari G, Massobrio M. Diagnosis of pelvic endometriosis with use of macroscopic versus histologic findings. *Fertility and Sterility* 2005;**84**:12-5.

Martin 2006

Martin DC. Applying STARD criteria to the laparoscopic identification of endometriosis (abstract) . Fertility and Sterility. 2006; Vol. 86 Suppl 2:270.

Matsuzaki 2006

Matsuzaki S, Canis M, Pouly JL, Rabischong B, Botchorishvili R, Mage G. Relationship between delay of surgical diagnosis and severity of disease in patients with symptomatic deep infiltrating endometriosis. *Fertility and Sterility* 2006;**86**:1314-6.

May 2010

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):651-74.

May 2011

May KE, Villar J, Kirtley S, Kennedy SH, Becker CM. Endometrial alterations in endometriosis: a systematic review of putative biomarkers. *Human Reproduction Update* 2011;**17**(5):637-53.

McGraw-Hill Dictionary of Medicine 2006

Segen JC (author). McGraw-Hill Concise Dictionary of Modern Medicine. The McGraw-Hill Companies, Inc, 2006.

Medeiros 2009

Medeiros LRF, Rosa DD, Bozzetti MC, Fachel JMG, Furness S, Garry R, et al. Laparoscopy versus laparotomy for benign ovarian tumour. *Cochrane Database of Systematic Reviews* 2009, Issue 2. [DOI: 10.1002/14651858.CD004751.pub3]

Medeiros 2015

Medeiros LR, Rosa MI, Silva BR, Reis ME, Simon CS, Dondossola ER, et al. Accuracy of magnetic resonance in deeply infiltrating endometriosis: a systematic review and meta-analysis. *Archives of Gynecology and Obstetrics* 2015;**291**(3):611-21.

Moore 2002

Moore J, Copley S, Morris J, LIndsell D, Golding S, Kennedy S. A systematic review of the accuracy of ultrasound in the diagnosis of endometriosis. *Ultrasound in Obstetrics & Gynecology* 2002;**20**:630-4.

Nisenblat 2016a

Nisenblat V, Bossuyt PMM, Shaikh R, Arora D, Farquhar C, Jordan V, et al. Blood biomarkers for the non-invasive diagnosis of endometriosis. *Cochrane Database of Systematic Reviews* 2016, Issue 5. [DOI: 10.1002/14651858.CD012179]

Nisenblat 2016b

Nisenblat V, Bossuyt PMM, Farquhar C, Johnson N, Hull ML. Imaging modalities for the non-invasive diagnosis of

endometriosis. *Cochrane Database of Systematic Reviews* 2016, Issue 2. [DOI: 10.1002/14651858.CD009591.pub2]

Nyholt 2012

Nyholt DR, Low SK, Anderson CA, Painter JN, Uno S, Morris AP, et al. Genome-wide association meta-analysis identifies new endometriosis risk loci. *Nature Genetics* 2012;**44**(12):1355-9.

Rahimoglu 2014

Rahmioglu N, Fassbender A, Vitonis AF, Tworoger SS, Hummelshoj L, D'Hooghe TM, et al. World Endometriosis Research Foundation Endometriosis Phenome and biobanking harmonization project: III. Fluid biospecimen collection, processing, and storage in endometriosis research. *Fertility and Sterility* 2014;**102**(5):1233-43.

Redwine 2003

Redwine DB. Invisible' microscopic endometriosis: a review. *Gynecologic and Obstetric Investigation* 2003;**55**:63-7.

Rogers 2009

Rogers PA, D'Hooghe TM, Fazleabas A, Gargett CE, Giudice LC, Montgomery GW, et al. Priorities for Endometriosis Research: recommendations from an international consensus workshop. *Reproductive Sciences* 2009;**16**(4):335-46.

Rutjes 2005

Rutjes AWS, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PMM. Case–control and two-gate designs in diagnostic accuracy studies. *Clinical Chemistry* 2005;**51**(8):1335-41.

Sampson 1927

Sampson JA. Peritoneal endometriosis due to menstrual dissemination of endometrial tissue into the peritoneal cavity. *American Journal of Obstetrics and Gynecology* 1927;**14**:442-69.

Simoens 2012

Simoens S, Dunselman G, Dirksen C, Hummelshoj L, Bokor A, Brandes I, et al. The burden of endometriosis: costs and quality of life of women with endometriosis and treated in referral centres. *Human Reproduction* 2012;**27**(5):1292-9.

Sinaii 2002

Sinaii N, Cleary SD, Ballweg ML, Nieman LK, Stratton P. High rates of autoimmune and endocrine disorders, fibromyalgia, chronic fatigue syndrome and atopic diseases among women with endometriosis: A survey analysis. *Human Reproduction* 2002;**17**(10):2715-24.

SOGC 2010

Society of Obstetricians Gynaecologists of Canada. Endometriosis: diagnosis and management. SOGC clinical practice guideline no. 244. *Journal of Obstetrics and Gynaecology Canada* 2010;**32**:S1-S28.

Somigliana 2006

Somigliana E, Vigano P, Parazzini F, Stoppelli S, Giambattista E, Vercellini P. Association between endometriosis and cancer: A comprehensive review and a critical analysis of clinical and epidemiological evidence. *Gynecologic Oncology* 2006;**101**(2):331-41.

Spaczynski 2003

Spaczynski RZ, Duleba AJ. Diagnosis of endometriosis. *Seminars in Reproductive Medicine* 2003;**21**:193-208.

Stegmann 2008

Stegmann BJ, Sinaii N, Liu S, Segars J, Merino M, Nieman LK, et al. Using location, color, size, and depth to characterize and identify endometriosis lesions in a cohort of 133 women. *Fertility and Sterility* 2008;**89**:1632-6.

The Gale Encyclopedia of Medicine 2008

Donna Olendorf (Editor), Christine Jeryan (Editor), Karen Boyden (Editor), Gale Group (Corporate Author). The Gale Encyclopedia of Medicine (5 volume set). The Gale Group, Inc, 2008.

Vercellini 1996

Vercellini P, Trespidi L, De Giorgi O, Cortesi I, Parazzini F, Crosignani GP. Endometriosis and pelvic pain: relation to disease stage and localization. *Fertility and Sterility* 1996;**65**:299-304.

Vigano 2004

Vigano P, Parazzini F, Somigliana E, Vercellini P. Endometriosis: epidemiology and aetiological factors. *Best Practice & Research. Clinical Obstetrics and Gynaecology* 2004;**18**:177-200.

Vitonis 2014

Vitonis AF, Vincent K, Rahmioglu N, Fassbender A, Buck Louis GM, Hummelshoj L, et al. WERF EPHect Working Group. World Endometriosis Research Foundation Endometriosis Phenome and biobanking harmonization project: II. Clinical and covariate phenotype data collection in endometriosis research. *Fertility and Sterility* 2014;**102**(5):1223-32.

Whiting 2005

Whiting PF, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. *BMC Medical Research Methodology* 2005;**5**:19.

Whiting 2011

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. the QUADAS-2 Group. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Annals of Internal Medicine* 2011;**155**(8):529-36.

Wykes 2004

Wykes CB, Clark TJ, Khan KS. Accuracy of laparoscopy in the diagnosis of endometriosis: a systematic quantitative review. *BJOG - an International Journal of Obstetrics and Gynaecology* 2004;**111**:1204-12.

Yeung 2009

Yeung PP Jr, Shwayder J, Pasic RP. Laparoscopic management of endometriosis: comprehensive review of best evidence. *Journal of Minimally Invasive Gynecology* 2009;**16**:269-81.

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Cho 2012

Study characteristics	
Patient sampling	**Primary objectives** To investigate proteins secreted in urine of women with endometriosis using proteomic techniques in order to identify potential markers for the clinical diagnosis of endometriosis; to evaluate urinary VDBP in women with endometriosis **Study population** Women who underwent laparoscopy for various indications including pelvic masses, pelvic pain, suspicious endometriosis, infertility and diagnostic evaluation **Selection criteria** Inclusion criteria: pre-menopausal age; exclusion criteria: previous hormone or GnRH agonist use, adenomyosis, endometrial cancer, hyperplasia or endometrial polyps, infectious diseases, chronic or acute inflammatory diseases, malignancy, autoimmune disease and cardiovascular disease **Study design** Cross-sectional, single-gate design, prospective collection of samples
Patient characteristics and setting	**Clinical presentation** Pelvic masses, pelvic pain, suspicious endometriosis, infertility **Ag:** Mean age 34.22 ± 6.88 years (endometriosis group), 32.76 ± 10.26 years (control group) **Number of participants enrolled** 95 women **Number of participants available for analysis** 95 women (in follicular or luteal cycle phase, numbers not specified) **Setting** Gangnam Severance Hospital, Yonsei University College of Medicine **Place of study** Seoul, Korea **Period of study** January 2008 - October 2010 **Language** English
Index tests	**Index test** Urinary VDBP-Cr x serum CA-125 **Details of the index test procedure as stated** Urinary VDBP was measured using specific commercial sandwich ELISA assays according to manufacturer's protocols (ALPCO Diagnostics, Salem, NH, USA); urine VDBP values were normalized to urine Cr concentrations; serum CA-125 were measured using CA-125 II electro chemiluminescence immunoassay with the Roche/Hitachi Modular Analytics E170 system (Roche Diagnostics, Tokyo, Japan); sample handling described **Threshold for positive result** Cut-off value > 2755, not pre-specified **Examiners:** No information provided; unclear if blinded to the result of reference standard **Interobserver variability** Not reported
Target condition and reference standard(s)	**Target condition:** Endometriosis **Prevalence of target condition in the sample:** n = 57/95 (60%): stage I-II 5, stage III-IV 52; controls n = 38 **Reference standard:** Laparoscopy and histology **Description of positive case definition by reference standard as reported:** Visual inspection, confirmed by histopathology; staging according to the rASRM classification

Cho 2012 *(Continued)*

	Examiners: No information provided
Flow and timing	**Time interval between index test and reference standard:** Blood was collected preoperatively, urine was collected after induction of anaesthesia **Withdrawals:** None reported
Comparative	
Key conclusions by the authors	Urinary VDBP levels are elevated in women with endometriosis, but they have limited value as a potential diagnostic biomarker for endometriosis (sensitivity 58%, specificity 76%)
Conflict of interest	The authors reported no conflict of interests; supported by the Basic Science Research Program of NRF of Korea by the Ministry of Education, Science and Technology (2010-0023323)
Notes	The reported diagnostic estimates for urinary or blood test alone are presented in other reviews from this series

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
Was a 'two-gate' design avoided?	Yes		
		Unclear	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it prespecified?	No		
Was a cycle phase considered in interpretation of the result of index test?	Yes		
Did the study provide a clear prespecified definition of what was considered to be a positive result of index test?			
Was the index test performed by a single operator or interpreted by consensus in a joint session?			
Were the same clinical data available when the index test results			

Cho 2012 *(Continued)*

were interpreted as would be available when the test is used in practice?			
		High	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

el Sharkwy 2013

Study characteristics	
Patient sampling	**Primary objectives** To evaluate the diagnostic value of serum measurement of IL-6 combined with the presence of nerve fibres in the functional layer of endometrium for diagnosis of minimal-mild endometriosis **Study population** Women undergoing laparoscopy for evaluation of infertility or pelvic pain at the authors' institution **Selection criteria** Inclusion criteria: reproductive age (18 - 36 years), follicular cycle phase, regular menstrual cycle; exclusion criteria: any current infection (genital or systemic), any medication within 1/12 prior to laparoscopy, previous surgery for endometriosis, smoking or drinking alcohol **Study Design** Cross-sectional, single-gate design, prospective recruitment and collection of samples
Patient characteristics and setting	**Clinical presentation** Dysmenorrhoea - 64/114, dyspareunia - 17/114, dyschezia - 6/114, pelvic pin - 35/114, infertility - 91/114 **Age** Mean age 31 ± 1.1 years (endometriosis group), 29 ± 0.6 years (controls) **Number of participants enrolled** 114 women

el Sharkwy 2013 *(Continued)*

	Number of participants available for analysis 78 women (only minimal-mild endometriosis included; all in follicular cycle phase) **Setting** Department of O&G, Zagazig University Hospital **Place of study** Zagazig, Egypt **Period of study** December 2010 - April 2012 **Language** English
Index tests	**Index test** IL-6 in serum + PGP 9.5 in endometrium **Details of the index test procedure as stated** Serum IL-6 level assessed using a commercially available ELISA (DRG, Germany); endometrial PGP 9.5 assessed using immunohistochemistry (assessment using Olympus microscope, normal skin as positive control, rabbit immunoglobulin fraction as a negative control); sample processing described **Threshold for positive result** IL-6 > 15.4 pg/ml, PGP 9.5 - presence of nerve fibres in functional layer of endometrium; not pre-specified **Examiners** IL-6 - no information provided; endometrial nerve fibres - experienced gynaecologist and two experienced pathologists; unclear if blinded to the result of reference standard **Interobserver variability** Nerve fibre counting - 96% correlations between the two pathologists
Target condition and reference standard(s)	**Target condition** Endometriosis **Prevalence of target condition in the sample** n = 74/114 (65%): stage I-II 38, stage III-IV 36; controls n = 40 **Reference standard** Laparoscopy n = 114 (100%) **Description of positive case definition by reference standard test as reported** Visual inspection; staging according to the rASRM classification **Examiners** Three experienced gynaecologists in endometriosis
Flow and timing	**Time interval between index test and reference standard** Blood and endometrial samples were obtained on the day of surgery **Withdrawals** 36 participants with moderate-severe disease were not included in final analysis
Comparative	
Key conclusions by the authors	Combination of both serum IL-6 and presence of nerve fibres in the endometrium is more reliable method for diagnosis of minimal-mild endometriosis than in single test
Conflict of interest	The authors declared no conflict of interest
Notes	The reported separate data for endometrial or blood biomarkers alone are presented in other reviews from this series The evaluations were performed only for minimal-mild disease

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			

el Sharkwy 2013 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
Was a 'two-gate' design avoided?	Yes		
		Unclear	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it prespecified?	No		
Was a cycle phase considered in interpretation of the result of index test?	Yes		
Did the study provide a clear prespecified definition of what was considered to be a positive result of index test?			
Was the index test performed by a single operator or interpreted by consensus in a joint session?			
Were the same clinical data available when the index test results were interpreted as would be available when the test is used in practice?			
		High	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			

el Sharkwy 2013 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
	High

Gagné 2003

Study characteristics	
Patient sampling	**Primary objectives** To determine whether the proportion of several leukocyte subsets is modulated in the endometrium of women with endometriosis and, if yes, whether it can be used for diagnostic purposes **Study population** Women scheduled to undergo laparoscopy or laparotomy at one of the eight clinical institutions in Montreal **Selection criteria** Inclusion criteria: women of premenopausal age who had never been pregnant, luteal phase of the menstrual cycle (based on the last period and further confirmed by histology), regular cycles (21 - 35 days), not acute salpingitis, no hormonal treatment or intrauterine device in previous three months. **Study Design** Multi-centre study of two-gate design, prospective recruitment, random sample of women [participation rate 94%]
Patient characteristics and setting	**Clinical presentation** Infertility (7% controls, 16% cases); pain (19% controls, 33% cases); pelvic mass (8% controls, 13% cases); fibroids (9% controls, 15% cases); menorrhagia (2% controls, 4% cases); tubal ligation (60% controls, 25% cases); hysterectomy (19% controls, 32% cases); diagnostic laparoscopy (20% controls, 43% cases); history of endometriosis (3% controls, 16% cases) **Age** Random sampling from a population with mean age of 37.3 ± 6.4 years **Number of participants enrolled** 368 women **Number of participants available for analysis** 368 women (in luteal phase of menstrual cycle) **Setting** Biotech firm - MetrioGene BioSciences (a subsidiary of PROCREA BioSciences) **Place of study** Montreal, Canada **Period of study** July 1997 - May 2001 **Language** English
Index tests	**Index test**: Predictive model including: clinical history (length of menses) + serum CA-125 level + endometrial leukocytes (CD3+, CD16+, CD3-HLADR-, CD3-CD45RA-, CD3+CD16-, CD3+CD56-, CD56-CD16+, CD16b+) **Details of the index test procedure as stated** Clinical history was collected using a questionnaire in which a clinical profile as well as information concerning personal habits, menstrual characteristics, crude evaluation of the intensity of pain, and contraception and parity were obtained by the investigators; serum CA-125 level determined by using a one step-sandwich radioimmunoassay (Fujirebio America Inc.) with assay sensitivity 0.4 U/ml; endometrial leukocyte subsets determined by Coulter EPICS XL

Gagné 2003 *(Continued)*

	flow cytometer (Beckman/Coulter) with an argon laser operating at 488 nm and detectors at 525, 575, 610, and 675 nm to measure FITC, PE, ECD, and PerCP emission, respectively; the percentage of cells with the markers of interest was evaluated within the CD45+ cells; sample handling and laboratory procedure described in details; predictive model by a multiple logistic regression with subsequent bootstrap method validation by drawing 200 replicate samples with replacement from the original data set **Threshold for positive result** Predictive model - predictive probability 0.61, not pre-specified; CA-125 > 12.8 U/ml and >35 U/ml; endometrial leukocytes: cut-offs different for each subset selected by ROC, not pre-specified **Examiners** No information provided; unclear if were blinded to the result of reference standard **Interobserver variability** CA-125: Inter- and intra-assay variations < 5%
Target condition and reference standard(s)	**Target condition** Endometriosis **Prevalence of target condition in the sample** n = 173/368 (47%): stage I-II 78%, stage III-IV 22%; controls n = 195 **Reference standard** Laparoscopy/Laparotomy n = 368 (100%) **Description of positive case definition by reference standard test as reported** Cases were defined by the presence of endometriotic lesions confirmed at the time of surgical examination; staging according to the ASRM system **Examiners** Gynaecologists collaborating in the study were trained surgeons experienced with the management of endometriosis who were skilled in detecting and identifying all forms of endometriotic lesions
Flow and timing	**Time interval between index test and reference standard** Blood and endometrial samples were obtained on the day of surgery, clinical data were obtained preoperatively **Withdrawals** None
Comparative	
Key conclusions by the authors	The predictive model represents a novel diagnostic tool to identify women with a high likelihood of suffering from endometriosis
Conflict of interest	All the authors except RM are (or were) employees of PROCREA BioSciences; supported by the Industrial Research Assistance Program (IRAP) from NSERC grant #15453Q and internal resources at PROCREA BioSciences
Notes	The reported data for blood biomarkers or endometrial test alone are presented in other reviews from this series The reported diagnostic estimates per severity of endometriosis are not presented in this review

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		

Gagné 2003 *(Continued)*

Was a 'two-gate' design avoided?	No		
		High	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Was a cycle phase considered in interpretation of the result of index test?	Yes		
Did the study provide a clear pre-specified definition of what was considered to be a positive result of index test?			
Was the index test performed by a single operator or interpreted by consensus in a joint session?			
Were the same clinical data available when the index test results were interpreted as would be available when the test is used in practice?			
		High	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			

Gagné 2003 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
	Low

Guerriero 1996a

Study characteristics	
Patient sampling	**Primary objectives** To evaluate the accuracy of CA-19.9 plasma levels (with or without CA-125 levels) combined with transvaginal ultrasonography in the differential diagnosis of endometriosis **Study population** Women undergoing laparoscopy or laparotomy for persistent adnexal mass at the authors' institution **Selection criteria** Inclusion criteria: premenopausal, non-pregnant (only moderate-severe endometriosis included) **Study Design** Cross-sectional, single-gate design, prospective recruitment and collection of samples, consecutive series
Patient characteristics and setting	**Clinical presentation** Pelvic mass - 100%, infertility - 53% **Age** Mean age 33.3 ± 9.6 years **Number of participants enrolled** 118 women **Number of participants available for analysis** 118 women (only moderate-severe endometriosis included; all in follicular cycle phase) **Setting** Department of O&G, University of Cagliari **Place of study** Cagliari, Italy **Period of study** November 1994 - November 1995 **Language** English
Index tests	**Index test**: CA-19.9 ± CA-125 in serum + Transvaginal Ultrasonography (TVUS) **Details of the index test procedure as stated** Serum CA-125 levels assessed by immunoradiometric assay (CIS Bio International, Gif sur Yvette, France), limit of detection 0.5 U/ml; serum CA-19.9 levels assessed by immunoradiometric assay (CIS Bio International, Gif sur Yvette, France), limit of detection 1.5 U/ml; TVUS performed with a 5 MHz endovaginal probe (Acuson XP/10 OB ultrasound system), procedure described in details **Threshold for positive result** CA-125 ≥ 25 IU/ml, pre-specified; CA-19.9 ≥ 12 U/ml, not pre-specified; TVUS - presence of round, intra-ovarian homogenous, hypoechoic "tissue," with a

Guerriero 1996a *(Continued)*

	clear demarcation from the parenchyma and without papillary proliferations - referenced to the primary source, pre-specified **Examiners** For blood assays - no information provided, unclear if were blinded to the results of reference standard; for TVUS - same physician blinded to reference standard **Interobserver variability** Intra- and inter-assay CV for CA-125 3.9% and 4.2%; for CA-19.9 4.6% and 5.3%
Target condition and reference standard(s)	**Target condition** Ovarian endometriosis **Prevalence of target condition in the sample** n = 39/118 (33%): all stage III-IV; controls - 79 **Reference standard** Laparoscopy n = 99/ Laparotomy n = 19 (n = 118, 100%) + histology **Description of positive case definition by reference standard test as reported** Visual inspection with careful assessment of the ovaries, followed by histopathological diagnosis; surgical staging according to the rAFS classification **Examiners** No information provided
Flow and timing	**Time interval between index test and reference standard** Blood was collected on the day of surgery, TVUS was performed several days prior **Withdrawals** None
Comparative	
Key conclusions by the authors	Transvaginal ultrasonography used alone is the most cost-effective method in the preoperative differential diagnosis of endometrioma
Conflict of interest	Not reported
Notes	The reported diagnostic estimates for blood biomarkers or TVUS alone are presented in other reviews from this series

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Was a 'two-gate' design avoided?	Yes		
		Low	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it prespecified?	No		

Guerriero 1996a *(Continued)*

Was a cycle phase considered in interpretation of the result of index test?	Yes		
Did the study provide a clear prespecified definition of what was considered to be a positive result of index test?	Yes		
Was the index test performed by a single operator or interpreted by consensus in a joint session?	Yes		
Were the same clinical data available when the index test results were interpreted as would be available when the test is used in practice?	Unclear		
		High	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Guerriero 1996b

Study characteristics	
Patient sampling	**Primary objectives** To assess the role of transvaginal ultrasonography in combination with CA-125 plasma levels in diagnosis of endometrioma **Study population** Women who were submitted to laparoscopy or laparotomy because of the presence of a persistent adnexal mass

Guerriero 1996b *(Continued)*

	Selection criteria Inclusion criteria: premenopausal, non-pregnant women **Study Design** Cross-sectional, single-gate design, prospective recruitment and collection of samples, consecutive series
Patient characteristics and setting	**Clinical presentation** Adnexal mass **Age:** Age range 20-49 years, mean age not provided **Number of enrolled** 101 women **Number of available for analysis** 101 women (only moderate-severe endometriosis included; all in follicular cycle phase) **Setting** University Hospital, University of Cagliari **Place of study** Cagliari, Italy **Period of study** November 1993 - October 1994 **Language** English
Index tests	**Index test** CA-125 in serum + Transvaginal Ultrasonography (TVUS) **Details of the index test procedure as stated** Serum CA-125 levels assessed by immunoradiometric assay (CIS Bio International, Gif sur Yvette, France), limit of detection 0.5 U/ml; TVUS performed with a 5 MHz endovaginal probe (Acuson XP/10 OB ultrasound system), procedure described in details **Threshold for positive result** CA-125: ≥20 IU/ml ≥25 IU/ml, ≥35 IU/ml pre-specified; TVUS - presence of round, homogenous, hypoechoic "tissue," within the ovary - referenced to the primary source, pre-specified **Examiners** For blood assays - no information provided; unclear if were blinded to the results of reference standard; for TVUS - same physician blinded to reference standard **Interobserver variability** Intra- and inter-assay CV for CA-125 3.9% and 4.2%
Target condition and reference standard(s)	**Target condition** Ovarian endometriosis **Prevalence of target condition in the sample** n = 29/101 (29%): all stage III-IV; controls n = 72 **Reference standard** Laparoscopy/ Laparotomy (number for each group not reported) + histopathology **Description of positive case definition by reference test as reported:** Visual inspection confirmed on histopathology; histological criteria reported; surgical procedure described; surgical staging according to the rAFS classification **Examiners** No information provided
Flow and timing	**Time interval between index test and reference standard** Blood was collected on the day of surgery, TVUS was performed several days prior **Withdrawals** None
Comparative	
Key conclusions by the authors	Transvaginal ultrasonography used alone has a better predictive capacity in differentiating endometrioma from other adnexal masses than combined methods.
Conflict of interest	Not reported

Guerriero 1996b *(Continued)*

Notes	The reported diagnostic estimates for blood biomarker or TVUS alone are presented in other reviews from this series

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Was a 'two-gate' design avoided?	Yes		
		Low	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it prespecified?	Yes		
Was a cycle phase considered in interpretation of the result of index test?	Yes		
Did the study provide a clear prespecified definition of what was considered to be a positive result of index test?	Yes		
Was the index test performed by a single operator or interpreted by consensus in a joint session?	Yes		
Were the same clinical data available when the index test results were interpreted as would be available when the test is used in practice?	Unclear		
		Unclear	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Guerriero 1996b *(Continued)*

		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Hudelist 2009

Study characteristics	
Patient sampling	**Primary objectives** To evaluate the accuracy of routine clinical examination (per vaginam, PV) combined with transvaginal sonography (TVS) for presurgical, non-invasive diagnosis of endometriosis **Study population** Women with suspected endometriosis, who were either referred to the pelvic pain clinic or self-referred **Selection criteria** Inclusion criteria: availability of the complete past medical, social, obstetrical and gynaecological history and the woman's consent; exclusion criteria: history of gynaecological cancer, previous surgery for DIE involving rectal surgery or dissection of the POD or RVS or other disease entities including resection of the bladder or anterior rectal wall and virginity of the woman **Study design** Cross-sectional, single-gate design, prospective recruitment and collection of samples, consecutive series
Patient characteristics and setting	**Clinical presentation** Suspected endometriosis: dysmenorrhoea - 77.5%, dyspareunia - 34.5%, chronic pelvic pain - 22%, dyschezia - 14.5% or subfertility - 17% **Age** Median age 33 years, range 16–45 years **Number of enrolled** 223 women **Number of available for analysis** 200 women (cycle phase not reported) **Setting** Tertiary referral service Villach General Hospital (endometriosis centre) **Place of study** Villach, Austria; Worthing and Chertsey, UK **Period of study** September 2007 - June 2008 **Language** English
Index tests	**Index test** Clinical examination + TVUS **Details of the index test procedure as stated** Clinical examination included speculum and bimanual PV examination; TVUS performed with either a Logic 9 (GE) or a Accuvix XQ (Accuvix) scanner using a 5–9 MHz endovaginal transducers, procedure described in details **Description of positive case definition by index test as reported** Clinical examination - a palpable nodularity or stiffened or thickened area or a palpable cystic expansion with topographic-anatomical

Hudelist 2009 *(Continued)*

	correlation to the following sites: left and right USLs, vagina, RVS, POD, the rectosigmoid and the urinary bladder (posterior wall); TVUS - endometrioma: presence of a cyst or multiple cysts containing diffuse low-level echoes, DIE: regular or irregular hypoechogenic nodular structure, cystic mass or hypoechogenic linear thickening with regular or irregular margins, described for each site (USL, vaginal wall, RVS, bladder, rectosigmoid colon); POD complete obliteration: uterus, adnexa and rectosigmoid colon adherent, with disappearance of the peritoneal structure, POD incomplete obliteration: peritoneal limits partially identified with the presence or absence of fluid collection **Examiners:** All combined PV examination and TVS (PV followed by TVS) were performed by one of the three examiners with extensive experience in TVS for DIE (> five years) **Interobserver variability:** Not provided
Target condition and reference standard(s)	**Target condition:** DIE - separate anatomical sites; ovarian endometriosis **Prevalence of target condition in the sample** Pelvic endometriosis n = 135/200 (67.5%); DIE n = 64/200 (32%); ovarian endometriosis n = 49/200 (24.5%) **Reference standard** Laparoscopy n = 200 (100%) + histopathology **Description of positive case definition by reference test as reported** Visual inspection confirmed on histopathology, histological criteria specified and referenced to primary source; surgical procedure described in details **Examiners:** No information provided; surgery performed in endometriosis referral centre
Flow and timing	**Time interval between index test and reference standard:** Both tests were performed within two months before surgery **Withdrawals:** 23 women were excluded because they did not meet the inclusion criteria
Comparative	
Key conclusions by the authors	The combination of PV and TVS accurately predicts the presence of endometriosis affecting the ovaries, vagina, rectum, USL, RVS and POD in women with suspected endometriosis. We suggest the routine combination of PV and TVS as an essential part of the standard primary assessment of pelvic pain women with suspected endometriosis
Conflict of interest	Not reported
Notes	The accuracy estimates for endometrioma and USL are reported by the authors but not presented in the review, because this was a lesion-specific analysis The reported diagnostic performance of vaginal examination alone is not included in this review The diagnostic estimates for bladder endometriosis are reported but not included in the review

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		

Hudelist 2009 *(Continued)*

Was a 'two-gate' design avoided?	Yes		
		Low	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was a cycle phase considered in interpretation of the result of index test?	Unclear		
Did the study provide a clear pre-specified definition of what was considered to be a positive result of index test?	Yes		
Was the index test performed by a single operator or interpreted by consensus in a joint session?	No		
Were the same clinical data available when the index test results were interpreted as would be available when the test is used in practice?	No		
		High	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		

Hudelist 2009 *(Continued)*

Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
	Low

Koninckx 1996

Study characteristics	
Patient sampling	**Primary objectives** To evaluate clinical examination during menstruation and plasma CA-125 concentration to diagnose endometriosis **Study population** Women scheduled for laparoscopy for suspected endometriosis **Selection criteria** Exclusion criteria: hormonal treatment or medical treatment for endometriosis in the three months preceding laparoscopy, refusal for a clinical examination during menstruation (only DIE considered) **Study design** Cross-sectional single-gate, prospective recruitment and collection of samples, consecutive series
Patient characteristics and setting	**Clinical presentation** Infertility (n = 33), pain (n = 13), infertility + pain (n = 6), hydrosalpinx (n = 1), ovarian cyst (n= 2) **Age** Range 20 - 45 years (personal communication with the author) **Number of participants enrolled** 61 women **Number of participants available for analysis** 55 women (only DIE or endometrioma or severe pelvic adhesions included; all in menstrual, follicular and early luteal phase of menstrual cycle) **Setting** Division of endoscopic surgery, University Hospital Gasthiusberg, University of Leuven **Place of study** Leuven, Belgium **Period of study** Not stated **Language** English
Index tests	**Index test** Clinical examination in menstrual phase (menstrual nodularities) + CA-125 in mid follicular phase **Details of the index test procedure as stated** Clinical examination included pelvic bimanual examination with careful assessment of pelvic nodularities and their tenderness; CA-125 assay using a second generation IRMA kit (CA-125 II, Centocor, Malvern, Pa; all kits from the same production batch) **Threshold for positive result** Clinical examination - presence of induration (with one or more small nodules) or painful nodularities (larger spherical nodule), pre-specified and described in details; CA-125 >35 U/ml, not pre-specified **Examiners** Clinical examination: one of the two authors, always preoperatively and hence blinded to the results of reference standard; CA-125 - no information provided; unclear if were blinded to reference standard **Interobserver variability** CA-125: intra-and inter-assay variation < 5% and < 8%

Koninckx 1996 *(Continued)*

Target condition and reference standard(s)	**Target condition** DIE, ovarian endometrioma or severe pelvic adhesions **Prevalence of target condition in the sample** n = 38/55 (69%): stage I-II 29, stage III-IV 9; DIE 13, endometrioma 9, deep endometriosis + severe cul de sac adhesions + endometrioma 24; controls n = 17) **Reference standard** Laparoscopy n = 55 (100%) **Description of positive case definition by reference standard test as reported** Visual inspection, deep endometriosis classified as type I - III, reference to the source with diagnostic criteria and described; staging according to the rAFS classification . **Examiners** Not stated; unclear if were blinded to the result of pelvic examination
Flow and timing	**Time interval between index test and reference standard** The tests were performed within four months before surgery (personal communication with the author) **Withdrawals** In six women (11%) the surgery was cancelled for various reasons
Comparative	
Key conclusions by the authors	Clinical examination during menstruation can diagnose reliably deep endometriosis, cystic ovarian endometriosis or cul de sac adhesions. This test, preferentially combined with a follicular phase CA-125 assay, should be used to decide whether a preparation for bowel surgery should be given
Conflict of interest	Not reported
Notes	The reported diagnostic estimates for blood biomarker alone are presented in other reviews from this series The presented diagnostic estimates are for DIE or ovarian endometrioma or severe cul de sac adhesions

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Was a 'two-gate' design avoided?	Yes		
		Low	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		

Koninckx 1996 *(Continued)*

Was a cycle phase considered in interpretation of the result of index test?	Yes		
Did the study provide a clear pre-specified definition of what was considered to be a positive result of index test?			
Was the index test performed by a single operator or interpreted by consensus in a joint session?			
Were the same clinical data available when the index test results were interpreted as would be available when the test is used in practice?			
		High	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
		High	

Marasinghe 2014

Study characteristics	
Patient sampling	**Primary objectives** To compare the performance of history and examination findings combined with transvaginal ultrasound 'soft marker' evaluation of ovarian mobility, for the prediction of fixed ovaries secondary to endometriosis at laparoscopy

Marasinghe 2014 *(Continued)*

	Study population Women scheduled for laparoscopy for chronic pelvic pain and or subfertility **Selection criteria** Exclusion criteria: previous surgical diagnosis of endometriosis or pelvic adhesions, confirmed diagnosis of genital abnormalities, those who did not consent to vaginal examinations **Study design** Cross-sectional single-gate, prospective recruitment and collection of samples, consecutive series
Patient characteristics and setting	**Clinical presentation** Infertility 83%, chronic pelvic pain 17%, dysmenorrhoea 46%, dyspareunia 31% **Age** Mean age 33.3 ± 5.1 years, range 22 - 48 years **Number of participants enrolled** 110 women **Number of participants available for analysis** 106 women **Setting** University Gynecology unit, National Hospital of Sri Lanka, tertiary referral centre **Place of study** Colombo, Sri Lanka **Period of study** December 2009 - March 2010 **Language** English
Index tests	**Index test** Clinical history + examination + TVUS **Details of the index test procedure as stated** Clinical history - interview with history of dysmenorrhoea and dyspareunia, severity was assessed using a visual analogue scale ranging 1-10 with a score of one considered as 'no pain'; examination included pelvic bimanual examination to detect the presence of pelvic tenderness, a fixed retroverted uterus, tender uterosacral ligaments and deeply infiltrating nodules on the uterosacral ligaments or in the cul-de-sac; TVUS - Assessment was with gentle pressure on the ovary with the transvaginal probe, fixed or mobile ovaries were diagnosed by assessing the ovaries in relation to the uterus and ipsilateral internal iliac vessels **Threshold for positive result** Clinical history - dysmenorrhoea or dyspareunia; examination - at least one of the abovementioned examination features; TVUS: at least one 'fixed' ovary **Examiners** Clinical history and examination: one clinician with > four years experience in gynaecology; TVUS - single examiner with > four years experience in ultrasound blinded to the clinical data **Interobserver variability** Each tests was performed by a single operator
Target condition and reference standard(s)	**Target condition** Endometriosis and other peri-ovarian adhesions **Prevalence of target condition in the sample** n = 32/106 (30%): stage I-II 19, stage III-IV 13; other peri-ovarian adhesions n = 5; controls n = 17) **Reference standard** Laparoscopy n = 106 (100%) ± histology **Description of positive case definition by reference standard test as reported** Visual inspection, endometriotic adhesions causing a fixed ovary - defined as ovaries fixed to the uterus or ovarian fossa and when it could not be elevated from the ovarian fossa by using a blunt probe; histology was considered in selected cases, histological criteria provided; staging according to the rAFS classification . **Examiners** Single examiner with 15 years of laparoscopic experience, blinded to the results of index test
Flow and timing	**Time interval between index test and reference standard** The tests were performed within two weeks before surgery

Marasinghe 2014 *(Continued)*

	Withdrawals In four women at least one ovary could not be visualised on transvaginal scanning and they were excluded
Comparative	
Key conclusions by the authors	A combination of clinical and transvaginal ultrasound based 'soft marker' of ovarian mobility provides a valid method for identifying fixed ovaries secondary to endometriosis
Conflict of interest	Not reported
Notes	The primary outcome of the study was endometriosis with fixed ovaries or other peri-ovarian adhesions

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Was a 'two-gate' design avoided?	Yes		
		Low	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was a cycle phase considered in interpretation of the result of index test?			
Did the study provide a clear pre-specified definition of what was considered to be a positive result of index test?	Yes		
Was the index test performed by a single operator or interpreted by consensus in a joint session?	Yes		
Were the same clinical data available when the index test results were interpreted as	Yes		

Marasinghe 2014 *(Continued)*

would be available when the test is used in practice?			
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
		High	

Paiva 2014

Study characteristics	
Patient sampling	**Primary objectives** To develop a test to discriminate between women suffering from pelvic pain associated with presence or absence of endometriosis, using symptom visual analogue scale (VAS) scores, demographic and lifestyle factors and known and novel plasma biomarkers **Study population** Women undergoing laparoscopy for evaluation of chronic pelvic pain, dysmenorrhoea, or dyspareunia **Selection criteria** Exclusion criteria: women on current hormonal therapy, failure to complete questionnaire, **Study design** Cross-sectional single-gate, prospective collection of samples
Patient characteristics and setting	**Clinical presentation** Pelvic pain, dysmenorrhoea, dyspareunia **Age** Mean age 27 years, range 18 - 44 years (endometriosis group) and 30 years, range 19 - 43 years (controls) **Number of participants enrolled** 172 women **Number of participants available for analysis** 101 women (in menstrual, proliferative or secretory cycle phase)

Paiva 2014 *(Continued)*

	Setting Department of O&G, Royal Women's Hospital, University of Melbourne **Place of study** Melbourne, Australia **Period of study** May 2006 - February 2009 **Language** English
Index tests	**Index test** Predictive model including: clinical history (parity, ever had IUD, hx of endometriosis, alcohol intake, dyspareunia) + serum CA-125 **Details of the index test procedure as stated** Clinical history obtained with a questionnaire including demographic details, LMP, risk factors for endometriosis and symptom severity on form of visual analogue scale (VAS); serum CA-125 was measured using 2-plex magnetic human circulating cancer biomarker panel kit (Millipore, USA), detection limit 0.26 pg/ml; laboratory methods and sample processing described; predictive model was constructed using multivariate logistic regression model (initial selection of 14 parameters) **Threshold for positive result** Not reported **Examiners** No information provided; unclear if blinded to the results of reference standard **Interobserver variability** CA-125: the intra- and inter-assay CV < 10%
Target condition and reference standard(s)	**Target condition** Endometriosis **Prevalence of target condition in the sample** n = 69/101 (68%): stage I-II 45, stage III-IV 24; controls n = 32 **Reference standard** Laparoscopy n = 101 (100%) + histopathology **Description of positive case definition by reference standard test as reported** Visual inspection confirmed by histological demonstration of endometrial glands and stroma; staging according to the rASRM classification **Examiners** No information provided
Flow and timing	**Time interval between index test and reference standard** Blood samples and questionnaire were obtained preoperatively on the same day **Withdrawals** 71 participants were excluded: 16 due to current hormone treatment, 31 - not completed questionnaire, 24 - no samples available due to laboratory freezer failure
Comparative	
Key conclusions by the authors	Combining symptom scores, historical measures and CA-125 provides a reasonable means to discriminate between women with pelvic pain associated with presence or absence of endometriosis, but greater specificity is needed before such a model could replace laparoscopy
Conflict of interest	The authors declared no conflict of interests; the study was supported by several research grants
Notes	The predictive model included history of endometriosis, hence can be used for primary diagnosis as well as for assessment of recurrence

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			

Paiva 2014 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
Was a 'two-gate' design avoided?	Yes		
		Unclear	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Was a cycle phase considered in interpretation of the result of index test?	Yes		
Did the study provide a clear pre-specified definition of what was considered to be a positive result of index test?	Yes		
Was the index test performed by a single operator or interpreted by consensus in a joint session?			
Were the same clinical data available when the index test results were interpreted as would be available when the test is used in practice?			
		High	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			

Paiva 2014 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
	Low

Yun 2014

Study characteristics	
Patient sampling	**Primary objectives** To validate and investigate the clinical value of urinary enolase I (NNE) in women with endometriosis. **Study population** Women who underwent laparoscopy for diagnostic evaluation of pelvic masses, pelvic pain, suspicious endometriosis and infertility **Selection criteria** Exclusion criteria: postmenopausal status, previous use of hormone or gonadotropin-releasing hormone agonist, adenomyosis, endometrial cancer or hyperplasia, endometrial polyps, infectious diseases, chronic or acute inflammatory diseases, malignancy, autoimmune disease and cardiovascular disease **Study design** Cross-sectional, single-gate design, prospective collection of samples
Patient characteristics and setting	**Clinical presentation:** Pelvic masses, pelvic pain, suspicious endometriosis and infertility **Age:** Mean age 31.48 ± 6.30 for endometriosis group, 29.35 ± 6.87 for control group **Number of enrolled:** 59 women **Number of available for analysis:** 59 women (in follicular or luteal cycle phase, numbers not specified; only moderate/severe endometriosis) **Setting:** Gongnam Severance Hospital, Yonsei University College of Medicine **Place of study:** Seoul, Republic of Korea **Period of study:** January 2009 - December 2011 **Language:** English
Index tests	**Index test** Urinary enolase I (NNE-Cr) + serum CA-125 **Description of positive case definition by index test as reported** Urinary enolase I was measured with commercial ELISA according to the manufacturer's protocols (USCN Life Science & Technology Company, TX) with minimal detectable concentration of 0.312 ng/ml; urinary NNE values were normalised to urinary Cr concentrations; serum CA-125 was measured using CA-125 II electro chemiluminescence immunoassay with Roche/Hitachi Modular Analytics E170 system (Roche Diagnostics, Japan); sample handling described **Threshold for positive result** Cut-off value > 27.23, not pre-specified **Examiners:** No information provided, unclear if were blinded to the results of reference standard

Yun 2014 *(Continued)*

	Interobserver variability Not provided
Target condition and reference standard(s)	**Target condition** Endometriosis **Prevalence of target condition in the sample** n = 39/59 (55%): all stage III-IV; controls n = 20 **Reference standard:** Laparoscopy and histology **Description of positive case definition by reference test as reported** Visual inspection confirmed on histopathology; staging according to the rASRM classification **Examiners** Not information provided
Flow and timing	**Time interval between index test and reference standard** Blood was collected preoperatively, urine was collected after induction of anaesthesia **Withdrawals** No withdrawals reported
Comparative	
Key conclusions by the authors	Urinary enolase I was significantly increased in women with endometriosis, though the findings undermine its capacity as a diagnostic marker. May have potential as one of the combined markers.
Conflict of interest	The authors declared no conflict of interest; supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0023323)
Notes	The reported diagnostic estimates for urine or blood test only are presented in other reviews from this series Only information for severe disease available

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
Was a 'two-gate' design avoided?	Yes		
		Unclear	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it prespecified?	No		

Yun 2014 *(Continued)*

Was a cycle phase considered in interpretation of the result of index test?	Yes		
Did the study provide a clear prespecified definition of what was considered to be a positive result of index test?			
Was the index test performed by a single operator or interpreted by consensus in a joint session?			
Were the same clinical data available when the index test results were interpreted as would be available when the test is used in practice?			
		High	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Zeng 2005

Study characteristics	
Patient sampling	**Primary objectives** To evaluate the diagnostic value of examining endometrial biopsy specimens for aromatase cytochrome P450 and CA-125 for endometriosis

Zeng 2005 *(Continued)*

	Study population Women undergoing laparoscopy or laparotomy for pelvic pain or infertility **Selection criteria** Inclusion criteria: reproductive age regular menstrual cycle; exclusion criteria: hormonal treatment for 3/12 prior reproductive age, preoperative diagnosis of uterine fibroids, adenomyosis. **Study design** Cross-sectional single-gate, prospective collection of samples
Patient characteristics and setting	**Clinical presentation** Infertility or pelvic pain **Age** Mean age 33 ± 4 years, range 26 - 40 years (endometriosis), 32 ± 4 years, range 25-39 years (controls) **Number of participants enrolled** 58 women **Number of participants available for analysis** 58 women (31 women in follicular and 27 women in luteal cycle phase) **Setting** Department of O&G, Third Xiangya Hospital, Central South University **Place of study** Changsha, China **Period of study** March 2003 - February 2004 **Language** Chinese
Index tests	**Index test** CA-125 in serum + P450 aromatase in endometrium **Details of the index test procedure as stated** Serum CA-125 was determined by chemiluminescence assay; endometrial aromatase protein was evaluated by immunohistochemistry, 2nd generation assay (ElivisionTM plus kit), positive IHC staining indicated by presence of brown particles within the cytoplasm; sample handling or laboratory technique not described **Threshold for positive result** CA-125 >35 U/ml, not pre-specified; P450 aromatase: positive or negative test **Examiners** No information provided, unclear if blinded to the result of reference standard **Interobserver variability** Not stated
Target condition and reference standard(s)	**Target condition** Endometriosis **Prevalence of target condition in the sample** n = 36/58 (62%): stage I-II 20, stage III-IV 16; controls n = 22 **Reference standard** Laparoscopy/Laparotomy n = 58 (100%) **Description of positive case definition by reference standard test as reported** Visual inspection; staging according to rAFS classification **Examiners** Not stated
Flow and timing	**Time interval between index test and reference standard** Blood and endometrial samples were collected on the day of surgery **Withdrawals** None
Comparative	
Key conclusions by the authors	The combination assay of aromatase cytochrome P450 in eutopic endometrium and CA-125 can be used as a diagnostic test for endometriosis, especially for the early stage of endometriosis, which is superior to the assay of CA-125

Zeng 2005 *(Continued)*

Conflict of interest	Not reported
Notes	Translated from Chinese, some information may have been misinterpreted The reported diagnostic estimates for endometrial or blood markers alone are presented in other reviews from this series

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
Was a 'two-gate' design avoided?	Yes		
		Unclear	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it prespecified?	No		
Was a cycle phase considered in interpretation of the result of index test?	Unclear		
Did the study provide a clear pre-specified definition of what was considered to be a positive result of index test?			
Was the index test performed by a single operator or interpreted by consensus in a joint session?			
Were the same clinical data available when the index test results were interpreted as would be available when the test is used in practice?			
		High	**Unclear**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		

Zeng 2005 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

(r)AFS: (revised) American Fertility Society; **(r)ASRM**: (revised) American Society for Reproductive Medicine; **CA-125**: cancer antigen; **CV**: coefficient of variation; **DIE**: deep infiltrating endometriosis; **ECD**: Electron Coupled Dye; **ELISA**: enzyme-linked immunosorbent assay; **FITC**: fluorescein isothiocyanate; **GnRH**: gonadotropin-releasing hormone; **IL**: interleukin; **IUD**: intrauterine device; **LMP**: last menstrual period; **NNE**: non neuronal enolase;**PE**: phycoerythrin; **PerCP**: peridinin; chlorophyll protein; **PGP**: permeability glycoprotein; **POD**: pouch of Douglas; **PV**: per vaginam; **RVS**: rectovaginal septum; **TVS**: transvaginal sonography; **TVUS**: transvaginal ultrasound; **USL**: uterosacral ligament; **VAS**: visual analogue scale; **VDBP**: vitamin-D-binding protein; **VDBPCr**: VDBP level corrected for creatinine.

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Abrao 2007	Index test outside inclusion criteria (no combined diagnostic estimates available)
Adamyan 1993	Index test outside inclusion criteria (no combined diagnostic estimates available)
Alcazar 2011	Index test outside inclusion criteria (no combined diagnostic estimates available)
Badawy 1984	Index test outside inclusion criteria (no combined diagnostic estimates available)
Bazot 2009	Index test outside inclusion criteria (no combined diagnostic estimates available)
Borboletto 1995	Review article
Cho 2007	Index test outside inclusion criteria (no combined diagnostic estimates available)
da Silva 2014	Index test outside inclusion criteria (no combined diagnostic estimates available)
Dias 2012	Index test outside inclusion criteria (no combined diagnostic estimates available)
Eskenazi 2001	Index test outside inclusion criteria (no combined diagnostic estimates available)
Fedele 1988	Index test outside inclusion criteria (no combined diagnostic estimates available)
Guerriero 1997	Index test outside inclusion criteria (only 'lesion-level' analysis for imaging component of the test)

Study	Reason for exclusion
Hudelist 2011b	Index test outside inclusion criteria (no combined diagnostic estimates available)
Kocbek 2014	Index test outside inclusion criteria (no combined diagnostic estimates available)
Kuessel 2014	Index test outside inclusion criteria (no combined diagnostic estimates available)
Lee 2014	Index test outside inclusion criteria (no combined diagnostic estimates available)
Nezhat 1994	Index test outside inclusion criteria (no combined diagnostic estimates available)
Szubert 2014	Insufficient diagnostic test accuracy information (unable to construct 2 x 2 tables)
Weerakiet 2000	Study design, population and index test outside inclusion criteria (retrospective selection of cases; postmenopausal women were included; only 'lesion-level' analysis)
Wolfler 2005	Insufficient diagnostic test accuracy information (unable to construct 2 x 2 tables)
Yong 2013	Target condition outside inclusion criteria (not endometriosis but "abnormal laparoscopy in women with pelvic pain")

DATA

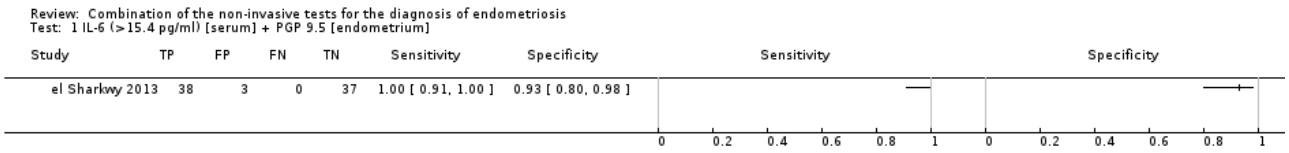
Presented below are all the data for all of the tests entered into the review.

Table Tests. Data tables by test

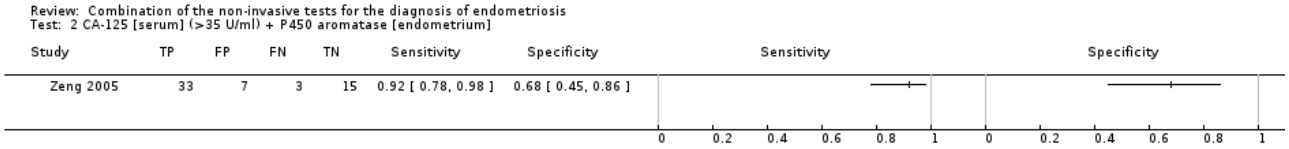
Test	No. of studies	No. of participants
1 IL-6 (>15.4 pg/ml) [serum] + PGP 9.5 [endometrium]	1	78
2 CA-125 [serum] (>35 U/ml) + P450 aromatase [endometrium]	1	58
3 VDBP-Cr [urine] x CA-125 [serum] (>2755)	1	95
4 NNE_Cr [urine] + CA-125 [serum] (>27.23)	1	59
5 Hx (dysmenorrhoea, dyspareunia) + PV examination + TVUS (fixed ovary)	1	106
6 Hx (length of menses) + CA-125 [serum] (>35 U/ml) + leukocytes [endometrium]	1	368
7 Hx (parity, past IUD, past endometriosis, alcohol intake, dyspareunia) + CA-125 [serum]	1	101
8 PV examination (menstrual nodularities) + CA-125 [serum] (>35 IU/ml) for DIE, endometrioma or severe adhesions	1	41
9 PV examination (menstrual nodularities) OR CA-125 [serum] (>35 IU/ml) for DIE, endometrioma or severe adhesions	1	41

Test	No. of studies	No. of participants
10 PV examination (menstrual nodularities) + CA-125 [serum] (>35 IU/ml) for DIE	1	30
11 PV examination (menstrual nodularities) OR CA-125 [serum] (>35 IU/ml) for DIE	1	30
12 PV examination (menstrual nodularities) + CA-125 [serum] (>35 IU/ml) for endometrioma	1	26
13 PV examination (menstrual nodularities) OR CA-125 [serum] (>35 IU/ml) for endometrioma	1	26
14 TVUS + CA-125 [serum] (≥25 U/ml) + CA-19.9 [serum] (≥12 U/ml)	1	118
15 TVUS + (CA-125 [serum] (≥25 U/ml) OR CA-19.9 [serum] (≥12 U/ml))	1	118
16 TVUS + CA-19.9 [serum] (≥12 U/ml)	1	118
17 TVUS OR CA-19.9 [serum] (≥12 U/ml)	1	118
18 TVUS + CA-125 [serum] (≥20 U/ml)	1	101
19 TVUS OR CA-125 [serum] (≥20 U/ml)	1	101
20 TVUS + CA-125 [serum] (≥25 U/ml)	1	101
21 TVUS OR CA-125 [serum] (≥25 U/ml)	1	101
22 TVUS + CA-125 [serum] (≥35 U/ml)	1	101
23 TVUS OR CA-125 [serum] (≥35 U/ml)	1	101
24 PV examination + TVUS for POD obliteration	1	200
25 PV examination + TVUS for vaginal endometriosis	1	200
26 PV examination + TVUS for RVS endometriosis	1	200
27 PV examination + TVUS for rectal endometriosis	1	200

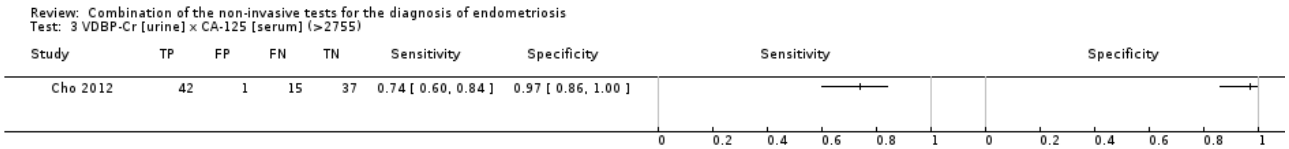
Test 1. IL-6 (>15.4 pg/ml) [serum] + PGP 9.5 [endometrium].



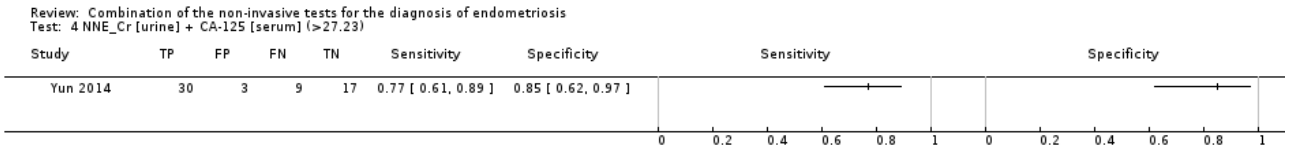
Test 2. CA-125 [serum] (>35 U/ml) + P450 aromatase [endometrium].



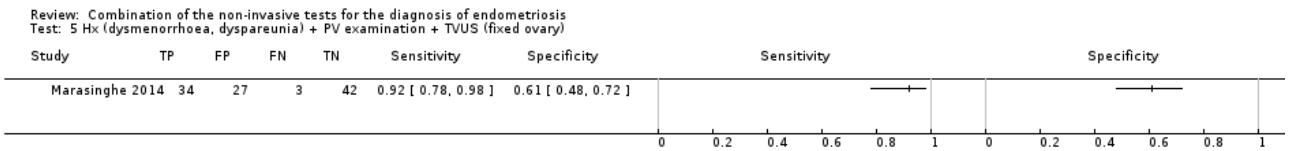
Test 3. VDBP-Cr [urine] x CA-125 [serum] (>2755).



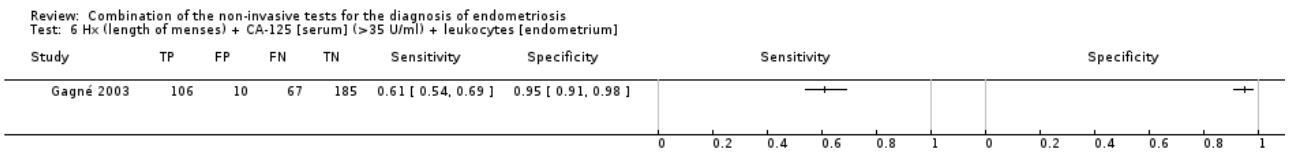
Test 4. NNE_Cr [urine] + CA-125 [serum] (>27.23).



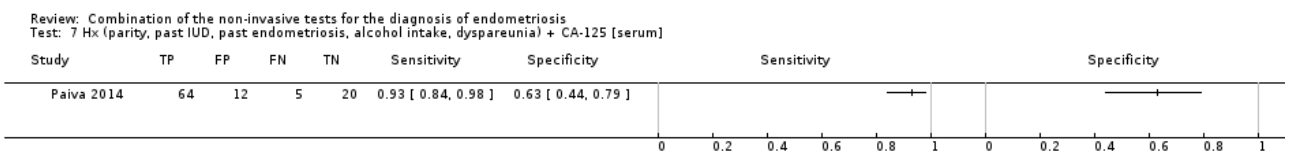
Test 5. Hx (dysmenorrhoea, dyspareunia) + PV examination + TVUS (fixed ovary).



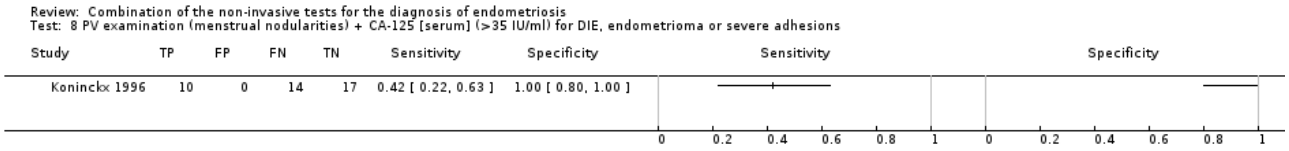
Test 6. Hx (length of menses) + CA-125 [serum] (>35 U/ml) + leukocytes [endometrium].



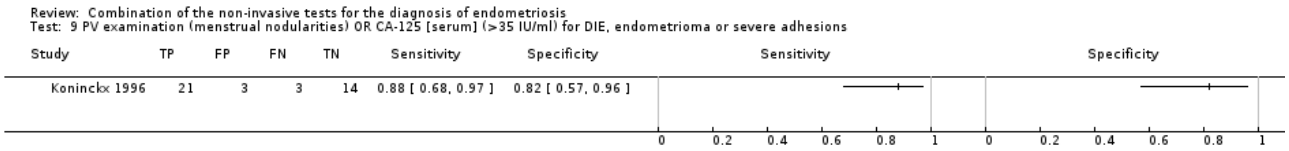
Test 7. Hx (parity, past IUD, past endometriosis, alcohol intake, dyspareunia) + CA-125 [serum].



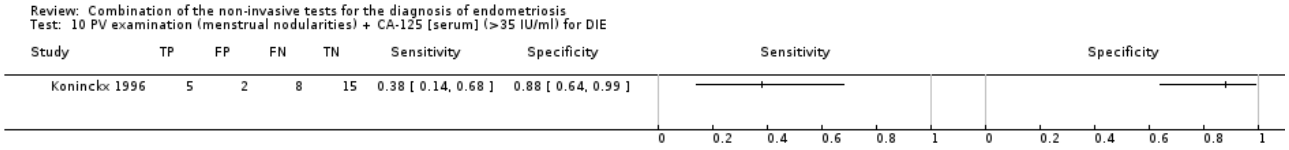
Test 8. PV examination (menstrual nodularities) + CA-125 [serum] (>35 IU/ml) for DIE, endometrioma or severe adhesions.



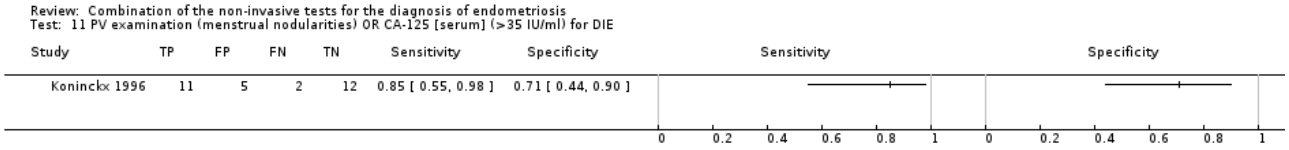
Test 9. PV examination (menstrual nodularities) OR CA-125 [serum] (>35 IU/ml) for DIE, endometrioma or severe adhesions.



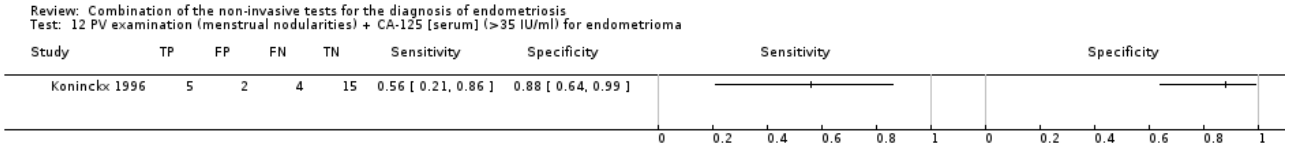
Test 10. PV examination (menstrual nodularities) + CA-125 [serum] (>35 IU/ml) for DIE.



Test 11. PV examination (menstrual nodularities) OR CA-125 [serum] (>35 IU/ml) for DIE.



Test 12. PV examination (menstrual nodularities) + CA-125 [serum] (>35 IU/ml) for endometrioma.



Test 13. PV examination (menstrual nodularities) OR CA-125 [serum] (>35 IU/ml) for endometrioma.

Review: Combination of the non-invasive tests for the diagnosis of endometriosis
Test: 13 PV examination (menstrual nodularities) OR CA-125 [serum] (>35 IU/ml) for endometrioma

Study	TP	FP	FN	TN	Sensitivity	Specificity
Koninckx 1996	8	6	1	11	0.89 [0.52, 1.00]	0.65 [0.38, 0.86]

Test 14. TVUS + CA-125 [serum] (≥25 U/ml) + CA-19.9 [serum] (≥12 U/ml).

Review: Combination of the non-invasive tests for the diagnosis of endometriosis
Test: 14 TVUS + CA-125 [serum] (≥25 U/ml) + CA-19.9 [serum] (≥12 U/ml)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Guerriero 1996a	19	1	20	78	0.49 [0.32, 0.65]	0.99 [0.93, 1.00]

Test 15. TVUS + (CA-125 [serum] (≥25 U/ml) OR CA-19.9 [serum] (≥12 U/ml)).

Review: Combination of the non-invasive tests for the diagnosis of endometriosis
Test: 15 TVUS + (CA-125 [serum] (≥25 U/ml) OR CA-19.9 [serum] (≥12 U/ml))

Study	TP	FP	FN	TN	Sensitivity	Specificity
Guerriero 1996a	31	2	8	77	0.79 [0.64, 0.91]	0.97 [0.91, 1.00]

Test 16. TVUS + CA-19.9 [serum] (≥12 U/ml).

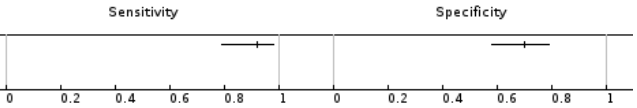
Review: Combination of the non-invasive tests for the diagnosis of endometriosis
Test: 16 TVUS + CA-19.9 [serum] (≥12 U/ml)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Guerriero 1996a	21	2	18	77	0.54 [0.37, 0.70]	0.97 [0.91, 1.00]

Test 17. TVUS OR CA-19.9 [serum] (≥12 U/ml).

Review: Combination of the non-invasive tests for the diagnosis of endometriosis
Test: 17 TVUS OR CA-19.9 [serum] (≥12 U/ml)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Guerriero 1996a	36	24	3	55	0.92 [0.79, 0.98]	0.70 [0.58, 0.79]

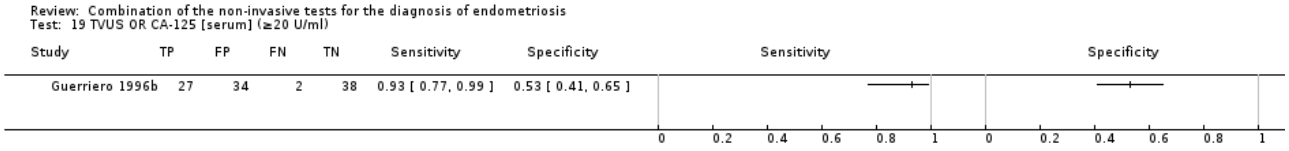


Test 18. TVUS + CA-125 [serum] (≥20 U/ml).

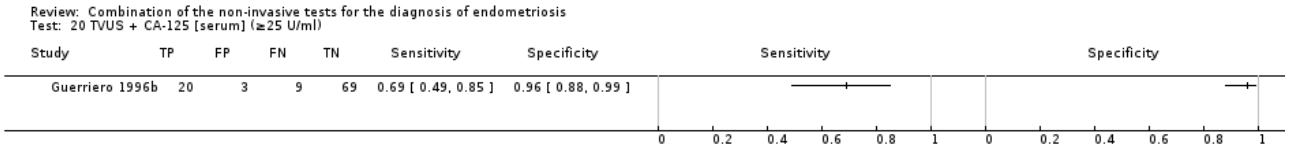
Review: Combination of the non-invasive tests for the diagnosis of endometriosis
Test: 18 TVUS + CA-125 [serum] (≥20 U/ml)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Guerriero 1996b	20	3	9	69	0.69 [0.49, 0.85]	0.96 [0.88, 0.99]

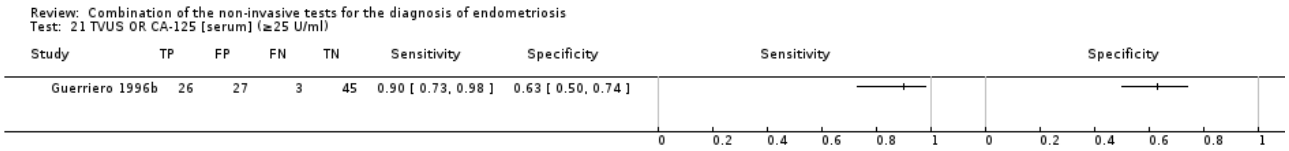
Test 19. TVUS OR CA-125 [serum] (≥20 U/ml).



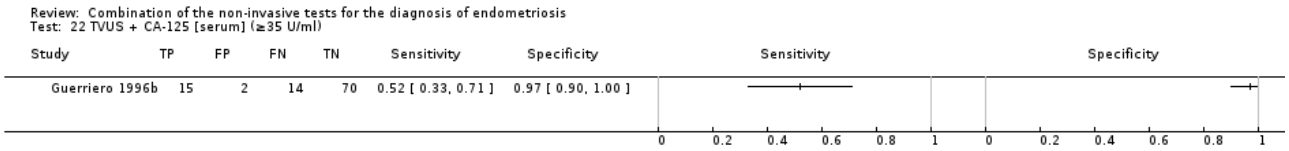
Test 20. TVUS + CA-125 [serum] (≥25 U/ml).



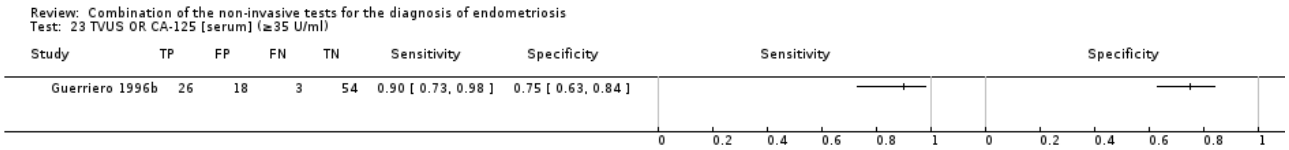
Test 21. TVUS OR CA-125 [serum] (≥25 U/ml).



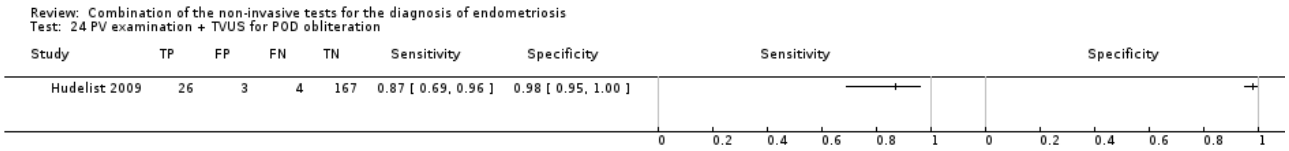
Test 22. TVUS + CA-125 [serum] (≥35 U/ml).



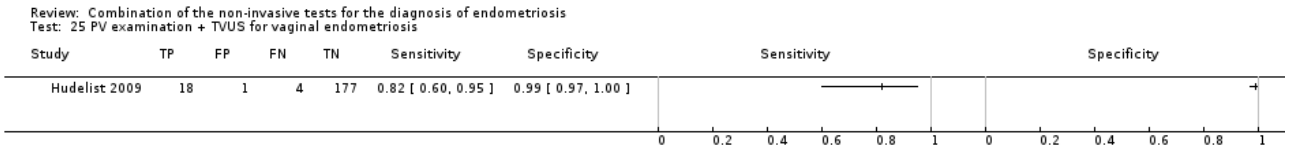
Test 23. TVUS OR CA-125 [serum] (≥35 U/ml).



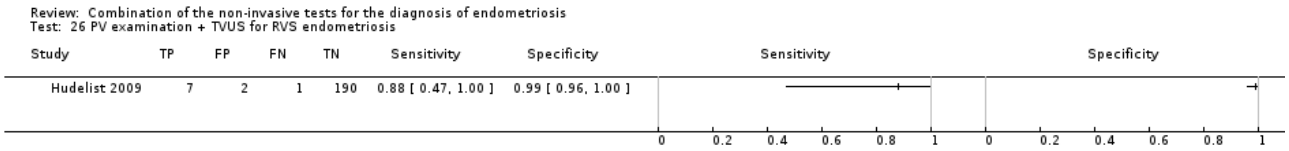
Test 24. PV examination + TVUS for POD obliteration.



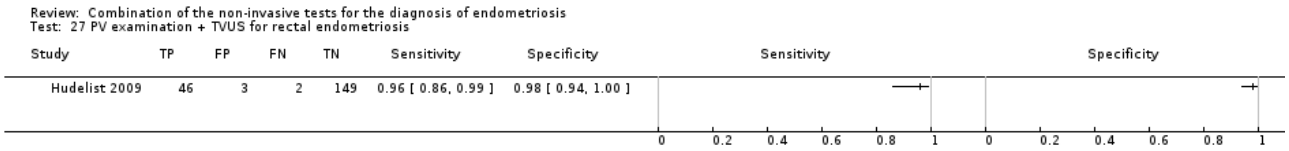
Test 25. PV examination + TVUS for vaginal endometriosis.



Test 26. PV examination + TVUS for RVS endometriosis.



Test 27. PV examination + TVUS for rectal endometriosis.



ADDITIONAL TABLES

Table 1. Staging of endometriosis, rASRM classification

Location of endometriosis	Extent	Depth		
		< 1 cm	1-3 cm	> 3 cm
Peritoneum	Superficial	1	2	4
	Deep	2	4	6
Ovary	R Superficial	1	2	4
	Deep	4	16	20
	L Superficial	1	2	4

Table 1. Staging of endometriosis, rASRM classification *(Continued)*

	Deep	4	16	20
Posterior Cul de sac Obliteration		**Partial**	**Complete**	
		4	40	
Adhesions		**< 1/3 Enclosure**	**1/3-2/3 Enclosure**	**> 2/3 Enclosure**
Ovary	R Filmy	1	2	4
	Dense	4	8	16
	L Filmy	1	2	4
	Dense	4	8	16
Tube	R Filmy	1	2	4
	Dense	4*	8*	16
	L Filmy	1	2	4
	Dense	4*	8*	16

* If the fimbriated end of the fallopian tube is completely enclosed, change the point assignment to 16

American Society for Reproductive Medicine 1997

Table 2. Combination of the non-invasive tests for endometriosis evaluated in this review

N	Test
1	IL-6 (>15.4 pg/ml) [serum] + PGP 9.5 [endometrium]
2	CA-125 [serum] (>35 U/ml) + P450 aromatase [endometrium]
3	VDBP-Cr [urine] x CA-125 [serum] (>2755)
4	NNE_Cr [urine] + CA-125 [serum] (>27.23)
5	History (dysmenorrhoea, dyspareunia) + PV examination + TVUS (fixed ovary)
6	History (length of menses) + CA-125 [serum] (>35 U/ml) + leukocytes [endometrium]
7	History (parity, past IUD, past endometriosis, alcohol intake, dyspareunia) + CA-125 [serum]
8	PV examination (menstrual nodularities) + CA125 (>35 U/ml) [serum]
9	PV examination (menstrual nodularities) OR CA125 (>35 U/ml) [serum]
10	TVUS + CA-125 [serum] (≥25 U/ml) + CA-19.9 [serum] (≥12 U/ml)
11	TVUS + (CA-125 [serum] (≥25 U/ml) OR CA-19.9 [serum] (≥12 U/ml))

Table 2. Combination of the non-invasive tests for endometriosis evaluated in this review *(Continued)*

12	TVUS + CA-19.9 [serum] (≥12 U/ml))
13	TVUS OR CA-19.9 [serum] (≥12 U/ml))
14	TVUS + CA-125 [serum] (≥20 U/ml; ≥25 U/ml; ≥35 U/ml)
15	PV examination + TVUS

CA-125: cancer antigen; **IL**: interleukin; **IUD:** intrauterine device;**NNE**: non neuronal enolase;**PV**: per vaginam; **TVUS**: transvaginal ultrasound; **VDBP**: vitamin-D-binding protein

Table 3. Risk of bias and applicability judgments for the quality assessment of diagnostic accuracy studies (QUADAS-2)

Domain 1 - Patient selection	
Description	Describe methods of patient selection and included women
Type of bias assessed	Selection bias, spectrum bias
Review Question	Women of reproductive age with clinically suspected endometriosis (symptoms, clinical examination ± presence of pelvic mass), scheduled for surgical exploration of pelvic/abdominal cavity for confirmation of the diagnosis ± treatment
Informaton collected	Study objectives, study population, selection (inclusion/ exclusion criteria), study design, clinical presentation, age, number of enrolled and number of available for analysis, setting, place and period of the study
Signalling question 1	***Was a consecutive or random sample of patients enrolled?***
Yes	If a consecutive sample or a random sample of the eligible participants was included in the study
No	If non-consecutive sample or non-random sample of the eligible participants was included in the study
Unclear	If this information was unclear
Signalling question 2	***Did the study avoid inappropriate exclusions?***
Yes	If inclusion/exclusion criteria were presented and all women with suspected endometriosis were included, with an exception for those who a) had a history of medical conditions or were on medical therapy that would have potentially interfered with interpretation of index test (e.g. malignancy, pregnancy, autoimmune disorders, infectious diseases, treatment with hormonal or immunomodulator substances); b) refused to participate in the study; or c) were unfit for surgery
No	If the study excluded the participants based on education level, psychosocial factors, genetic testing or phenotype or excluded participants with any co-morbidities commonly present in general population, including a population that could have undergone a testing for endometriosis in clinical setting (hypertension, asthma, obesity, benign gastro-intestinal or renal disease, etc)
Unclear	If the study did not provide clear definition of the selection (inclusion/exclusion) criteria and 'no' judgement was not applicable
Signalling question 3	***Was a 'two-gate' design avoided?***

Table 3. Risk of bias and applicability judgments for the quality assessment of diagnostic accuracy studies (QUADAS-2) *(Continued)*

Yes	If the study had a single set of inclusion criteria, defined by the clinical presentation (i.e. only participants in whom the target condition is suspected) - a 'single-gate design'.
No	If the study had more than one set of inclusion criteria in respect to clinical presentation (i.e. participants suspected of target condition and participants with alternative diagnosis in whom the target condition would not be suspected in clinical practice) - a 'two-gate' study design
Unclear	If it was unclear whether a 'two-gate design' was avoided or not
Risk of bias	**Could the selection of patients have introduced bias?**
High	If 'no' classification for any of the above three questions
Low	If 'yes' classification for all the above three questions
Unclear	If 'unclear' classification for three of the above questions
Concerns about applicability	**Are there concerns that the included patients do not match the review question?**
High	If the study population differed from the population defined in the review question in terms of demographic features and co-morbidity (e.g. studies with multiple sets of inclusion criteria with respect to clinical presentation including either healthy controls or alternative diagnosis controls that would not have undergone index test in real practice). Further, if target condition diagnosed in the study population was not representative of the entire spectrum of disease, such as limited spectrum of severity (e.g. only mild forms) or limited type of endometriosis (e.g. only DIE)
Low	If the study includes only clinically relevant population that would have undergone index test in real practice and includes representative form of target condition
Unclear	If this information was unclear (e.g. severity of endometriosis was not reported)
Domain 2 - Index test	
Description	Describe the index test, how it was conducted and interpreted
Type of bias assessed	Test review bias, clinical review bias, interobserver variation bias
Review Question	Any type of test that combines several different testing modalities with and without clinical history or examination
Informaton collected	Index test name, description of positive case definition by index test as reported, threshold for positive result, examiners (number, level of expertise, blinding), interobserver variability, conflict of interests
Signalling question 1	***Were the index test results interpreted without knowledge of the results of the reference standard?***
Yes	If the operators performing/interpreting index test were unaware of the results of reference standard
No	If the operators performing/interpreting index test were not blinded to the results of reference standard
Unclear	If this information was unclear

Table 3. Risk of bias and applicability judgments for the quality assessment of diagnostic accuracy studies (QUADAS-2) *(Continued)*

Signalling question 2	***If a threshold was used, was it pre-specified? [only for the studies that included biomarker testing]***
Yes	If study clearly provided a threshold for positive result and was defined before execution/interpretation of index test
No	If a threshold for positive result was not provided or not defined prior to test execution
Unclear	If it was unclear whether a threshold was pre-specified or not
Signalling question 3	***Was a menstrual cycle phase considered in interpreting the index test? [only for the studies that included biomarker testing]***
Yes	If all the included participants were in the same phase of menstrual cycle or if the study reported subgroup analyses per cycle phase or if study reported the pooled estimates after impact of the cycle phase on biomarker expression was not detected
No	If study included participants in different phases of menstrual cycle, but effect of cycle phase on index test was not assessed
Unclear	If the cycle phase was not reported
Signalling question 4	***Did the study provide a clear pre-specified definition of what was considered to be a "positive" result of index test? [only for the studies that included imaging modalities]***
Yes	If study provided clear definition of positive findings and this was defined before execution/interpretation of index test
No	If definition of the positive result was not provided or if study described the findings derived from the index test and not defined prior to its execution
Unclear	If it was unclear whether the criteria were pre-specified or not
Signalling question 5	***Was the index test performed by a single operator or interpreted by consensus in a joint session? [only for the studies that included imaging modalities]***
Yes	If test was performed/interpreted either by single operator or interpreted after collegial discussion of the case
No	If test was performed/interpreted by various operators in different participants
Unclear	If this information was unclear
Signalling question 6	***Were the same clinical data available when the index test results were interpreted as would be available when the test is used in practice? [only for the studies that included imaging modalities]***
Yes	If operators performing/interpreting the test were aware of suspected endometriosis or of the clinical history, but were not aware of the results of other imaging tests or of previous diagnosis of endometriosis, including the results of previous surgeries
No	If operators performing/interpreting the test were informed on previously or recently surgically diagnosed endometriosis or were not blinded to the results of other imaging tests or tests raising suspicion for endometriosis
Unclear	If this information was unclear

Table 3. Risk of bias and applicability judgments for the quality assessment of diagnostic accuracy studies (QUADAS-2) *(Continued)*

Risk of bias	**Could the conduct or interpretation of the index test have introduced bias?**
High	If 'no' classification for any of the first three questions [for the studies that included biomarker testing] or if 'no' classification for any of the following: signalling questions 1, 4, 5, 6 [for studies that included imaging modalities]
Low	If 'yes' classification for all the relevant questions: signalling questions 1 - 3 [for the studies that included biomarker testing] or signalling questions 1, 4, 5, 6 [for studies that included imaging modalities]
Unclear	If 'unclear' classification for any of the relevant questions
Concerns about applicability	**Are there concerns that the index test, its conduct, or interpretation differ from the review question?**
High	We did not consider the studies where index tests other than combinations of different testing modalities were included or where index test looked at other target conditions not specified in the review (e.g. studies aimed at classifying pelvic masses as benign and malignant), therefore none of the included studies was classified as 'high concern'
Low	We considered all types of combinations of different testing modalities as eligible, therefore all the included studies were classified as 'low concern', unless 'unclear' judgement was applicable
Unclear	If study reported, but did not present sufficient information on any of the following: laboratory method, sample handling, reagents used, radiological protocol or equipment (where applicable), experience of the test operators
Domain 3 - Reference standard	
Description	Describe the reference standard, how it was conducted and interpreted
Type of bias assessed	Verification bias, bias in estimation of diagnostic accuracy due to inadequate reference standard
Review Question	Target condition - pelvic endometriosis, ovarian endometriosis, DIE. Reference standard - visualisation of endometriosis at surgery (laparoscopy or laparotomy) with or without histological confirmation
Informaton collected	Target condition, prevalence of target condition in the sample, reference standard, description of positive case definition by reference test as reported, examiners (number, level of expertise, blinding)
Signalling question 1	***Is the reference standards likely to correctly classify the target condition?***
Yes	If the study reported at least one of the following: surgical procedure was described in sufficient details or criteria for positive reference standard were stated or diagnosis was confirmed by histopathology or the procedure was performed by the team with high level of expertise in diagnosis/surgical treatment of target condition, including tertiary referral centres for endometriosis
No	If reference standard did not classify target condition correctly; considering the inclusion criteria and a nature of the reference standard, none of the studies were classified as 'no' for this item
Unclear	If information on execution of the reference standard, its interpretation or operators was unclear
Signalling question 2	***Were the reference standard results interpreted without knowledge of the results of the index tests?***

Table 3. Risk of bias and applicability judgments for the quality assessment of diagnostic accuracy studies (QUADAS-2) *(Continued)*

Yes	If operators performing the reference test were unaware of the results of index test
No	If operators performing the reference test were aware of the results of index test
Unclear	If this information was unclear
Risk of bias	**Could the reference standard, its conduct, or its interpretation have introduced bias?**
High	If 'no' classification for any of the above two questions
Low	If 'yes' classification for all the above two questions The Robinson Institute, University of Adelaide
Unclear	If 'unclear' classification for any of the above two questions and 'high risk' judgement was not applicable
Concerns about applicability	**Are there concerns that the target condition as defined by the reference standard does not match the question?**
High	We excluded the studies where participants did not undergo surgery for diagnosis of endometriosis, therefore none of the included studies were classified as 'high concern'
Low	Considering the inclusion criteria, all the studies were classified as 'low concern', therefore all the included studies were classified as 'low concern', unless 'unclear' judgement was applicable
Unclear	Only studies were laparoscopy/ laparotomy served as a reference test were included; therefore none of the included studies was classified as 'unclear concern'
Domain 4 - Flow and timing	
Description	Describe any participants who did not receive the index tests or reference standard or who were excluded from the 2 x 2 table, describe the interval and any interventions between index tests and the reference standard
Type of bias assessed	Disease progression bias, bias of diagnostic performance due to missing data
Review Question	Less than 12 months interval between index test and reference standard - endometriosis may progress over the time, so we had chosen an arbitrary time interval of '12' months as an acceptable time interval between the index test and surgical confirmation of diagnosis
Informaton collected	Time interval between index test and reference standard, withdrawals (overall number of reported and if were explained)
Signalling question 1	***Was there an appropriate interval between index test (sample collection) and reference standard?***
Yes	If time interval was reported and was less than 12 months
No	We excluded all the studies where time interval was longer than 12 months, therefore none of the included studies were classified as 'no' for this item
Unclear	if time interval was not stated clearly, but authors description allowed to assume that the interval was reasonably short
Signalling question 2	***Did all patients receive the same reference standard?***

Table 3. Risk of bias and applicability judgments for the quality assessment of diagnostic accuracy studies (QUADAS-2) *(Continued)*

Yes	If all participants underwent laparoscopy/laparotomy as a reference standard. Considering the inclusion criteria, all the studies were classified as 'yes' for this item, as anticipated
No	If all participants did not undergo surgery or had alternative reference standard or if only a subset of participants had surgery as reference standard, but the information on this population was not available in isolation
Unclear	If this information was unclear. Considering the inclusion criteria, none of the included studies were classified as 'unclear' for this item
Signalling question 3	***Were all patients included in the analysis?***
Yes	If all the participants were included in the analysis or if the participants were excluded because they did not meet inclusion criteria prior to execution of index test or if the withdrawals were less than 5% of the enrolled population (arbitrary selected cut-off)
No	If any participants were excluded from the analysis because of un interpretable results, inability to undergo either index test or reference standard or unclear reasons
Unclear	If this information was unclear
Risk of bias	**Could the patient flow have introduced bias?**
High	If 'no' classification for any of the above three questions
Low	If 'yes' classification for all the above three questions
Unclear	If 'unclear' classification for any of the above three questions and 'high risk' judgement was not applicable

DIE: deep infiltrating endometriosis

APPENDICES

Appendix 1. Biomarkers search strategy for CENTRAL (OVID platform)

Database: EBM Reviews - Cochrane Central Register of Controlled Trials <July 2015 (3.09.2015)>

1 (biomarker$ or marker$).tw. (23692)

2 Laboratory Test$.tw. (2793)

3 growth factor$.tw. (5448)

4 scatter factor$.tw. (8)

5 cytokine$.tw. (6264)

6 hepatocyte growth factor.tw. (111)

7 (FGF or fibroblast growth factor$).tw. (433)

8 (PDGF or platelet derived growth factor$).tw. (250)

9 (EGF or epidermal growth factor$).tw. (1077)

10 (IGF-I or insulin-like growth factor$ or IGF1).tw. (2132)

11 (TGF-a or transforming growth factor alfa or TGFa).tw. (519)

12 (TGF-b or transforming growth factor beta or TGFb).tw. (236)

13 (EGFR or epidermal growth factor receptor$).tw. (1905)

14 (VEGF or vascular endothelial growth factor$).tw. (1532)

15 exp Luteinizing Hormone/bl [Blood] (151)

16 leptin$.tw. (1399)

17 exp Progesterone/bl [Blood] (58)

18 Proteolytic enzyme$.tw. (136)

19 exp matrix metalloproteinase 1/ or exp matrix metalloproteinase 2/ or exp matrix metalloproteinase 3/ or exp matrix metalloproteinase 9/ (292)

20 matrix metalloproteinase$.tw. (676)

21 MMP$.tw. (905)

22 TIMP$.tw. (229)

23 exp "tissue inhibitor of metalloproteinase-1"/ or exp "tissue inhibitor of metalloproteinase-2"/ (101)

24 exp Glycoproteins/ (10108)

25 (Ca-125 or Ca125 or cancer antigen 125).tw. (305)

26 (Ca-19-9 or Ca19-9 or cancer antigen 19-9).tw. (71)

27 (PP 14 or PP14).tw. (23)

28 serum placental protein$.tw. (6)

29 exp Follistatin/ (13)

30 Osteopontin$.tw. (80)

31 exp intercellular adhesion molecule-1/ or exp selectins/ (929)

32 soluble intercellular adhesion.tw. (256)

33 Soluble adhesion molecule$.tw. (89)

34 sICAM.tw. (319)

35 sVCAM$.tw. (223)

36 (sEcadherin or soluble E-cadherin).tw. (4)

37 (sEselectin or soluble E-selectin).tw. (99)

38 exp t-lymphocytes/ or exp natural killer t-cells/ (2645)

39 Immune cells alteration$.tw. (1)

40 (T helper$ or T supressor$ or T helper$ T supressor$ ratio).tw. (445)

41 Total complement level$.tw. (0)

42 Autoantibodies.tw. (428)

43 exp Antibodies, Antiphospholipid/ (85)

44 Anti-endometrial.tw. (0)

45 Antiphospholipid$.tw. (152)

46 exp hla antigens/ or exp hla-a1 antigen/ or exp hla-a2 antigen/ (563)

47 (HLA or human leucocyte antigen$).tw. (1724)

48 Anti-laminin-1.tw. (0)

49 Anti-thyroid.tw. (49)

50 Anti-Thomsen Friedenreich antigen$.tw. (0)

51 Anti-transferrin.tw. (0)

52 Anti-LDL.tw. (3)

53 (Anti-2HSG or Heremans-Schmidt glycoprotein).tw. (0)

54 interleukin$.tw. (7276)

55 (MCP-I or monocyte chemoattractant protein-I).tw. (0)

56 (MIF or migration inhibitory factor$).tw. (75)

57 (TNF-a or tumour necrosis factor$ alfa).tw. (3923)

58 Fas ligand$.tw. (47)

59 Endometrial marker$.tw. (2)

60 CAMs.tw. (53)

61 cell adhesion molecule$.tw. (568)

62 exp Integrins/ (781)

63 Integrin$.tw. (248)

64 Selectin$.tw. (2183)

65 Cadherin$.tw. (71)

66 Aromatase P450.tw. (3)

67 estrogen receptor$.tw. (1252)

68 progesterone receptor$.tw. (531)

69 MTMMP$.tw. (0)

70 cyr61.tw. (1)

71 exp Cysteine-Rich Protein 61/ (1)

72 cysteine-rich heparin-binding protein$.tw. (0)

73 (ANXA 1 or ANXA1).tw. (3)

74 (Annexin 1 or Annexin1).tw. (2)

75 (PGP 9?5 or PGP9?5 or protein gene product$).tw. (18)

76 serum marker$.tw. (411)

77 neural marker$.tw. (9)

78 cell surface marker$.tw. (46)

79 inflammatory marker$.tw. (1739)

80 microarray$.tw. (501)

81 microRNA$.tw. (103)

82 proteomic$.tw. (176)

83 genomic$.tw. (526)

84 (endometri$ adj2 biops$).tw. (464)

85 Follistatin$.tw. (26)

86 Vascular Endothelial Growth Factor A/ (560)

87 Vitamin D-Binding Protein/ (18)

88 exp Cytokines/ (13960)

89 exp interleukins/ or exp interleukin-1/ or exp interleukin-6/ or exp interleukin-8/ or exp interleukin-12/ or exp interleukin-13/ (4413)

90 exp Epidermal Growth Factor/ (91)

91 exp Fibroblast Growth Factors/ (197)

92 Platelet-Derived Growth Factor/ (99)

93 Keratin-19/ (19)

94 exp Clinical Laboratory Techniques/ (35164)

95 (Luteinizing Hormone$ or LH).tw. (2935)

96 cytokeratin-19.tw. (25)

97 (VDBP or vitamin D-binding protein$).tw. (44)

98 urinary peptide$.tw. (8)

99 VDBP-Cr.tw. (0)

100 urinary VDBP corrected for creatinine expression.tw. (0)

101 urinary marker$.tw. (67)

102 or/1-101 (90390)

103 Endometriosis/di [Diagnosis] (6)

104 102 or 103 (90394)

105 exp Endometriosis/ (469)

106 Endometrio$.tw. (1026)

107 105 or 106 (1067)

108 104 and 107 (226)

109 (animals not (humans and animals)).sh. (1)

110 108 not 109 (226)

Appendix 2. Biomarkers search strategy for MEDLINE (OVID platform)

Database: MEDLINE (Ovid) <1946 to February, week 2 2015 (16.2.2015)>

1 (biomarker$ or marker$).tw. (605002)

2 Laboratory Test$.tw. (29839)

3 growth factor$.tw. (272049)

4 scatter factor$.tw. (1287)

5 cytokine$.tw. (250618)

6 hepatocyte growth factor.tw. (8053)

7 (FGF or fibroblast growth factor$).tw. (31798)

8 (PDGF or platelet derived growth factor$).tw. (19864)

9 (EGF or epidermal growth factor$).tw. (58069)

10 (IGF-I or insulin-like growth factor$ or IGF1).tw. (43539)

11 (TGF-a or transforming growth factor alfa or TGFa).tw. (281)

12 (TGF-b or transforming growth factor beta or TGFb).tw. (28842)

13 (EGFR or epidermal growth factor receptor$).tw. (41719)

14 (VEGF or vascular endothelial growth factor$).tw. (53588)

15 exp Luteinizing Hormone/bl [Blood] (24587)

16 leptin$.tw. (24994)

17 exp Progesterone/bl [Blood] (18412)

18 Proteolytic enzyme$.tw. (9768)

19 exp matrix metalloproteinase 1/ or exp matrix metalloproteinase 2/ or exp matrix metalloproteinase 3/ or exp matrix metalloproteinase 9/ (22968)

20 matrix metalloproteinase$.tw. (34522)

21 MMP$.tw. (44439)

22 TIMP$.tw. (10777)

23 exp "tissue inhibitor of metalloproteinase-1"/ or exp "tissue inhibitor of metalloproteinase-2"/ (6146)

24 exp Glycoproteins/ (637149)

25 (Ca-125 or Ca125 or cancer antigen 125).tw. (6761)

26 (Ca-19-9 or Ca19-9 or cancer antigen 19-9).tw. (4194)

27 (PP 14 or PP14).tw. (229)

28 serum placental protein$.tw. (33)

29 exp Follistatin/ (1134)

30 Osteopontin$.tw. (6769)

31 exp intercellular adhesion molecule-1/ or exp selectins/ (25302)

32 soluble intercellular adhesion.tw. (1588)

33 Soluble adhesion molecule$.tw. (779)

34 sICAM.tw. (2258)

35 sVCAM$.tw. (1277)

36 (sEcadherin or soluble E-cadherin).tw. (95)

37 (sEselectin or soluble E-selectin).tw. (689)

38 exp t-lymphocytes/ or exp natural killer t-cells/ (272580)

39 Immune cells alteration$.tw. (1)

40 (T helper$ or T supressor$ or T helper$ T supressor$ ratio).tw. (21275)

41 Total complement level$.tw. (23)

42 Autoantibodies.tw. (33457)

43 exp Antibodies, Antiphospholipid/ (7522)

44 Anti-endometrial.tw. (23)

45 Antiphospholipid$.tw. (9974)

46 exp hla antigens/ or exp hla-a1 antigen/ or exp hla-a2 antigen/ (64462)

47 (HLA or human leucocyte antigen$).tw. (80501)

48 Anti-laminin-1.tw. (33)

49 Anti-thyroid.tw. (1414)

50 Anti-Thomsen Friedenreich antigen$.tw. (6)

51 Anti-transferrin.tw. (275)

52 Anti-LDL.tw. (181)

53 (Anti-2HSG or Heremans-Schmidt glycoprotein).tw. (3)

54 interleukin$.tw. (175195)

55 (MCP-I or monocyte chemoattractant protein-I).tw. (44)

56 (MIF or migration inhibitory factor$).tw. (4479)

57 (TNF-a or tumour necrosis factor$ alfa).tw. (1344)

58 Fas ligand$.tw. (6032)

59 Endometrial marker$.tw. (11)

60 CAMs.tw. (1756)

61 cell adhesion molecule$.tw. (20903)

62 exp Integrins/ (44414)

63 Integrin$.tw. (39960)

64 Selectin$.tw. (55426)

65 Cadherin$.tw. (20780)

66 Aromatase P450.tw. (180)

67 estrogen receptor$.tw. (38819)

68 progesterone receptor$.tw. (16623)

69 MTMMP$.tw. (7)

70 cyr61.tw. (559)

71 exp Cysteine-Rich Protein 61/ (386)

72 cysteine-rich heparin-binding protein$.tw. (9)

73 (ANXA 1 or ANXA1).tw. (313)

74 (Annexin 1 or Annexin1).tw. (339)

75 (PGP 9?5 or PGP9?5 or protein gene product$).tw. (2096)

76 serum marker$.tw. (5429)

77 neural marker$.tw. (925)

78 cell surface marker$.tw. (4456)

79 inflammatory marker$.tw. (10916)

80 microarray$.tw. (75404)

81 microRNA$.tw. (29731)

82 proteomic$.tw. (45292)

83 genomic$.tw. (190985)

84 (endometri$ adj2 biops$).tw. (3411)

85 Follistatin$.tw. (1663)

86 Vascular Endothelial Growth Factor A/ (35738)

87 Vitamin D-Binding Protein/ (1282)

88 exp Cytokines/ (547522)

89 exp interleukins/ or exp interleukin-1/ or exp interleukin-6/ or exp interleukin-8/ or exp interleukin-12/ or exp interleukin-13/ (188479)

90 exp Epidermal Growth Factor/ (21298)

91 exp Fibroblast Growth Factors/ (25075)

92 Platelet-Derived Growth Factor/ (11030)

93 Keratin-19/ (1090)

94 exp Clinical Laboratory Techniques/ (2132820)

95 (Luteinizing Hormone$ or LH).tw. (56679)

96 cytokeratin-19.tw. (1469)

97 (VDBP or vitamin D-binding protein$).tw. (1158)

98 urinary peptide$.tw. (137)

99 VDBP-Cr.tw. (1)

100 urinary VDBP corrected for creatinine expression.tw. (1)

101 urinary marker$.tw. (638)

102 or/1-101 (4086291)

103 Endometriosis/di [Diagnosis] (3354)

104 102 or 103 (4088946)

105 exp Endometriosis/ (17244)

106 Endometrio$.tw. (21492)

107 105 or 106 (24940)

108 104 and 107 (10490)

109 (animals not (humans and animals)).sh. (3892900)

110 108 not 109 (10113)

Additional search February 2015 - May 2015

Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) <1946 to Present (3.9.2015)>

1 (biomarker$ or marker$).tw. (652345)

2 Laboratory Test$.tw. (31389)

3 growth factor$.tw. (287701)

4 scatter factor$.tw. (1326)

5 cytokine$.tw. (267766)

6 hepatocyte growth factor.tw. (8585)

7 (FGF or fibroblast growth factor$).tw. (33674)

8 (PDGF or platelet derived growth factor$).tw. (20842)

9 (EGF or epidermal growth factor$).tw. (61625)

10 (IGF-I or insulin-like growth factor$ or IGF1).tw. (45386)

11 (TGF-a or transforming growth factor alfa or TGFa).tw. (306)

12 (TGF-b or transforming growth factor beta or TGFb).tw. (30559)

13 (EGFR or epidermal growth factor receptor$).tw. (46446)

14 (VEGF or vascular endothelial growth factor$).tw. (58203)

15 exp Luteinizing Hormone/bl [Blood] (24870)

16 leptin$.tw. (26783)

17 exp Progesterone/bl [Blood] (18699)

18 Proteolytic enzyme$.tw. (9992)

19 exp matrix metalloproteinase 1/ or exp matrix metalloproteinase 2/ or exp matrix metalloproteinase 3/ or exp matrix metalloproteinase 9/ (24504)

20 matrix metalloproteinase$.tw. (37055)

21 MMP$.tw. (47849)

22 TIMP$.tw. (11419)

23 exp "tissue inhibitor of metalloproteinase-1"/ or exp "tissue inhibitor of metalloproteinase-2"/ (6447)

24 exp Glycoproteins/ (662211)

25 (Ca-125 or Ca125 or cancer antigen 125).tw. (7058)

26 (Ca-19-9 or Ca19-9 or cancer antigen 19-9).tw. (4399)

27 (PP 14 or PP14).tw. (232)

28 serum placental protein$.tw. (34)

29 exp Follistatin/ (1180)

30 Osteopontin$.tw. (7267)

31 exp intercellular adhesion molecule-1/ or exp selectins/ (26225)

32 soluble intercellular adhesion.tw. (1663)

33 Soluble adhesion molecule$.tw. (795)

34 sICAM.tw. (2374)

35 sVCAM$.tw. (1360)

36 (sEcadherin or soluble E-cadherin).tw. (97)

37 (sEselectin or soluble E-selectin).tw. (713)

38 exp t-lymphocytes/ or exp natural killer t-cells/ (284378)

39 Immune cells alteration$.tw. (1)

40 (T helper$ or T supressor$ or T helper$ T supressor$ ratio).tw. (22494)

41 Total complement level$.tw. (24)

42 Autoantibodies.tw. (35161)

43 exp Antibodies, Antiphospholipid/ (7759)

44 Anti-endometrial.tw. (22)

45 Antiphospholipid$.tw. (10351)

46 exp hla antigens/ or exp hla-a1 antigen/ or exp hla-a2 antigen/ (66724)

47 (HLA or human leucocyte antigen$).tw. (83856)

48 Anti-laminin-1.tw. (33)

49 Anti-thyroid.tw. (1478)

50 Anti-Thomsen Friedenreich antigen$.tw. (8)

51 Anti-transferrin.tw. (284)

52 Anti-LDL.tw. (183)

53 (Anti-2HSG or Heremans-Schmidt glycoprotein).tw. (3)

54 interleukin$.tw. (184697)

55 (MCP-I or monocyte chemoattractant protein-I).tw. (46)

56 (MIF or migration inhibitory factor$).tw. (4718)

57 (TNF-a or tumour necrosis factor$ alfa).tw. (1428)

58 Fas ligand$.tw. (6204)

59 Endometrial marker$.tw. (11)

60 CAMs.tw. (1823)

61 cell adhesion molecule$.tw. (22033)

62 exp Integrins/ (46487)

63 Integrin$.tw. (42447)

64 Selectin$.tw. (58540)

65 Cadherin$.tw. (22688)

66 Aromatase P450.tw. (182)

67 estrogen receptor$.tw. (41210)

68 progesterone receptor$.tw. (17437)

69 MTMMP$.tw. (7)

70 cyr61.tw. (620)

71 exp Cysteine-Rich Protein 61/ (425)

72 cysteine-rich heparin-binding protein$.tw. (9)

73 (ANXA 1 or ANXA1).tw. (355)

74 (Annexin 1 or Annexin1).tw. (358)

75 (PGP 9?5 or PGP9?5 or protein gene product$).tw. (2190)

76 serum marker$.tw. (5721)

77 neural marker$.tw. (1026)

78 cell surface marker$.tw. (4751)

79 inflammatory marker$.tw. (12244)

80 microarray$.tw. (81764)

81 microRNA$.tw. (35967)

82 proteomic$.tw. (49911)

83 genomic$.tw. (205064)

84 (endometri$ adj2 biops$).tw. (3518)

85 Follistatin$.tw. (1762)

86 Vascular Endothelial Growth Factor A/ (38477)

87 Vitamin D-Binding Protein/ (1356)

88 exp Cytokines/ (575020)

89 exp interleukins/ or exp interleukin-1/ or exp interleukin-6/ or exp interleukin-8/ or exp interleukin-12/ or exp interleukin-13/ (197567)

90 exp Epidermal Growth Factor/ (21875)

91 exp Fibroblast Growth Factors/ (26259)

92 Platelet-Derived Growth Factor/ (11355)

93 Keratin-19/ (1179)

94 exp Clinical Laboratory Techniques/ (2203416)

95 (Luteinizing Hormone$ or LH).tw. (57796)

96 cytokeratin-19.tw. (1538)

97 (VDBP or vitamin D-binding protein$).tw. (1262)

98 urinary peptide$.tw. (148)

99 VDBP-Cr.tw. (1)

100 urinary VDBP corrected for creatinine expression.tw. (1)

101 urinary marker$.tw. (679)

102 or/1-101 (4283825)

103 Endometriosis/di [Diagnosis] (3449)

104 102 or 103 (4286552)

105 exp Endometriosis/ (17833)

106 Endometrio$.tw. (22478)

107 105 or 106 (26003)

108 104 and 107 (10936)

109 (animals not (humans and animals)).sh. (4004321)

110 108 not 109 (10539)

111 (201501$ or 201502$ or 201503$ or 201504$).ed. (322721)

112 110 and 111 (215)

Appendix 3. Biomarkers search strategy for EMBASE (OVID platform)

Database: EMBASE (Ovid) <1980 to 2015 Week 07 (16.02.2015)>

1 Laboratory Test$.tw. (41662)

2 growth factor$.tw. (318593)

3 scatter factor$.tw. (1388)

4 cytokine$.tw. (322134)

5 hepatocyte growth factor.tw. (9594)

6 (FGF or fibroblast growth factor$).tw. (37191)

7 (PDGF or platelet derived growth factor$).tw. (23530)

8 (EGF or epidermal growth factor$).tw. (69553)

9 (IGF-I or insulin-like growth factor$ or IGF1).tw. (49806)

10 (TGF-a or transforming growth factor alfa or TGFa).tw. (542)

11 (TGF-b or transforming growth factor beta or TGFb).tw. (30820)

12 (EGFR or epidermal growth factor receptor$).tw. (64664)

13 (VEGF or vascular endothelial growth factor$).tw. (73191)

14 exp luteinizing hormone/ec [Endogenous Compound] (21924)

15 leptin$.tw. (32576)

16 exp progesterone blood level/ or exp progesterone urine level/ (6285)

17 Proteolytic enzyme$.tw. (9643)

18 exp matrix metalloproteinase/ (19364)

19 matrix metalloproteinase$.tw. (41445)

20 MMP$.tw. (58466)

21 TIMP$.tw. (14174)

22 exp "tissue inhibitor of metalloproteinase 2"/ (4824)

23 exp "tissue inhibitor of metalloproteinase 1"/ (8779)

24 exp glycoprotein/ec [Endogenous Compound] (246077)

25 (Ca-125 or Ca125 or cancer antigen 125).tw. (9536)

26 (Ca-19-9 or Ca19-9 or cancer antigen 19-9).tw. (6054)

27 (PP 14 or PP14).tw. (244)

28 serum placental protein$.tw. (43)

29 exp follistatin/ (2148)

30 Osteopontin$.tw. (8475)

31 exp intercellular adhesion molecule 1/ (32066)

32 exp selectin/ (3082)

33 soluble intercellular adhesion.tw. (1788)

34 Soluble adhesion molecule$.tw. (919)

35 sICAM.tw. (2888)

36 sVCAM$.tw. (1793)

37 (sEcadherin or soluble E-cadherin).tw. (120)

38 (sEselectin or soluble E-selectin).tw. (822)

39 exp T lymphocyte/ (374675)

40 exp natural killer T cell/ (5800)

41 Immune cells alteration$.tw. (6)

42 (T helper$ or T supressor$ or T helper$ T supressor$ ratio).tw. (24786)

43 Total complement level$.tw. (20)

44 Autoantibodies.tw. (42037)

45 exp phospholipid antibody/ (9920)

46 Anti-endometrial.tw. (23)

47 Antiphospholipid$.tw. (13777)

48 exp HLA antigen/ (81011)

49 exp HLA A1 antigen/ (597)

50 exp HLA A2 antigen/ (3288)

51 (HLA or human leucocyte antigen$).tw. (104497)

52 Anti-laminin-1.tw. (43)

53 Anti-thyroid.tw. (1873)

54 Anti-Thomsen Friedenreich antigen$.tw. (5)

55 Anti-transferrin.tw. (290)

56 Anti-LDL.tw. (186)

57 (Anti-2HSG or Heremans-Schmidt glycoprotein).tw. (4)

58 interleukin$.tw. (199692)

59 (MCP-I or monocyte chemoattractant protein-I).tw. (112)

60 (MIF or migration inhibitory factor$).tw. (5063)

61 (TNF-a or tumour necrosis factor$ alfa).tw. (5998)

62 Fas ligand$.tw. (6708)

63 Endometrial marker$.tw. (18)

64 CAMs.tw. (2100)

65 cell adhesion molecule$.tw. (24039)

66 exp integrin/ (29036)

67 Integrin$.tw. (48293)

68 Selectin$.tw. (67300)

69 Cadherin$.tw. (27150)

70 Aromatase P450.tw. (202)

71 estrogen receptor$.tw. (46656)

72 progesterone receptor$.tw. (19861)

73 MTMMP$.tw. (15)

74 cyr61.tw. (755)

75 exp cysteine rich protein 61/ (753)

76 cysteine-rich heparin-binding protein$.tw. (12)

77 (ANXA 1 or ANXA1).tw. (452)

78 (Annexin 1 or Annexin1).tw. (425)

79 (PGP 9?5 or PGP9?5 or protein gene product$).tw. (2620)

80 serum marker$.tw. (7720)

81 neural marker$.tw. (1119)

82 cell surface marker$.tw. (5851)

83 inflammatory marker$.tw. (17339)

84 microarray$.tw. (101846)

85 microRNA$.tw. (40082)

86 proteomic$.tw. (55191)

87 genomic$.tw. (217184)

88 (endometri$ adj2 biops$).tw. (4369)

89 Follistatin$.tw. (1945)

90 exp vasculotropin/ (69810)

91 Vascular Endothelial Growth Factor A.tw. (2275)

92 exp vitamin D binding protein/ (2064)

93 exp cytokine/ (1034772)

94 exp interleukin derivative/ (2790)

95 exp interleukin 1/ (48499)

96 exp interleukin 6/ (136328)

97 exp interleukin 8/ (48884)

98 exp interleukin 12/ (31842)

99 exp interleukin 13/ (13584)

100 exp epidermal growth factor/ (32130)

101 exp fibroblast growth factor/ (13858)

102 cytokeratin 19/ (3601)

103 platelet derived growth factor/ (18930)

104 cytokeratin-19.tw. (1918)

105 (VDBP or vitamin D-binding protein$).tw. (1413)

106 urinary peptide$.tw. (174)

107 VDBP-Cr.tw. (1)

108 urinary VDBP corrected for creatinine expression.tw. (1)

109 urinary marker$.tw. (830)

110 exp blood analysis/ (118854)

111 exp endometrium biopsy/ (4988)

112 exp urinalysis/ or exp biological marker/ (210153)

113 (biomarker or biomarkers).tw. (159748)

114 or/1-113 (2734501)

115 endometriosis/di [Diagnosis] (4979)

116 114 or 115 (2738583)

117 exp endometriosis/ (25923)

118 Endometriosis.tw. (22110)

119 117 or 118 (27911)

120 116 and 119 (10326)

121 Animal/ not Human/ (1204497)

122 120 not 121 (10279)

Additional search February 2015 - May 2015

Embase <1980 to 2015 Week 35 (3.09.2015)>

1 Laboratory Test$.tw. (44290)

2 growth factor$.tw. (335543)

3 scatter factor$.tw. (1407)

4 cytokine$.tw. (343623)

5 hepatocyte growth factor.tw. (10104)

6 (FGF or fibroblast growth factor$).tw. (39159)

7 (PDGF or platelet derived growth factor$).tw. (24591)

8 (EGF or epidermal growth factor$).tw. (73599)

9 (IGF-I or insulin-like growth factor$ or IGF1).tw. (51838)

10 (TGF-a or transforming growth factor alfa or TGFa).tw. (583)

11 (TGF-b or transforming growth factor beta or TGFb).tw. (32580)

12 (EGFR or epidermal growth factor receptor$).tw. (71526)

13 (VEGF or vascular endothelial growth factor$).tw. (79087)

14 exp luteinizing hormone/ec [Endogenous Compound] (22767)

15 leptin$.tw. (34921)

16 exp progesterone blood level/ or exp progesterone urine level/ (6534)

17 Proteolytic enzyme$.tw. (9903)

18 exp matrix metalloproteinase/ (20462)

19 matrix metalloproteinase$.tw. (44380)

20 MMP$.tw. (63208)

21 TIMP$.tw. (15146)

22 exp "tissue inhibitor of metalloproteinase 2"/ (5136)

23 exp "tissue inhibitor of metalloproteinase 1"/ (9381)

24 exp glycoprotein/ec [Endogenous Compound] (260024)

25 (Ca-125 or Ca125 or cancer antigen 125).tw. (10051)

26 (Ca-19-9 or Ca19-9 or cancer antigen 19-9).tw. (6446)

27 (PP 14 or PP14).tw. (243)

28 serum placental protein$.tw. (44)

29 exp follistatin/ (2283)

30 Osteopontin$.tw. (9173)

31 exp intercellular adhesion molecule 1/ (33492)

32 exp selectin/ (3217)

33 soluble intercellular adhesion.tw. (1865)

34 Soluble adhesion molecule$.tw. (944)

35 sICAM.tw. (3049)

36 sVCAM$.tw. (1924)

37 (sEcadherin or soluble E-cadherin).tw. (125)

38 (sEselectin or soluble E-selectin).tw. (861)

39 exp T lymphocyte/ (394405)

40 exp natural killer T cell/ (6310)

41 Immune cells alteration$.tw. (6)

42 (T helper$ or T supressor$ or T helper$ T supressor$ ratio).tw. (26082)

43 Total complement level$.tw. (20)

44 Autoantibodies.tw. (44153)

45 exp phospholipid antibody/ (10362)

46 Anti-endometrial.tw. (25)

47 Antiphospholipid$.tw. (14399)

48 exp HLA antigen/ (83748)

49 exp HLA A1 antigen/ (622)

50 exp HLA A2 antigen/ (3409)

51 (HLA or human leucocyte antigen$).tw. (109332)

52 Anti-laminin-1.tw. (43)

53 Anti-thyroid.tw. (2059)

54 Anti-Thomsen Friedenreich antigen$.tw. (7)

55 Anti-transferrin.tw. (297)

56 Anti-LDL.tw. (191)

57 (Anti-2HSG or Heremans-Schmidt glycoprotein).tw. (4)

58 interleukin$.tw. (210083)

59 (MCP-I or monocyte chemoattractant protein-I).tw. (114)

60 (MIF or migration inhibitory factor$).tw. (5342)

61 (TNF-a or tumour necrosis factor$ alfa).tw. (6488)

62 Fas ligand$.tw. (6895)

63 Endometrial marker$.tw. (18)

64 CAMs.tw. (2198)

65 cell adhesion molecule$.tw. (25207)

66 exp integrin/ (30330)

67 Integrin$.tw. (50938)

68 Selectin$.tw. (71624)

69 Cadherin$.tw. (29496)

70 Aromatase P450.tw. (207)

71 estrogen receptor$.tw. (49530)

72 progesterone receptor$.tw. (21068)

73 MTMMP$.tw. (16)

74 cyr61.tw. (822)

75 exp cysteine rich protein 61/ (829)

76 cysteine-rich heparin-binding protein$.tw. (12)

77 (ANXA 1 or ANXA1).tw. (500)

78 (Annexin 1 or Annexin1).tw. (440)

79 (PGP 9?5 or PGP9?5 or protein gene product$).tw. (2760)

80 serum marker$.tw. (8158)

81 neural marker$.tw. (1234)

82 cell surface marker$.tw. (6222)

83 inflammatory marker$.tw. (19492)

84 microarray$.tw. (110181)

85 microRNA$.tw. (47554)

86 proteomic$.tw. (60599)

87 genomic$.tw. (233444)

88 (endometri$ adj2 biops$).tw. (4589)

89 Follistatin$.tw. (2081)

90 exp vasculotropin/ (74115)

91 Vascular Endothelial Growth Factor A.tw. (2526)

92 exp vitamin D binding protein/ (2196)

93 exp cytokine/ (1094317)

94 exp interleukin derivative/ (3281)

95 exp interleukin 1/ (50850)

96 exp interleukin 6/ (147379)

97 exp interleukin 8/ (52281)

98 exp interleukin 12/ (33479)

99 exp interleukin 13/ (14685)

100 exp epidermal growth factor/ (33057)

101 exp fibroblast growth factor/ (14499)

102 cytokeratin 19/ (3886)

103 platelet derived growth factor/ (19655)

104 cytokeratin-19.tw. (2030)

105 (VDBP or vitamin D-binding protein$).tw. (1520)

106 urinary peptide$.tw. (189)

107 VDBP-Cr.tw. (1)

108 urinary VDBP corrected for creatinine expression.tw. (1)

109 urinary marker$.tw. (883)

110 exp blood analysis/ (124468)

111 exp endometrium biopsy/ (5197)

112 exp urinalysis/ or exp biological marker/ (232619)

113 (biomarker or biomarkers).tw. (182609)

114 or/1-113 (2911073)

115 endometriosis/di [Diagnosis] (5173)

116 114 or 115 (2915302)

117 exp endometriosis/ (27433)

118 Endometriosis.tw. (23449)

119 117 or 118 (29532)

120 116 and 119 (10922)

121 Animal/ not Human/ (1261620)

122 120 not 121 (10862)

123 (201501$ or 201502$ or 201503$ or 201504$).em. (49200)

124 122 and 123 (34)

Appendix 4. Biomarkers search strategy for CINAHL (EBSCO platform)

Database: CINAHL Plus with Full Text (EBSCOhost) <1980 to 20.04.2015>

#	Query	Results
S97	S3 AND S96	1131
S96	S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42 OR S43 OR S44 OR S45 OR S46 OR S47 OR S48 OR S49 OR S50 OR S51 OR S52 OR S53 OR S54 OR S55 OR S56 OR S57 OR S58 OR S59 OR S60 OR S61 OR S62 OR S63 OR S64 OR S65 OR S66 OR S67 OR S68 OR S69 OR S70 OR S71 OR S72 OR S73 OR S74 OR S75 OR S76 OR S77 OR S78 OR S79 OR S80 OR S81 OR S82 OR S83 OR S84 OR S85 OR S86 OR S87 OR S88 OR S89 OR S90 OR S91 OR S92 OR S93 OR S94 OR S95	341775
S95	TX urinary peptide*	1598
S94	TX (VDBP or vitamin D-binding protein*)	134
S93	TX cytokeratin-19	109

(Continued)

S92	TX (Luteinizing Hormone* or LH)	18041
S91	(MH "Diagnosis, Laboratory+")	101773
S90	"Keratin-19"	2
S89	(MH "Platelet-Derived Growth Factor")	394
S88	(MH "Epidermal Growth Factors")	1264
S87	(MH "Interleukins")	6584
S86	(MH "Cytokines")	6860
S85	TX Vitamin D-Binding Protein	131
S84	(MH "Vascular Endothelial Growth Factor A")	194
S83	TX (endometri* N2 biops*)	432
S82	TX (endometri* adj2 biops*)	0
S81	TX genomic$	7487
S80	TX proteomic*	2434
S79	TX microRNA	824
S78	TX microarray	3123
S77	TX (PGP 95 or PGP95 or protein gene product*)	9925
S76	TX (Annexin 1 or Annexin1)	472
S75	TX (ANXA 1 or ANXA1)	41
S74	TX cysteine-rich heparin-binding protein*	12
S73	(MH "Protein Array Analysis")	73
S72	TX cyr61	34
S71	TX MTMMP*	0
S70	TX progesterone receptor*	1927
S69	TX estrogen receptor*	5193
S68	TX Aromatase P450	38
S67	TX Cadherin*	900
S66	TX Selectin*	28411
S65	TX Integrin*	1587

(Continued)

S64	TX cell adhesion molecule*	1578
S63	TX CAMs	550
S62	TX Endometrial marker*	54
S61	TX Fas ligand	338
S60	TX (TNF-a or tumour necrosis factor* alfa)	1489
S59	TX (MIF or migration inhibitory factor*)	399
S58	TX (MCP-I or monocyte chemoattractant protein-I)	13
S57	TX interleukin	13809
S56	TX (Anti-2HSG or Heremans-Schmidt glycoprotein)	7
S55	TX Anti-LDL	9
S54	TX Anti-transferrin	3
S53	TX Anti-Thomsen Friedenreich antigen*	1
S52	TX Anti-thyroid	109
S51	TX Anti-laminin-1	15
S50	TX (HLA or human leucocyte antigen*)	4202
S49	(MM "HLA Antigens")	638
S48	TX Antiphospholipid*	1249
S47	TX Anti-endometrial	34
S46	(MH "Antibodies/BL/DU")	1294
S45	TX Autoantibodies	4385
S43	TX Total complement level	3
S42	TX (T helper* or T supressor*)	2341
S41	TX Immune cells alteration*	24
S40	TX natural killer t-cells	669
S39	(MM "T Lymphocytes")	2404
S38	TX (sEselectin or soluble E-selectin)	91
S37	TX (sEcadherin or soluble E-cadherin)	8
S36	TX sVCAM	100

(Continued)

S35	TX sICAM	173
S34	TX Soluble adhesion molecule	368
S33	TX soluble intercellular adhesion	237
S32	(MM "Cell Adhesion Molecules")	52
S31	TX Osteopontin*	416
S30	TX Follistatin	74
S29	TX serum placental protein*	11
S28	TX (Ca-19-9 or Ca19-9 or cancer antigen 19-9)	262
S27	TX (Ca-125 or Ca125 or cancer antigen 125)	831
S26	(MM "Glycoproteins/BL/DU")	224
S25	TX tissue inhibitor of metalloproteinase	423
S24	TX TIMP*	1845
S23	TX MMP*	4244
S22	TX matrix metalloproteinase*	3325
S21	TX Proteolytic enzyme*	1461
S20	(MM "Progesterone/BL/DU")	51
S19	TX leptin*	3258
S18	(MM "Luteinizing Hormone/BL/DU")	38
S17	TX (VEGF or vascular endothelial growth factor*)	7166
S16	TX (EGFR or epidermal growth factor receptor*)	6188
S15	TX (TGF-b or transforming growth factor beta or TGFb)	2972
S14	TX (TGF-a or transforming growth factor alfa or TGFa)	464
S13	TX (IGF-I or insulin-like growth factor* or IGF1)	3588
S12	TX (EGF or epidermal growth factor*)	6250
S11	TX (PDGF or platelet derived growth factor*)	3195
S10	TX (FGF or fibroblast growth factor*)	3395
S9	TX hepatocyte growth factor*	880
S8	TX cytokine*	20821

(Continued)

S7	TX scatter factor*	1864
S6	TX growth factor*	76163
S5	TX Laboratory Test*	82732
S4	TX (biomarker* or marker*)	84857
S3	S1 OR S2	2841
S2	TX Endometrio*	2841
S1	(MM "Endometriosis")	889
S4	TX (biomarker* or marker*)	61,794
S3	S1 OR S2	2,174
S2	TX Endometrio*	2,174
S1	(MM "Endometriosis")	1,306

Appendix 5. Biomarkers search strategy for other databases

Searches for clinical studies

Database: PsycINFO (Ovid) <1806 to April Week 2 2015 (20.04.2015)>

Search strategy:

1. endometriosis.tw. (174)

Database: Web of Science Core Collection (Thomson Reuters) <1900 to Present (20.04.2015)>

Search strategy:

1. Topic=(endometrio*) AND Topic=(diagnos* OR test* OR imag*); Timespan=All Years (7425)

Database: LILACS <20.04.2015>

Search strategy:

1. (tw:(endometriosis)) AND (tw:(diagnos*)) (420)

Database: OAIster (WorldCat.org) <20.04.2015>

Search strategy:

1. endometriosis and (marker* or biomarker*) (11)

2. endometriosis and diagnos* (446)

Database: TRIP <20.04.2015>

Search strategy:

1. (endometriosis and diagnos*) (1648)

Searches of trial registers for ongoing and registered trials

Database: ClinicalTrials.gov (US NIH) <20.04.2015>

Search strategy:

1. endometriosis (220)

2. endometriosis AND diagnosis (22)

Database: WHO International Clinical Trials Registry Platform (ICTRP) <20.04.2015>

Search strategy:

1. endometriosis (523)

Searches for the reviews as source of references to potentially relevant studies

Database: MEDION <10.01.2014>

Search strategy:

ICP Code female genital system (including breast), Signssymp medical imaging, laboratory tests, histology and cytology, endoscopy and laparoscopy. Filter: systematic reviews of diagnostic studies. (2)

Database: DARE (CRD) <20.04.2015>

Search strategy:

1. endometriosis (99)

PubMed, a 'Systematic Review' search under the 'Clinical Queries' link <20.04.2015>

Search strategy:

1. (endometriosis) AND systematic[sb] (418)

Category: Diagnosis; Scope: Broad

Searches for papers recently published and not yet indexed in the major databases

Search engine: PubMed <20.10.2014 to 20.04.2015>

Search strategy:

1. marker (14979) 2. test (61151) 3. diagnos* (69743) 4. biomarker (10806) 5. or/1-4 (7943) Filters: Publication date from 2014/10/20 to 2015/04/20	Index test(s) set
6. Endometriosis (584) Filters: Publication date from 2014/10/20 to 2015/04/20	Target condition set
7. 5 and 6 (267) Filters: Publication date from 2014/10/20 to 2015/04/20	Combined sets

Appendix 6. Imaging search strategy for CENTRAL

Database: EBM Reviews - Cochrane Central Register of Controlled Trials <April 2015 (20.4.2015)>

1. exp magnetic resonance imaging or exp ultrasonography or exp Imaging, Three-Dimensional or exp radiography (772)	Index test(s) set
2. (ultraso* or magnetic resonance imaging or MRI or imag*).tw. (36)	
3. diagnos* (106503)	
4. [mh diagnosis] (257329)	
5. or/1-4 (310878)	
6. exp endometriosis (142)	Target condition set
7. endometrio*.tw. (22)	
8. [mh endometriosis] (553)	
9. or/6-8 (681)	
10. 5 and 9 (465)	Combined sets
11. (animals not (humans and animals)).sh. (36)	
12. 10 not 11 (445)	

Appendix 7. Imaging search strategy for MEDLINE

1. exp magnetic resonance imaging/ or exp ultrasonography/ or exp Imaging, Three-Dimensional/ or exp radiography/ (1114639)	Index test(s) set
2. ultraso$.tw. or magnetic resonance imaging.tw. or MRI.tw. or imag$.tw. (1020000)	
3. diagnos$.tw. (1750239)	
4. or/1-3 (3048652)	
5. exp Endometriosis/ (17415)	Target condition set
6. Endometrio$.tw. (21775)	
7. or/5-6 (25236)	
8. 4 and 7 (8107)	Combined sets
9. (animals not (humans and animals)).sh. (3931867)	
10. 8 not 9 (7391)	

Appendix 8. Imaging search strategy for EMBASE

1. Ecography/exp or radiodiagnosis/exp (1988601)	Index test(s) set
2. 'magnetic resonance imaging':ab,ti or MRI:ab,ti or imag*:ab,ti or ultraso*:de,ab,ti (1370683)	
3. diagnos*:ab,ti (2373625)	

4. 'diagnostic accuracy':de or'diagnostic test accuracy study':de or 'diagnostic value':de (298281)	
5. or/1-4 (4437871)	
6. Endometrio*:de,ab,ti (37439)	Target condition set
7. 'endometriosis'/exp/dm_di (4976)	
8. or/6-7 (37439)	
9. #5 and #8 (13500)	Combined sets
10. animal:de not (animal:de and human:de) (3861389)	
11. #9 not #10 (12161)	

Appendix 9. Imaging search strategy for CINAHL

#	Query	Results	
S9	S3 AND S8 *Search modes - Boolean/Phrase* *Search Screen - Advanced Search*	668	Combined sets
S8	S4 OR S5 OR S6 OR S7	258011	
S7	TX imag*	258011	

S6	TX ultraso*	58570
S5	TX (magnetic resonance imaging or MRI)	58387
S4	TX (biomarker* or marker*)	84857
S3	S1 or S2	2841
S2	TX Endometrio*	2841
S1	(MM "Endometriosis")	889

Appendix 10. Imaging search strategy for other databases

Database: Web of Science Core Collection (Thomson Reuters) <1900 to Present (20.04.2015)>

Search strategy:

1. Topic=(endometrio*) AND Topic=(diagnos* OR test* OR imag*); Timespan=All Years (7425)

Database: LILACS <20.04.2015>

Search strategy:

1. (tw:(endometriosis)) AND (tw:(diagnos*)) (420)

Database: OAIster (WorldCat.org) <20.04.2015>

Search strategy:

1. endometriosis and (marker* or biomarker*) (11)

2. endometriosis and diagnos* (446)

Database: TRIP <20.04.2015>

Search strategy:

1. (endometriosis and diagnos*) (1648)

Searches of trial registers for ongoing and registered trials

Database: ClinicalTrials.gov (US NIH) <20.04.2015>

Search strategy:

1. endometriosis (220)

2. endometriosis AND diagnosis (22)

Database: WHO International Clinical Trials Registry Platform (ICTRP) <20.04.2015>

Search strategy:

1. endometriosis (523)

Searches for the reviews as source of references to potentially relevant studies

Database: MEDION <10.01.2014>

Search strategy:

ICP Code – female genital system (including breast), Signssymp – medical imaging, endoscopy and laparoscopy. Filter: systematic reviews of diagnostic studies (190)

Database: DARE (CRD) <20.04.2015>

Search strategy:

1. endometriosis (99)

PubMed, a 'Systematic Review' search under the 'Clinical Queries' link <20.04.2015>

Search strategy:

1. (endometriosis) AND systematic[sb] (418)

Category: Diagnosis; Scope: Broad

Searches for papers recently published and not yet indexed in the major databases

Search engine: PubMed <20.10.2014 to 20.04.2015>

Search strategy:

1. marker (14979) 2. test (61151) 3. diagnos* (69743) 4. biomarker (10806) 5. or/1-4 (7943) Filters: Publication date from 2014/10/20 to 2015/04/20	Index test(s) set
6. Endometriosis (584) Filters: Publication date from 2014/10/20 to 2015/04/20	Target condition set
7. 5 and 6 (267) Filters: Publication date from 2014/10/20 to 2015/04/20	Combined sets

CONTRIBUTIONS OF AUTHORS

Vicki Nisenblat co-ordinated the production of the protocol and the review series and was involved in literature search, quality appraisal and data extraction for the included studies and took a primary role in writing the review. Louise Hull participated in co-ordination of the protocol and the review series. Lucy Prentice participated in literature search, quality appraisal and data extraction for the included studies and contributed to the first draft of the review. Neil Johnson contributed to the conception and the design of the review and was involved in quality appraisal, data extraction for the included studies and critical revision of the manuscript. Patrick Bossuyt provided advice on statistical methods for the review. Cindy Farquhar critically reviewed the methodological aspects and participated in the study design. All the authors contributed to the revision and drafting of the review.

DECLARATIONS OF INTEREST

Cindy Farquhar is a director/shareholder of a fertility/gynaecology clinic and undertakes private practice within those premises.

Neil Johnson is involved in research funded by Abb-Vie. He has received support to attend conferences from MSD, Merck-Serono and Bayer. He has been on an advisory board for Vifor Pharma.

No other authors have any conflict of interest to declare.

SOURCES OF SUPPORT

Internal sources

- Cochrane Gynaecology and Fertility Group, University of Auckland, New Zealand.

  Technical support
- The Robinson Institute, University of Adelaide, Australia.

  Access to academic resources

External sources

- None, Other.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

General scope: this review is a part of the review series arising from the same generic protocol. The following sections were adjusted to the main topic of the review as following.

Background: the section on the index test was modified and all the irrelevant information to combination of different testing modalities was removed. The "Rationale" section was updated and now includes a clearer definition of triage diagnostic tests.

Objectives

1. During revision of the literature on the subject, we identified a substantial body of studies looking at the biomarkers, expression of which did not change by presence of endometriosis (no statistically significant difference was found in women with and without the disease). We believe that presenting this type of data, obtained from the adequately designed studies is important for both the clinicians and the researchers in the field, which has been explained in the Background section under "Rationale", in the Methods section under "Criteria for considering studies for this review - Index test" and added to "Objectives" as a secondary objective: '2.To assess the biomarkers which were not affected by endometriosis and hence were unlikely to discriminate between women with and without the disease'.
2. The list of the sources of heterogeneity has been updated.

Methods

1. Criteria for considering studies for this review were updated as following.
   a. Types of studies: we removed the 'cohort' and 'case control' classifications and introduced the concepts of 'single-gate design' and 'two-gate' design'. This was defined as the presence of a single or multiple sets of inclusion criteria by clinical condition or by reference standard. We found this classification more informative in the description of diagnostic studies, all of which are cross-sectional in nature. We limited the inclusion criteria to the studies with a single set of inclusion criteria by reference standard (i.e. all women who underwent abdominal surgery), but included single or multiple sets of inclusion criteria by clinical presentation (i.e. women with suspected endometriosis or other indications for abdominal surgery), referring to these as 'single-gate design' and 'two-gate' design', respectively.
   b. Likewise, we removed the terminology 'prospective studies' and introduced 'studies performed on prospectively collected samples'. This decision was guided by the fact that most diagnostic studies are retrospective in nature, as they aim to compare the result of an index test with the result of a reference standard in the same group of participants, where the groups are classified by the

outcome of the reference standard. Also, the analysis of the index test could have been performed retrospectively in a single batch on stored samples after the prospective collection of samples. The timing of sample collection (before or after surgical treatment of the disease) from a preoperatively recruited population has more impact on the test result than the timing of the laboratory assay. Therefore, we included only studies where the biological sample was collected before the reference surgical procedure, i.e. 'prospectively collected', irrespective of the actual timing of test performance and abandoned labelling studies as 'prospective' or 'retrospective' to avoid confusion. This allowed us to include the studies from well-established high-quality tissue banks using well characterised archived samples as omission of these studies would have resulted in the loss of potentially valuable data. This is presented in the Methods under "Criteria for considering studies for this review". For the combined tests that include imaging modality, only prospectively recruited women and prospectively performed tests were considered eligible.

c. Index tests were modified to pertain only to the combined test of different testing modalities and the table listing the tests of interest (Table 2) was updated accordingly. As a summary review of the series, this review also presents the full list of the tests evaluated in each sister review.

d. Target condition now also includes deep pelvic endometriosis in view of the growing body of literature on this condition as a separate entity and its diagnostic importance to optimise the surgical approach.

e. Spectrum of disease: following an ad hoc observation, we included the studies that involved only selected populations of women with endometriosis (i.e. specific rASRM stages) in view of the emerging evidence on poor correlation of this classification with infertility and pain symptoms. Exclusion of such studies could result in the loss of potentially important diagnostic information from the otherwise eligible publications. Where possible, we aimed to address the impact of the inclusion of these studies in investigations of heterogeneity.

2. Search methods for identification of studies
   a. In the protocol we stated that the grey literature (unpublished studies including conference proceedings and reports) would be identified and defined specific search strategies. In practice, the paucity of relevant data that was available from abstracts made it impossible to apply the selection criteria and methodological quality judgement to these studies. Identification of these types of studies and attempts to obtain the necessary information directly from the study investigators was anticipated to increase the already labour intense work involved in preparation of this review. Therefore, by consensus between the key authors, we removed already identified unpublished studies and did not complete an intended search for unpublished material.
   b. The search strings were updated for all biomarkers excluding imaging (searched separately), applying the same principles as presented in the protocol.
3. Assessment of methodological quality: the QUADAS-2 tool was tailored for the topic of the review. The differences between the original QUADAS-2 tool and the designed for this review are outlined in the relevant section in the Methods.

Analysis

1. The section on statistical methods was amended and tailored to the types of tests included in the review.
2. We performed no sensitivity analyses and no assessment of heterogeneity due to insufficient data.
3. When a test performance was judged against the predetermined diagnostic criteria, only the point estimates of sensitivity and specificity were considered as we believe that presenting these metrics of test performance is the most helpful and informative way to summarise the diagnostic data. We acknowledge that the choice of the most helpful summary is subjective. There are tests where the point estimates did not reach the predetermined criteria, but the confidence intervals (CIs) contain the values above the thresholds for replacement or triage tests. These tests could have diagnostic value if the point values underestimated their diagnostic potential. For the tests where the point estimates reached the criteria for a replacement or triage test but the CIs contained values below the thresholds, point values could have overestimated the diagnostic performance of the test. If the range of the CIs rather than the point estimates of the data are used, the predetermined cut-off becomes meaningless. We did not consider CIs in qualifying the test performance, however, we utilised the CIs in interpreting the reliability of the obtained data.

The authors list and order changed to accurately reflect their contribution to the review.

NOTES

The initially planned single review on the non-invasive tests for diagnosis of endometriosis was split into five smaller reviews in order to facilitate data handling and interpretation, due to the abundance and diversity of the suggested tests. The review was generated from a generic protocol, which was designed for all the reviews in these series. The other reviews from the series include 1) Endometrial biomarkers for the non-invasive diagnosis of endometriosis; 2) Urinary biomarkers for the non-invasive diagnosis of endometriosis; 3) Imaging modalities for the non-invasive diagnosis of endometriosis; 4) Blood biomarkers for the non-invasive diagnosis of endometriosis.

INDEX TERMS

Medical Subject Headings (MeSH)

Aromatase [analysis]; Biomarkers [*analysis]; CA-125 Antigen [blood]; CA-19-9 Antigen [blood]; Endometriosis [*diagnosis] [diagnostic imaging]; Interleukin-6 [blood]; Leukocytes [cytology]; Ovarian Diseases [*diagnosis] [diagnostic imaging]; Pelvis

[diagnostic imaging]; Peritoneal Diseases [*diagnosis] [diagnostic imaging]; Phosphopyruvate Hydratase [urine]; Sensitivity and Specificity; Ubiquitin Thiolesterase [analysis]; Ultrasonography; Vitamin D-Binding Protein [urine]

MeSH check words

Female; Humans