Menopausal status, ultrasound and biomarker tests in combination for the diagnosis of ovarian cancer in symptomatic women

Davenport, Clare; Rai Talapadi, Nirmala; Sharma, Pawana; Deeks, Jon; Berhane, Sarah; Mallett, Sue; Saha, Pratyusha ; Champaneria, Rita ; Bayliss, Susan; Snell, Kym; Sundar, Sudha

DOI:
10.1002/14651858.CD011964.pub2

Cochrane Library

Cochrane Database of Systematic Reviews

Menopausal status, ultrasound and biomarker tests in combination for the diagnosis of ovarian cancer in symptomatic women (Review)

Davenport C, Rai N, Sharma P, Deeks JJ, Berhane S, Mallett S, Saha P, Champaneria R, Bayliss SE, Snell KIE, Sundar S

Davenport C, Rai N, Sharma P, Deeks JJ, Berhane S, Mallett S, Saha P, Champaneria R, Bayliss SE, Snell KIE, Sundar S.
Menopausal status, ultrasound and biomarker tests in combination for the diagnosis of ovarian cancer in symptomatic women.
Cochrane Database of Systematic Reviews 2022, Issue 7. Art. No.: CD011964.
DOI: 10.1002/14651858.CD011964.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Review]

Menopausal status, ultrasound and biomarker tests in combination for the diagnosis of ovarian cancer in symptomatic women

Clare Davenport[1], Nirmala Rai[2a], Pawana Sharma[1b], Jonathan J Deeks[1], Sarah Berhane[3], Sue Mallett[4], Pratyusha Saha[5], Rita Champaneria[6], Susan E Bayliss[7], Kym IE Snell[8], Sudha Sundar[9]

[1]Test Evaluation Research Group, Institute of Applied Health Research, University of Birmingham, Birmingham, UK. [2]School of Cancer Sciences, University of Birmingham, Birmingham, UK. [3]NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham, Birmingham, UK. [4]UCL Centre for Medical Imaging, Division of Medicine, Faculty of Medical Sciences, University College London, London, UK. [5]Medical School, University of Birmingham, Birmingham, UK. [6]Systematic Review Initiative, NHS Blood and Transplant, Oxford, UK. [7]Institute of Applied Health Research, University of Birmingham, Birmingham, UK. [8]Centre for Prognosis Research, School of Primary, Community and Social Care, Keele University, Keele, UK. [9]School of Cancer Sciences, University of Birmingham , Birmingham, UK

[a]Joint first author. [b]Joint first author

Contact: Clare Davenport, c.f.davenport@bham.ac.uk.

Editorial group: Cochrane Gynaecological, Neuro-oncology and Orphan Cancer Group.
Publication status and date: New, published in Issue 7, 2022.

Citation: Davenport C, Rai N, Sharma P, Deeks JJ, Berhane S, Mallett S, Saha P, Champaneria R, Bayliss SE, Snell KIE, Sundar S. Menopausal status, ultrasound and biomarker tests in combination for the diagnosis of ovarian cancer in symptomatic women. *Cochrane Database of Systematic Reviews* 2022, Issue 7. Art. No.: CD011964. DOI: 10.1002/14651858.CD011964.pub2.

ABSTRACT

Background

Ovarian cancer (OC) has the highest case fatality rate of all gynaecological cancers. Diagnostic delays are caused by non-specific symptoms. Existing systematic reviews have not comprehensively covered tests in current practice, not estimated accuracy separately in pre- and postmenopausal women, or used inappropriate meta-analytic methods.

Objectives

To establish the accuracy of combinations of menopausal status, ultrasound scan (USS) and biomarkers for the diagnosis of ovarian cancer in pre- and postmenopausal women and compare the accuracy of different test combinations.

Search methods

We searched CENTRAL, MEDLINE (Ovid), Embase (Ovid), five other databases and three trial registries from 1991 to 2015 and MEDLINE (Ovid) and Embase (Ovid) form June 2015 to June 2019. We also searched conference proceedings from the European Society of Gynaecological Oncology, International Gynecologic Cancer Society, American Society of Clinical Oncology and Society of Gynecologic Oncology, ZETOC and Conference Proceedings Citation Index (Web of Knowledge). We searched reference lists of included studies and published systematic reviews.

Selection criteria

We included cross-sectional diagnostic test accuracy studies evaluating single tests or comparing two or more tests, randomised trials comparing two or more tests, and studies validating multivariable models for the diagnosis of OC investigating test combinations, compared with a reference standard of histological confirmation or clinical follow-up in women with a pelvic mass (detected clinically or through USS) suspicious for OC.

Data collection and analysis

Two review authors independently extracted data and assessed quality using QUADAS-2. We used the bivariate hierarchical model to indirectly compare tests at commonly reported thresholds in pre- and postmenopausal women separately. We indirectly compared tests across all thresholds and estimated sensitivity at fixed specificities of 80% and 90% by fitting hierarchical summary receiver operating characteristic (HSROC) models in pre- and postmenopausal women separately.

Main results

We included 59 studies (32,059 women, 9545 cases of OC). Two tests evaluated the accuracy of a combination of menopausal status and USS findings (IOTA Logistic Regression Model 2 (LR2) and the Assessment of Different NEoplasias in the adneXa model (ADNEX)); one test evaluated the accuracy of a combination of menopausal status, USS findings and serum biomarker CA125 (Risk of Malignancy Index (RMI)); and one test evaluated the accuracy of a combination of menopausal status and two serum biomarkers (CA125 and HE4) (Risk of Ovarian Malignancy Algorithm (ROMA)). Most studies were at high or unclear risk of bias in participant, reference standard, and flow and timing domains. All studies were in hospital settings. Prevalence was 16% (RMI, ROMA), 22% (LR2) and 27% (ADNEX) in premenopausal women and 38% (RMI), 45% (ROMA), 52% (LR2) and 55% (ADNEX) in postmenopausal women. The prevalence of OC in the studies was considerably higher than would be expected in symptomatic women presenting in community-based settings, or in women referred from the community to hospital with a suspicion of OC. Studies were at high or unclear applicability because presenting features were not reported, or USS was performed by experienced ultrasonographers for RMI, LR2 and ADNEX.

The higher sensitivity and lower specificity observed in postmenopausal compared to premenopausal women across all index tests and at all thresholds may reflect highly selected patient cohorts in the included studies.

In premenopausal women, ROMA at a threshold of 13.1 (± 2), LR2 at a threshold to achieve a post-test probability of OC of 10% and ADNEX (post-test probability 10%) demonstrated a higher sensitivity (ROMA: 77.4%, 95% CI 72.7% to 81.5%; LR2: 83.3%, 95% CI 74.7% to 89.5%; ADNEX: 95.5%, 95% CI 91.0% to 97.8%) compared to RMI (57.2%, 95% CI 50.3% to 63.8%). The specificity of ROMA and ADNEX were lower in premenopausal women (ROMA: 84.3%, 95% CI 81.2% to 87.0%; ADNEX: 77.8%, 95% CI 67.4% to 85.5%) compared to RMI 92.5% (95% CI 90.3% to 94.2%). The specificity of LR2 was comparable to RMI (90.4%, 95% CI 84.6% to 94.1%).

In postmenopausal women, ROMA at a threshold of 27.7 (± 2), LR2 (post-test probability 10%) and ADNEX (post-test probability 10%) demonstrated a higher sensitivity (ROMA: 90.3%, 95% CI 87.5% to 92.6%; LR2: 94.8%, 95% CI 92.3% to 96.6%; ADNEX: 97.6%, 95% CI 95.6% to 98.7%) compared to RMI (78.4%, 95% CI 74.6% to 81.7%). Specificity of ROMA at a threshold of 27.7 (± 2) (81.5, 95% CI 76.5% to 85.5%) was comparable to RMI (85.4%, 95% CI 82.0% to 88.2%), whereas for LR2 (post-test probability 10%) and ADNEX (post-test probability 10%) specificity was lower (LR2: 60.6%, 95% CI 50.5% to 69.9%; ADNEX: 55.0%, 95% CI 42.8% to 66.6%).

Authors' conclusions

In specialist healthcare settings in both premenopausal and postmenopausal women, RMI has poor sensitivity. In premenopausal women, ROMA, LR2 and ADNEX offer better sensitivity (fewer missed cancers), but for ROMA and ADNEX this is off-set by a decrease in specificity and increase in false positives. In postmenopausal women, ROMA demonstrates a higher sensitivity and comparable specificity to RMI. ADNEX has the highest sensitivity in postmenopausal women, but reduced specificity. The prevalence of OC in included studies is representative of a highly selected referred population, rather than a population in whom referral is being considered. The comparative accuracy of tests observed here may not be transferable to non-specialist settings. Ultimately health systems need to balance accuracy and resource implications to identify the most suitable test.

PLAIN LANGUAGE SUMMARY

What is the accuracy of different combinations of ultrasound imaging and blood tests to diagnose ovarian cancer in women before and after the menopause?

Why is improving the diagnosis of ovarian cancer important?

Many women diagnosed with ovarian cancer (OC) die from the disease, because it has usually spread outside the tubes/ovaries at the time of diagnosis. Missing OC (a false-negative result) may need major surgery and a lower chance of survival. An incorrect diagnosis of OC (a false-positive result) may result in anxiety, unnecessary further tests and surgery.

What did we aim to do?

We aimed to find out how accurate ultrasounds and blood tests are for diagnosing OC in premenopausal women and postmenopausal women.

What did we study?

We included 59 studies that compared four tests: Risk of Malignancy Index (RMI) (ultrasound and CA125 blood test); Risk of Ovarian Malignancy Algorithm (ROMA) (CA125 and HE4 blood tests); the IOTA Logistic Regression model 2 (LR2) ultrasound and the Assessment of Different NEoplasias in the adneXa model (ADNEX) (CA125 blood test and ultrasound).

What were the main results?

Premenopausal women

The sensitivities (proportion of women *with* OC correctly identified) of ROMA (77.4%), LR2 (83.3%) and ADNEX (95.5%) are higher than RMI (57.2%).

The specificities (proportion of women *without* OC correctly identified) of ROMA (84.3%) and ADNEX (77.8%) were lower than RMI (92.5%) and LR2 (90.4%).

The results indicate that if these tests were to be used in hospital settings in a group of 1000 premenopausal women, of whom 30 (3%) actually have OC:

– for RMI 13 women, for ROMA 7 women, for LR2 5 women and for ADNEX 1 woman would have their cancer missed by the test (false-negative result);

– for RMI 73 women, for ROMA 152 women, for LR2 93 women and for ADNEX 215 women would test positive when they do not have OC (false-positive result).

Postmenopausal women

The sensitivities of ROMA (90.3%), LR2 (94.8%) and ADNEX (97.6%) are higher than RMI (78.4%).

The specificities of ROMA (81.5%) and RMI (85.4%) are higher than LR2 (60.6%) and ADNEX (55.0%).

The results of these studies indicate that if these tests were to be used in hospital settings in a group of 1000 postmenopausal women, of whom 30 (3%) actually have OC:

– for RMI 6 women, for ROMA 3 women, for LR2 2 women and for ADNEX 1 woman would have their cancer missed by the test (false-negative result);

– for RMI 142 women, for ROMA 179 women, for LR2 382 women and for ADNEX 437 women would test positive when they do not have OC (false-positive result).

How reliable are the results?

OC was diagnosed by histology (looking at surgically removed specimens under a microscope) or following up women for one year to see if they remained free of OC. In some studies, women with negative test results were not followed up for long enough to be sure a cancer had not been missed, and some studies excluded women with types of OC that are harder to diagnose. This may make tests appear more accurate than they are in practice.

Who do the results apply to?

Most studies were conducted in European hospitals in women with a confirmed pelvic mass. The occurrence of OC in included studies was much higher than seen in the community and so the accuracy of these tests may be different for women being tested in non-specialist healthcare settings.

What are the implications?

This review suggests that in both pre- and postmenopausal women referred to hospital with a pelvic mass, ADNEX appears to miss the fewest cases of OC and RMI misses the most cases of OC. RMI appears to result in the fewest incorrect diagnoses of OC and ADNEX results in the most incorrect diagnoses of OC. Incorrect diagnoses of OC, when no cancer is present (false-positive test), may result in anxiety, unnecessary further tests and surgery. When choosing which test to use, the potential for missed cancers must be balanced against unnecessary testing and surgery.

How up-to-date is this review?

The review includes studies published up to June 2019.

Summary of findings 1. Summary of findings for menopausal status, ultrasound scan and biomarker tests in pre- and postmenopausal women in secondary care (prevalence ovarian cancer 3%)

Review question	Menopausal status, ultrasound scan and biomarker tests in combination for the diagnosis of ovarian cancer in women with symptoms suspicious for ovarian cancer
Setting	Secondary care
Reference standards	Histology in women who have undergone surgery and clinical follow-up (> 6 months) in women with negative index tests results who do not undergo surgery
Study limitations	For the participant selection domain, 44/59 (75%) studies were at high or unclear risk of bias because of concerns about selective recruitment of women. 58/59 (92%) studies were at high or unclear applicability concern for the participant selection domain because study participants were not symptomatic women. For the index test domain, 9/42 (21%) of ROMA studies, 11/20 (55%) of RMI studies, 2/4 (50%) of ADNEX studies, and 5/5 (100%) of LR2 studies were at high risk of bias because of lack of blinding of the index test or for ROMA studies because of no predefined threshold. Applicability concern was high or unclear for all RMI, ADNEX and LR2 studies because ultrasound was conducted by specialist sonographers or this was unclear. For the reference standard domain, 2/59 studies were at high risk of bias because the minimum length of follow-up for index negatives was not reported or because of lack of blinding. Applicability concern was high or unclear in 50/59 (85%) studies because borderline tumours had been excluded from analysis or classification of borderline tumours for estimation of test accuracy was unclear. For the flow and timing domain, 45/59 (76%) studies were at unclear or high risk of bias because of no information about the interval between the index test and the reference standard or because not all participants receiving an index test received a reference standard.

Population	**Premenopausal women**					
Index test, threshold **Studies (participants)**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Absolute sensitivity difference (95% CI) compared to RMI**	**Absolute specificity difference (95% CI) compared to RMI**	**Consequences in a hypothetical cohort of 1000 women assuming a prevalence of 3%***	
					Number of women who would have their cancer missed (false-negatives) (95% CI)	**Number of women who would test positive when they do not have ovarian cancer (false-positives) (95% CI)**
RMI 200 17 (5233)	57.2 (50.3 to 63.8)	92.5 (90.3 to 94.2)	—	—	13 (11 to 15)	73 (56 to 94)
ROMA 13.1 (± 2)	77.4 (72.7 to 81.5)	84.3 (81.2 to 87.0)	20.2 (12.2 to 28.3);	–8.2 (–11.7 to –4.7); P < 0.0001	7 (6 to 8)	152 (126 to 182)

27 (4463)			P < 0.0001			
LR2 post-test probability ovarian cancer 10% 4 (2843)	83.3 (74.7 to 89.5)	90.4 (84.6 to 94.1)	26.2 (16.2 to 36.2); P < 0.0001	−2.1 (−7.2 to 2.9); P = 0.404	5 (3 to 8)	93 (57 to 149)
ADNEX post-test probability ovarian cancer 10% 4 (1696)	95.5 (91.0 to 97.8)	77.8 (67.4 to 85.5)	38.3 (30.9 to 45.8); P < 0.0001	−14.8 (−24.0 to −5.5); P = 0.002	1 (1 to 3)	215 (141 to 316)
Population	**Postmenopausal women**					
Index test, threshold **Studies (participants)**	**Sensitivity (95% CI)**	**Specificity** **(95% CI)**	**Absolute sensitivity difference (95% CI) compared to RMI**	**Absolute specificity difference (95% CI) compared to RMI**	**Consequences in a hypothetical cohort of 1000 women assuming a prevalence of 3%***	
					Number of women who would have their cancer missed (false-negatives) (95% CI)	**Number of women who would test positive when they do not have ovarian cancer (false-positives) (95% CI)**
RMI 200 17 (4369)	78.4 (74.6 to 81.7)	85.4 (82.0 to 88.2)	—	—	6 (5 to 8)	142 (114 to 175)
ROMA (27.7 (± 2)) 13 (2002)	90.3 (87.5 to 92.6)	81.5 (76.5 to 85.5)	11.9 (7.6 to 16.3); P < 0.0001	−3.9 (−9.4 to 1.5); P = 0.157	3 (2 to 4)	179 (141 to 228)
LR2 post-test probability ovarian cancer 10% 5 (2157)	94.8 (92.3 to 96.6)	60.6 (50.5 to 69.9)	16.4 (12.3 to 20.5); P < 0.0001	−24.8 (−35.1 to −14.5); P < 0.0001	2 (1 to 2)	382 (292 to 480)
ADNEX post-test probability ovarian cancer 10% 4 (1365)	97.6 (95.6 to 98.7)	55.0 (42.8 to 66.6)	19.2 (15.4 to 23.1); P < 0.0001	−30.4 (−42.9 to −17.9); P < 0.0001	1 (0 to 1)	437 (324 to 555)

*Estimate of disease prevalence (pretest probability) reflecting the NICE threshold for cancer referral from generalist to specialist settings in the UK (NICE 2017). Note this is considerably lower (3%) compared to the prevalence of ovarian cancer in included studies in the review (16% to 55%).
ADNEX: Assessment of Different NEoplasias in the adneXa model; CI: confidence interval; LR2: Logistic Regression model 2; RMI: Risk of Malignancy Index; ROMA: Risk of Ovarian Malignancy Algorithm.

BACKGROUND

The estimated lifetime risk of being diagnosed with ovarian cancer (OC) is 1 in 50 (2%) for females born after 1960 in the UK Office for National Statistics (ONS) (Office for National Statistics 2016; Smittenaar 2016). Increasing age is a risk factor for OC; with incidence rates highest in females between 75 and 79 years of age (Cancer Research UK 2017).

OC is the most common cause of mortality among all gynaecological cancers. In 2018, 295,414 women were diagnosed with OC and 184,799 women died worldwide (Bray 2018). The high case fatality rate is largely attributed to the advanced stage at diagnosis in most women with OC. Although overall survival is 35% at 10 years, one-year survival is only 51% in stage 4 disease, in comparison to 99% in stage 1 disease (Office for National Statistics 2016). Lack of awareness and recognition of pertinent symptoms and signs by patients and physicians is considered one of the main factors contributing to a delay in diagnosis. Diagnosis of OC is challenging because of variable presentation, the non-specific nature of symptoms (Fitch 2002), and low prevalence. The prevalence of OC in primary care has been estimated as 0.023% (Bankhead 2005; Hamilton 2009), whilst recent hospital audits suggest a prevalence of OC in secondary care of 10% (Rai 2015). The prevalence of OC in women undergoing surgery for ovarian pathology in tertiary care settings is in the region of 30% (Nunes 2014; Timmerman 2010; Timmerman 2016).

Diagnosis of OC in premenopausal women poses additional challenges. Most ovarian tumours detected in premenopausal women tend to be benign; only 1 in 1000 symptomatic ovarian cysts are malignant, increasing to 3 in 1000 at age 50 years (RCOG 2011).

Advances in surgical practice and chemotherapy in recent years have slightly improved survival, but a diagnosis of OC continues to be associated with a high mortality, largely attributed to an advanced stage at diagnosis.

Target condition being diagnosed

OC has various subtypes including, epithelial ovarian cancers (EOC), germ cell tumours, stromal cell tumours, metastatic cancers (from other primary sites) and tumours of low malignant potential (LMP) also known as borderline tumours. EOC are the most common type of OC in both pre- and postmenopausal women. More than 90% of OCs in postmenopausal women and 80% to 85% of OCs in premenopausal women are EOC; in premenopausal women, germ cell tumours account for 15% to 20% of OCs. Within the EOC group, high-grade serous carcinoma (HGSC) is the most common histological type. Other common epithelial histological types are mucinous, clear cell and endometrioid (Shepherd 2000). Morphological and genetic studies have helped to improve our understanding of ovarian carcinogenesis and tumour behaviour according to different histology types. The distal fallopian tube is the origin for serous ovarian carcinomas and ovarian clear cell cancers; the origin of endometrioid OCs has been linked to endometriosis (Wiegand 2010). A dualistic model has been proposed based on the behaviour of tumours (Shih 2004). Type 1 tumours are indolent and present at an early stage; a typical example is endometrioid cancer. Type 2 tumours are aggressive, high-grade carcinomas, most often diagnosed at an advanced stage; a typical example is high-grade serous OC. Type 1 and Type 2 tumours display markedly different and distinct genetic patterns (Cho 2009). This advancement in understanding has major research implications, especially regarding the role of biomarkers, either alone, or as part of a composite index tests, in the management of OC.

This review is concerned with primary OC of all histological types and stages, including borderline tumours. Metastatic disease (cancer found in the ovary, but originating in another organ) is outside the remit of this review.

Index test(s)

For the purpose of this review, combination tests are defined as tests which combine measures from more than one type of clinical information (e.g. age or menopausal status), biomarkers and ultrasound scan (USS) in any combination, and in any order. Table 1 provides details of index tests considered eligible for inclusion in this review.

Clinical information

The most important risk factor for OC is a family history of breast cancer or OC (American Cancer Society 2016). Approximately 15% to 20% of OC is caused by an inherited genetic mutation in genes such as *BRCA1* and *BRCA2* (Walsh 2011). For women with a *BRCA1* or *BRCA2* genetic mutation, the lifetime risk of ovarian, fallopian tube or peritoneal cancer is approximately 41% to 46% for *BRCA1* and 10% to 27% for *BRCA2* by age 70 years (Lancaster 2015). The importance of menopausal status as a risk factor for OC is a function of the increased risk of cancer associated with increasing age (Cancer Research UK 2017). Although ovarian cysts are more common in premenopausal women, due to the physiological function of the ovary, most are benign functional cysts that resolve spontaneously. Some persistent benign cysts, caused by abnormal growth of cells such as endometriosis, fibromas and cystadenomas, may require intervention, but the risk of malignancy is low at 1/1000 women compared to 3/1000 women at age 50 years (RCOG 2011).

Biochemical markers

Biochemical markers, also known as biomarkers, are substances secreted or shed by tumours into surrounding blood and body fluids and expressed in abnormal tissues. Biomarkers may be uniquely specific for some tumour subtypes, or non-specific. It has been noted that levels of some tumour markers may begin to rise as early as three years prior to diagnosis (Anderson 2009).

The most commonly used biomarker for OC is CA125, which is raised in many benign and physiological conditions (Moss 2005; Posadas 2004). CA125 operating at a threshold of 30 units/mL has a sensitivity of 81% and specificity of 75% for distinguishing benign from malignant tumours in mixed pre- and postmenopausal populations with adnexal masses (growths that occur in or near the uterus, ovaries, fallopian tubes and the connecting tissues) (Jacobs 1990). However, CA125 has a low sensitivity (50%) for early-stage OC (Jacobs 1989), and reduced specificity in premenopausal women.

The serum tumour marker Human Epididymis protein (HE4) is a glycoprotein belonging to the Whey acidic protein family (Hellstorm 2003), and was approved as a biomarker for OC by the US Food and Drug Administration (FDA) in 2008. HE4 is elevated in 8% of benign conditions compared to 29% for CA125 and hence has the potential to improve specificity especially in premenopausal

women (Moore 2012). HE4 secretion increases with age (Moore 2012), and is affected by different cellular types of OC, highest in endometrioid (100%), 93% of serous, 50% of clear cell and not elevated in mucinous types (Drapkin 2005). HE4 has similar sensitivity, but improved specificity compared to CA125 for OC, particularly in premenopausal women (Ferraro 2013; Holcomb 2011).

Ultrasound scan

USS enables visualisation of morphological details of ovarian cysts. The diagnostic potential of USS has improved with advancing technology and the availability of transvaginal ultrasound (TVS), 3D ultrasound and Doppler techniques to characterise blood flow. However, the use of ultrasound to characterise lesions is influenced by interference from surrounding tissue, variability of the macroscopic features and the subjective nature of interpretation that is operator-dependent. Various scores have been developed to make USS more objective (Geomini 2009). Morphological features, such as size, presence of bilateral lesions, presence and thickness of septum, presence of solid areas, excrescences and papillary structures within tumours, presence of metastases (spreading of a tumour to other parts of the body), presence of ascites (abnormal accumulation of fluid in the abdomen) and Doppler measurements of blood flow, have been combined in various ways.

The 'U' score records the presence of bilateral lesions, multilocularity, solid areas, metastases or ascites, where U = 0 indicates the absence of any of these features; U = 1 indicates the presence of any one of these features and U = 3 indicates the presence of two or more of these features (RCOG 2011). The U score is a component of the Risk of Malignancy Index (RMI) (see below). The International Ovarian Tumour Analysis (IOTA) proposed more-recent USS-based models as having better diagnostic accuracy in the preoperative evaluation of ovarian tumours than the U score, including the Logistic Regression model 2 (LR2) (Kaijser 2014).

Test combinations

OC is a heterogeneous tumour and consequently it is likely that a combination of tests (clinical information, USS and biomarkers) has the potential to improve diagnostic accuracy over any single test (clinical assessment, biomarker or imaging) alone. Several composite tests have subsequently been developed.

RMI is derived by multiplying the USS score (0 to 3) (1 point for each of the following characteristics: multilocular cysts, solid areas, metastases, ascites and bilateral lesions), menopausal status and CA125 in units per millilitre (RMI = U × M × CA125). RMI is the most widely used combination of tests. Four different versions of RMI (I to IV) have been developed, which differ in scores attributed to the result of each test component (Atkurk 2011). In addition, RMI IV includes a score for the size of the tumour. RMI I is the version currently recommended by the National Institute for Health and Care Excellence (NICE) (NICE 2011) and the Royal College of Obstetrics and Gynaecology (RCOG) (RCOG 2016), in both pre- and postmenopausal women. In this review, we included only RMI version I and use the term RMI as synonymous with RMI I.

Risk of Ovarian Malignancy Algorithm (ROMA) combines menopausal status and the biomarkers CA125 and HE4 in a multivariable model to estimate the probability (%) of malignancy in an adnexal mass. In subgroup analysis, the accuracy of ROMA was better for EOC compared to all OCs combined, in mixed populations compared to populations segregated by menopausal status (pre- or postmenopausal) and in late- compared to early-stage disease (Li 2012).

Two test combinations that integrate clinical information and USS findings to estimate the probability (%) of malignancy in an adnexal mass include the LR2 and (Assessment of Different NEoplasias in the adneXa model) ADNEX multivariable models. LR2 (superseding LR1) is a multivariable model to estimate the probability (%) of malignancy in an adnexal mass. The model combines clinical information (age) and USS findings (presence of ascites, presence of blood flow within a solid papillary projection, maximum diameter of the solid component of a mass, irregular cyst walls and the presence of acoustic shadows) (Timmerman 2010). The ADNEX multivariable model has been developed to estimate the probability of malignancy in an adnexal mass. The model combines clinical information (age, healthcare setting), USS characteristics (maximum mass diameter, proportion of solid tissue, number of papillary projections, presence of more than 10 cyst locules (cavities within an organ), acoustic shadows, presence of ascites) and CA125 levels and shows promise in the preoperative discrimination of benign, borderline, early and advanced malignancies in ovarian masses (van Calster 2014).

Clinical pathway

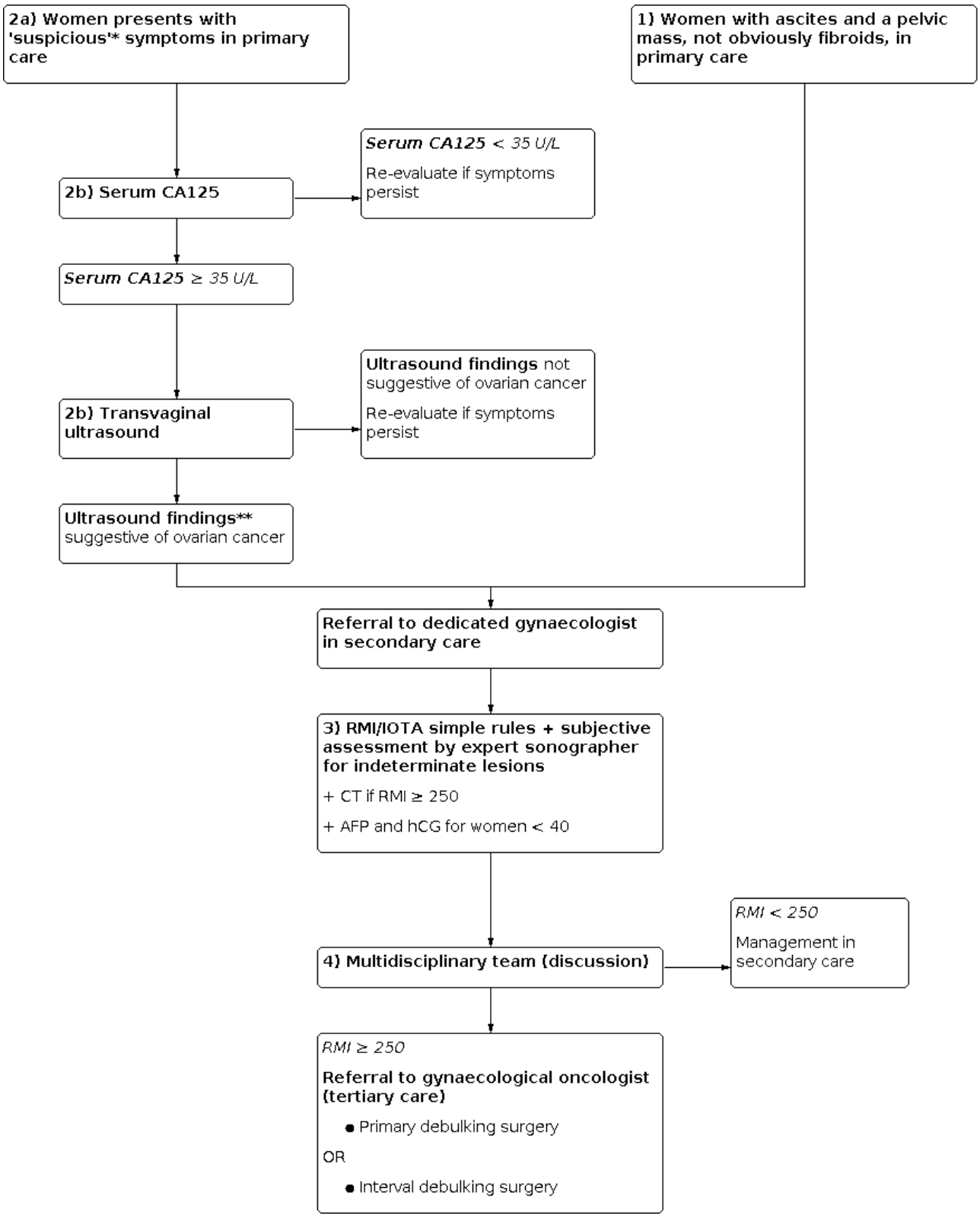
This review is concerned with women presenting with symptoms or signs (or both) in whom OC is being considered as a differential diagnosis. It is now recognised that women with OC may experience symptoms for a variable length of time prior to diagnosis (Hamilton 2009). Symptoms associated with OC include: abdominal bloating and distension; loss of appetite; early satiety; abdominal and pelvic pain; urinary urgency and frequency; vaginal and rectal bleeding; and change in bowel habit (constipation/diarrhoea).

In the UK, women with symptoms suspicious for OC may present in a generalist setting (primary care/family practice), or to hospital settings (secondary care or tertiary care (specialist gynaecological oncology units)). Symptoms should prompt investigations including the serum biomarker CA125, an USS, or both to determine whether an adnexal mass is present and the degree of suspicion for OC. It is recommended that women with a high index of suspicion for OC (a positive index test result) are referred to a gynaecological oncologist (tertiary care) for further management whereas those with a low index of suspicion for OC (a negative index test result) are referred to a designated gynaecologist in secondary care. International guidelines differ on the types of test and test positivity thresholds to used.

In the UK, NICE and RCOG recommend the following clinical pathway (NICE 2011; Figure 1).

Figure 1. UK recommended clinical pathway based on NICE and RCOG guidance *'Suspicious' symptoms: persistent (> 12 times per month) abdominal distension or bloating; early satiety/loss of appetite; urinary symptoms; abdominal or pelvic pain, weight loss; fatigue; change in bowel habit. **Ultrasound findings suggestive of ovarian cancer: laterality (any imbalance between masses observed in left compared to right ovary), multilocularity, solid areas, free fluid and distant metastasis. AFP: alpha fetoprotein; CT: computed tomography; hCG: human

chorionic gonadotrophin; IOTA: International Ovarian Tumour Analysis; NICE: National Institute for Health and Care Excellence; RCOG: Royal College of Obstetrics and Gynaecology; RMI: Risk of Malignancy Index.



- 1. Women with suspicious findings on clinical examination:

- women with ascites and a pelvic mass that is not obviously fibroids on clinical examination in a primary care setting should be immediately referred to secondary care.

- 2a. Women with suspicious symptoms:
  - women with persistent presence (more than 12 times per month) of abdominal distension or bloating, early satiety or loss of appetite, increased urinary urgency or frequency, and abdominal or pelvic pain, especially if aged over 50 years or women over 50 years presenting with unexplained weight loss, fatigue and change in bowel habit (symptoms suggestive of irritable bowel syndrome are rarely first diagnosed in women aged over 50 years).
- 2b. Women with suspicious symptoms should receive additional investigations: serum biomarker CA125 should be performed and, if 35 IU/mL or greater, a TVS scan should also be performed prior to referral to secondary care. Women with a high CA125 and presence of an adnexal mass on TVS scan should be urgently referred (within two weeks) to secondary care.
- 3. Once in secondary care, an algorithm combining menopausal status, USS features of the pelvic mass (laterality, multilocularity, solid areas, free fluid and distant metastasis) and the CA125 level is used to calculate the RMI I score. Alternatively, following referral from primary care, women may undergo USS as per IOTA criteria (RCOG 2016) TVS examination for a specific set of morphological features used to determine the malignant potential of a pelvic mass and, in the case of a mass which is indeterminate following IOTA assessment, a subjective assessment by an expert USS examiner (RCOG 2016).
- 4. Following either RMI or IOTA assessment and additional tests dictated by a woman's age (40 years or less: human chorionic gonadotrophin (hCG) and alpha fetoprotein (AFP) to detect germ cell tumours; or RMI score of 250 or greater: computed tomography (CT)), a multidisciplinary review team (MDT) is used to triage women for referral to a either a general gynaecologist (secondary care) or a gynaecological oncologist (tertiary care).

In the UK, it is estimated that 28% of women are referred via the two-week wait pathway (on the basis of symptoms and signs defined by guidelines as suspicious for cancer), 38% via general practitioner referral to gynaecologists, 26% via outpatients, 12% via other than gynaecology and 29% of women are diagnosed following an emergency presentation (Ellis-Brookes 2012). One multicentre study in the UK demonstrated variable adherence to the recent NICE guidance regarding the tests used and the impact of results on patient management (Rai 2015).

The American College of Obstetrics and Gynaecology recommends TVS as the initial test of choice if physical examination suggests the presence of an adnexal mass (ACOG 2016). Following TVS, referral to a gynaecological oncologist (tertiary care) is recommended in the presence of:

- elevated CA125 in combination with one or more of the following: a suspicious clinical history; suspicious TVS findings; elevation of other biomarkers; or
- an elevated risk score following assessment with LR2, RMI (OVA 1) or ROMA.

Referral to tertiary care is recommended for women suspected of having a germ cell tumour: elevated inhibin A/B, beta hCG, AFP, or L-lactate dehydrogenase.

No pan-European guideline for the investigation and management of suspected OC exists although variation in practice is recognised (Ledermann 2013).

Prior test(s)

As a minimum, women who are being considered for testing with the index tests because of a suspicion of OC will present with self-assessed symptoms. In addition, women may have had one or more clinical assessment (history and examination), biomarker tests and USS, depending on the point in the clinical pathway they present for testing with the index test.

Role of index test(s)

The index tests are used to decide whether women presenting with symptoms or signs (or both) suspicious for OC should receive further investigation and management in secondary care or specialist gynaecological oncology units (tertiary care).

Alternative test(s)

This review is concerned with initial investigations to diagnose OC that would be applicable in generalist and secondary-care settings. Combination tests including CT, magnetic resonance imaging (MRI), positron emission tomography (PET) and other complex imaging techniques are therefore beyond the scope of this review.

Four different versions of RMI (I to IV) have been developed (Atkurk 2011), which differ in scores attributed to the result of each test component. In addition, RMI IV includes a score for the size of the tumour. RMI I is the version currently recommended by NICE and the RCOG in both pre- and postmenopausal women and is the version of RMI that will be evaluated by this review (NICE 2011; RCOG 2016).

Rationale

The non-specific nature of symptoms associated with OC and the high prevalence of ovarian cysts of uncertain significance (30% of females with regular menstruation, 50% of females with irregular menstruation and 6% of postmenopausal females) (Duklewski 2009), continues to pose problems for early and accurate diagnosis. Combining different test types has the potential to improve accuracy over one test type used alone, but the most accurate combination of tests has yet to be determined. There is also a need to understand how test accuracy is influenced by patient characteristics so that test combinations can be appropriately targeted.

As part of a scoping review, 10 original systematic reviews were identified up to 2021 (Chacon 2019; Dodge 2012; Fakhar 2018; Geomini 2009; Kaijser 2014; Li 2012; Meys 2016; NICE 2011; Stukan 2015; Wang 2014). Six of the 10 reviews included ROMA, seven RMI and four LR2. The search date of the most recent review was 2018 (Chacon 2019). None of the reviews included ADNEX. Two reviews compared ROMA and RMI (Chacon 2019; Stukan 2015), and four compared RMI and LR2 (Dodge 2012; Kaijser 2014; Meys 2016; Stukan 2015), whilst six reviewed only single tests. Four of 10 reviews did not present results separately for pre- and

postmenopausal women. Nine of 10 reviews undertook meta-analysis, but only five used appropriate statistical methods.

OBJECTIVES

To establish the accuracy of combinations of menopausal status, ultrasound scan (USS) and biomarkers for the diagnosis of ovarian cancer in pre- and postmenopausal women and compare the accuracy of different test combinations.

Secondary objectives

We planned to investigate the following sources of heterogeneity.

Population

- Clinical setting (generalist/primary care/community/family practice) versus specialist setting (cancer unit/cancer centre/gynaecological oncology)
- Menopausal status (premenopausal versus postmenopausal)

Index tests

- Test positivity threshold
- Experience of the USS test operator (general sonographers versus specialist interest)

Target condition

- Histological subtype

Study quality

- For study participants not receiving surgery following a negative index test result (where clinical follow-up rather than histology is used as a reference standard for index test negatives): 12 months' follow-up versus less than 12 months' follow-up

METHODS

Criteria for considering studies for this review

Types of studies

We included diagnostic case-control studies (providing the control arm included women with benign ovarian pathology and these could be disaggregated from any healthy controls); diagnostic cross-sectional studies (retrospective and prospective data collection). We anticipated that in view of the low prevalence of OC, the majority of cross-sectional studies would recruit women who had already undergone the reference standard and index test results would be ascertained retrospectively. We also included studies externally validating multivariable models for the diagnosis of OC. We included comparative diagnostic test accuracy studies of any design (within-person or between-person comparisons). Studies were eligible if there were sufficient data to extract 2 × 2 tables on diagnostic test performance. We allowed inclusion of studies not providing verification of index test negatives where 2 × 2 tables could be constructed by imputation using setting-specific prevalence estimates. However, we did not identify eligible studies where index test negatives were not verified.

Participants

Women aged 18 years or older, irrespective of menopausal status. We excluded studies restricted exclusively to populations under 18. We excluded studies restricted to pregnant women, or women with a previous history of OC.

Prior tests

This review is concerned with women in whom a diagnosis of OC is suspected (i.e. women with symptoms or signs suggestive of OC). As a minimum, women should have self-referred to a healthcare professional on the basis of the presence of symptoms. Individual components of the test combinations (index tests) included in this review may be used alone in both generalist and specialist settings and so at the time women receive an index test, in addition to presentation with symptoms and signs, they may have had prior testing with one or more testing with one or more biomarkers or imaging with USS. We excluded studies explicitly describing included participants as asymptomatic, for example where the index test was being applied as a screening test, or where studies explicitly included asymptomatic participants and these could not be disaggregated from participants who were symptomatic. Where the prior presence of symptoms or signs was unclear or not reported, studies were included and this was reflected as part of the quality assessment of included studies (QUADAS-2) in the patient applicability domain.

Index tests

We included the following index tests in use in clinical practice at the time of undertaking our searches: any combination (two or more of the following test types): RMI (menopausal status, CA125 and USS examination); ROMA (menopausal status, CA125 and HE4), and the multivariable models LR2 and ADNEX (menopausal status and USS examination) (Table 1). We included studies where USS examination as part of RMI, LR2 and ADNEX was conducted by ultrasonographers with any experience: general sonographers or those with specialist training.

Target conditions

OC, all stages and types. We excluded studies where only one type of ovarian pathology was reported with the exception of EOC, as this is the most common (greater than 90% in postmenopausal women) of the OCs and is associated with the highest mortality. We excluded studies concerned exclusively with recurrent OC, OC which was metastatic from another primary cancer site, and studies where it was not possible to disaggregate participants with primary OC from metastatic or recurrent disease.

Reference standards

Histology in women who have undergone surgery and clinical follow-up in women with negative index test results (suggestive of no OC) who do not undergo surgery. For studies using clinical follow-up, the length of follow-up was considered as part of quality assessment; a minimum of one year of follow-up was considered of higher quality compared to less than one year of follow-up. We planned to investigate length of follow-up as a potential source of heterogeneity.

Search methods for identification of studies

Electronic searches

Original searches were conducted in 2015 to support a generic protocol for four separate reviews: USS, biomarkers, symptom scores and test combinations for the diagnosis of OC. With the exception of the symptom and symptom score search strategy, a

date restriction was applied (1991 onwards) to ensure applicability to current technology. For the symptom search strategy a date restriction of 2009 was applied, reflecting the existence of a comprehensive review of symptoms for the diagnosis of OC (NICE 2011). The 2015 strategies were designed to run across a range of databases: the Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE and MEDLINE In Process (Ovid), Embase (Ovid), CINAHL (EBSCO), the *Cochrane Database of Systematic Reviews* (CDSR), Database of Abstracts of Reviews of Effects (DARE), Health Technology Assessment Database (HTA) and SCI Science Citation Index (ISI Web of Knowledge).

We updated the search strategy in June 2019 specifically for this test combination review. The 2019 searches were a targeted update of evidence about RMI, ROMA, LR2 and ADNEX as these test combinations had emerged in the intervening period as the main contenders for use in clinical practice. For pragmatic reasons we restricted databases to MEDLINE and MEDLINE In Process (Ovid) and Embase (Ovid) for the 2019 update, combining terms for OC with terms to capture the index tests or their components (biochemical markers, symptom scores and USS) that were used in the original 2015 searches. The 2019 search was developed iteratively and evaluated for its performance in detecting key articles already deemed eligible for inclusion post-2015. Specifically, the following changes were made between the 2015 and 2019 search strategies to reflect changes in the review scope: the 2019 search strategy additionally included terms for the index tests of current clinical interest: RMI, ROMA, LR2 and ADNEX; used a reduced range of terms used to describe symptoms and symptom scores (as symptoms are not a major component of the index tests of current interest), and used a reduced range of biomarker terms reflecting those contained in the index tests of current interest. Changes were also made to terms used to describe the target condition (OC) in line with changes in the description of OC as a disease of the adnexa, rather than being a disease of tubal or ovarian origin. The search strategy used for the original 2015 searches as well as the 2019 targeted updated search strategy are shown in Appendix 1 and Appendix 2.

No language restrictions were applied.

Searching other resources

To identify ongoing and unpublished studies, we searched the following trials registers and conference abstracts and proceedings without date restrictions as part of the 2015 search strategy: ClinicalTrials.gov, UK Clinical Research Network Study Portfolio Database (UKCRN) and WHO International Clinical Trials Registry Platform (ICTRP). We searched conference proceedings from the European Society of Gynaecological Oncology (ESGO), International Gynecologic Cancer Society (IGCS), American Society of Clinical Oncology (ASCO) and Society of Gynecologic Oncology (SGO), supplemented by searches of the ZETOC and Conference Proceedings Citation Index (Web of Knowledge). For both the 2015 and 2019 search strategies, we drew on reference lists of existing systematic reviews and guidelines identified in the electronic searches as a source of primary studies.

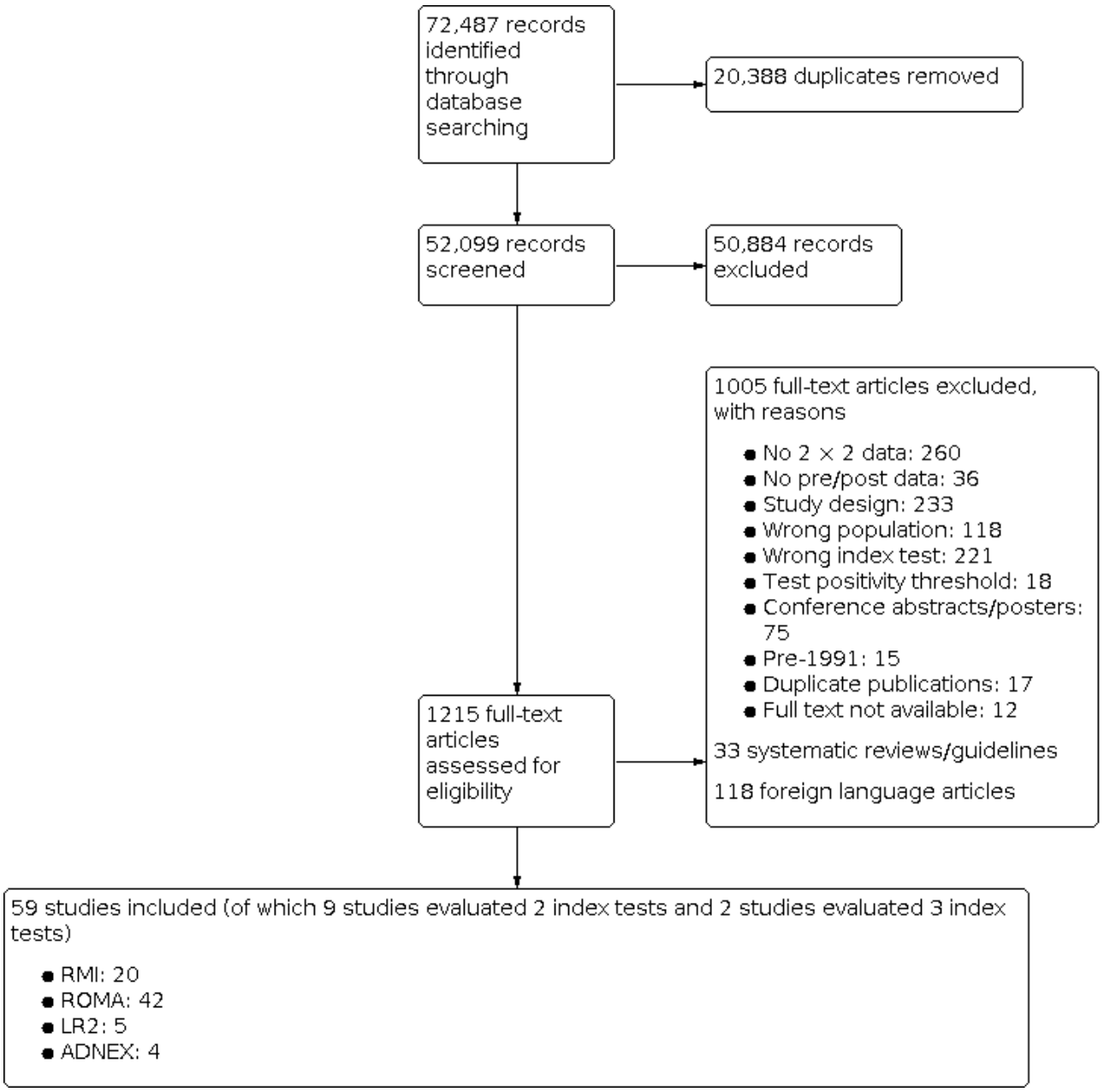
Data collection and analysis

Search results were managed in EndNote. After removal of duplicates, two review authors (from NR, RC, PSh, PSa) independently carried out study selection by reading the titles and abstracts and excluded obviously irrelevant studies at this stage. Two review authors (from NR, RC, PSh, PSa) independently read the full text of remaining studies. A third review author (CD, SS) resolved disagreements. Two review authors (NR, PSh, CD) independently extracted data into 2 × 2 tables and assessed quality. Another review author (RC or CD) double-checked characteristics of 30% of the studies. We resolved disagreements by discussion.

Selection of studies

We reviewed unique titles and abstracts against predefined selection criteria to select potentially relevant studies for full-text review. The results of the selection process and reasons for exclusion are documented and summarised using a PRISMA flow diagram (Figure 2).

Figure 2. PRISMA study flow diagram. ADNEX: Assessment of Different NEoplasias in the adneXa model; LR2: Logistic Regression Model 2; RMI: Risk of Malignancy Index; ROMA: Risk of Ovarian Malignancy Algorithm.



Data extraction and management

We used a predefined data collection form to extract the following data into an Excel database prior to entry into Review Manager 5 (Review Manager 2014): study design; country; setting; single or multicentre; method of recruitment; reasons for exclusion; number of participants; number of women with a diagnosis of OC and borderline ovarian tumours; age; menopausal status (directly or using age over 50 years or history of previous hysterectomy as a proxy for postmenopausal status); prior tests; index tests and index test threshold(s); expertise of index test operator (for symptoms and USS); reference standard (including where relevant duration of follow-up); stage, and histological subtype of OC. Either a clinician (NR) or review author (PSh, RC, CD) extracted data to derive a 2 × 2 table for each study; either a methodologist or statistician (CD, JD, SB) checked data.

Assessment of methodological quality

Quality assessment was undertaken using the QUADAS-2 checklist tailored according to the topic and detailed in Appendix 3 (Whiting 2011).

Tailoring of QUADAS-2 to the clinical topic required consideration of the following.

Patient selection domain

Studies were considered at high risk of bias if they excluded certain types of malignant or benign pathology that is known to affect the accuracy of index tests specifically for detecting primary OC. Examples include endometriosis (which, for example, causes a raised serum CA125) and borderline ovarian tumours (which are managed surgically, similar to malignant tumours, but may result in a negative index test result). Additionally, restricting populations by age was considered to place studies at high risk of bias because an increase in age is associated with a change in disease spectrum. For example, EOC is more prevalent in older women and germ cell tumours are more prevalent in younger women. It has also been shown that index test performance differs in different histological subtypes of OC and at different stages of malignancy Kobayashi 2012).

Menopausal status is a risk factor for OC. In addition the spectrum of disease (the type and severity of OC and the range of differential diagnoses) observed in postmenopausal women are different to those of premenopausal women. For example, in premenopausal women, the normal menstrual cycle and benign pathology, such as endometriosis, can result in false-positive test results. Therefore, we considered distinguishing test performance in pre- and postmenopausal women an important feature of studies. For this reason, the quality of studies that stratified test results by menopausal status is presented separately.

The target population for this review was symptomatic women receiving index tests because of a suspicion of OC on the basis of clinical history and examination. Therefore, studies were considered of high applicability concern if women were asymptomatic, and were selected for testing with index tests in secondary or tertiary care, following prior testing with one or more biomarker or USS.

Index test domain

The review included composite index tests comprised at least two of three different test types: clinical information (menopausal status), biochemical testing and USS examination. Studies were considered at high risk of bias if the USS component of index tests was not conducted blind to the results of other index test components (biochemical markers and clinical assessment). Similarly, studies were considered at high risk of bias if the USS component of composite index tests was not conducted and interpreted blind to the disease status/reference standard result. Studies that did not prespecify the test positivity threshold were considered at high risk of bias because this usually results in over-optimistic test accuracy estimates that are not replicable outside of the study sample. For quality assessment of index tests based on multivariable models (LR2 and ADNEX), QUADAS-2 was tailored by adding items taken from the PROBAST risk of bias tool for prognostic studies (Wolf 2019). These items were whether all model components and thresholds were prespecified and whether individual test components were assessed in a similar way (e.g. in similar healthcare settings or by individuals with similar levels of expertise). Assessment of applicability of index tests comprised consideration of whether the expertise of clinicians undertaking clinical assessment and USS examination was representative of a generalist setting.

Reference standard and target condition domain

We considered histological diagnosis or clinical follow-up for a minimum of 12 months as likely to classify correctly the target condition (therefore a low risk of bias). In studies using clinical follow-up, risk of bias was considered high if follow-up was less than six months. Concerning the applicability of the target condition, as defined by the reference standard; assessments were based on how authors had dealt with borderline tumours in their analysis and the implications this had for meta-analysis. Within the constraints of a 2 × 2 table and reflecting current clinical practice, we considered that borderline tumours should be classified as malignant for the purposes of estimation of test accuracy. Thus studies reporting results allowing grouping of borderline tumours with malignant for the purpose of meta-analysis were considered of low-applicability concern.

Flow and timing domain

We considered risk of bias high if the interval between index test and reference standard application was more than three months.

Statistical analysis and data synthesis

Summary

Exploratory analyses included plotting estimates of sensitivity and specificity grouped by test threshold on Forest plots and in summary ROC (receiver operating characteristic) plots.

Analyses were conducted in Stata version SE 17.0 (StataCorp 2019) and SAS software (version 9.4) (SAS 2015). Where there were adequate data available and it was considered reasonable to pool results, we performed meta-analyses using hierarchical models using the NLMIXED procedure in SAS (SAS 2015). Where meta-analysis was not considered appropriate due to clinical or methodological heterogeneity, or in the case of fewer than three studies, we used narrative synthesis.

Estimation of the accuracy of individual index tests

Since the characteristics measured by index tests could be extracted as 2 × 2 tables reported at common index test thresholds, we used the bivariate model including random effects (Chu 2006; Reitsma 2005). To estimate average sensitivity and specificity at fixed thresholds, we performed the analysis of each index test version by first restricting to studies that reported thresholds recommended in guidelines or used in clinical practice (or both), and second to those thresholds most commonly reported across included studies. In addition, for ROMA, we included studies using thresholds ± 2 units around the most commonly reported thresholds. We excluded thresholds based on particular values of sensitivity and specificity where no threshold in terms of index test operation was reported for the values of sensitivity and specificity used. We used random-effects univariate analyses (which ignore any correlation between sensitivity and specificity) where pooling was an appropriate approach but bivariate models failed to converge.

Comparison of index tests

In order to maximise use of data across studies using different thresholds, we undertook indirect comparisons of index tests by fitting HSROC models and estimating sensitivity at fixed vales of specificity (80% and 90%), reflecting clinical consensus about an acceptable false-positive rate (RCOG 2016). To illustrate the

comparative accuracy of index tests at specific test-operating thresholds that could be applied in clinical practice, we also undertook indirect comparisons of index tests using bivariate hierarchical models.

For the HSROC analysis (Rutter 2001), we used a covariate for test type and estimated a summary ROC curve for each index test across all included thresholds. Each included study contributed one threshold to the summary ROC curve. Where an individual study reported more than one threshold, we selected the most commonly reported threshold for that index test across all included studies for the meta-analyses. The selection of one threshold per study was only necessary for ROMA studies where the threshold pairs 31.1 (± 2 units) and 27.2 (± 2 units) were the most commonly reported across studies. Summary ROC curves which have a common shape were fitted to the data. We performed estimation of differences in accuracy using the NLMIXED procedure in Statistical Analysis System (SAS 2015) and the metandi macro (Takwoingi 2010). We computed P values for the difference in accuracy for each test compared to RMI (RMI being the test combination currently in routine use in the UK in both pre- and postmenopausal women) using Wald tests. We reported the difference in sensitivities at fixed specificities of 80% and 90% for each index test version compared to RMI with 95% confidence interval (CI).

For the bivariate hierarchical analysis, we undertook a comparison of index tests at the single most commonly reported threshold across studies, including a covariate for test type. Absolute differences in sensitivity/specificity and the corresponding P values for each pair-wise test comparison were reported from the model. Bivariate models were fitted using the *meqrlogit* command in Stata. Where appropriate, models were simplified by setting near-zero variance estimates of the random effects to zero (Takwoingi 2017). In cases where both random effects were set to zero, a fixed-effect logistic regression was fitted using the *blogit* command. Absolute differences in sensitivities/specificities and P values were derived from bivariate models using the *nlcom* command in Stata. This computes point estimates and standard errors using the delta method. We used random-effects univariate analyses (which ignore any correlation between sensitivity and specificity) where pooling was considered an appropriate approach, but bivariate models failed to converge.

We translated summary estimates of sensitivity and specificity into summary estimates of the absolute numbers of true-positives, false-negatives, false-positives and true-negatives using a hypothetical population of 1000 women using an estimate of disease prevalence (pretest probability) reflecting the NICE threshold for cancer referral from generalist to specialist settings in the UK of 3% (NICE 2017).

Investigations of heterogeneity

We investigated the effect on estimates of test accuracy of menopausal status (premenopausal or postmenopausal) and of classification of histologically borderline ovarian tumours as disease positive (grouped with histologically malignant ovarian tumours) or where classification of borderline ovarian tumours was unclear or these tumour types were excluded. Grouping of histologically borderline ovarian tumours with histologically malignant ovarian tumours was considered clinically appropriate (reflecting current clinical practice) whereas exclusion of histologically borderline ovarian tumours was considered methodologically inappropriate.

We performed estimation of differences in accuracy using the NLMIXED procedure in Statistical Analysis System (SAS 2015) by including menopausal status or borderline grouping as covariates in the bivariate model. We reported differences in accuracy using the ratio of Diagnostic odds ratios with 95% CI and computed associated P values using Wald tests.

We were unable to conduct separate meta-analyses for the following planned investigations of heterogeneity because of a lack of data:

- healthcare setting: generalist setting (primary care, community care, family practice) versus specialist setting (secondary care, tertiary care (cancer unit, cancer centre));
- target condition: histological subtype: EOC versus non-EOC; high-grade serous epithelial (type II) versus other epithelial (type I); early-stage (stage I/II) versus late-stage disease (stage III/IV).

Sensitivity analyses

We did not undertake any sensitivity analyses.

Assessment of reporting bias

We did not undertake any formal assessment of reporting bias in our review due to current uncertainty about how to assess reporting bias in diagnostic test accuracy reviews, especially in the presence of heterogeneity (Deeks 2005).

RESULTS

Results of the search

The search identified 72,487 references. After removal of 20,388 duplicates, there remained 52,099 unique records. After reviewing titles and abstracts, we obtained and screened full-text copies of 1215 potentially relevant reports, of which 59 studies reporting 71 data sets were deemed eligible for inclusion. Reasons for full-text study exclusions are detailed in Figure 2 and studies are listed in Appendix 4. Forty-nine studies assessed the accuracy of a single test, whilst 10 studies included a within-person comparison of two or more index tests (Al Musalhi 2016; Anton 2012; Krascsenitis 2016; Liest 2019; Lycke 2018; Meys 2017; Niemi 2017; Richards 2015; Sayasneh 2013a; Testa 2014). Test types and thresholds were too varied to permit separate meta-analyses of direct comparison studies.

Index tests and thresholds

Of the 71 data sets (59 studies; 32,059 participants, 9545 cases of OC), 17 evaluated the accuracy of RMI at a threshold of 200 and two at a threshold of 250 (10,283 participants, 2654 cases of OC); 42 evaluated the accuracy of ROMA (13,715 unique participants, 3944 cases of OC) at threshold pairs for pre- and postmenopausal women of 7.4 (± 2) (N = 12) and 25.3 (± 2) (N = 15); 12.5 and 14.4 (N = 3), 13.1 (± 2) (N = 27) and 27.7 (± 2) (N = 13); 11.4 (N = 11) and 29.9 (N = 12); five studies evaluated the accuracy of LR2 (5000 participants, 1743 cases of OC to achieve a post-test probability of OC of 10%); and four studies evaluated the accuracy of ADNEX (3061 participants, 1204 cases of OC) to achieve a post-test probability of OC of 3%, 5%, 10% and 15% (Table 2).

Characteristics of included studies

In summary, 41 studies were conducted in Europe, 12 in the Asia-Pacific region, five in North America and one in South America. Nineteen studies were multicentre. These tests can be carried out in primary care, by dedicated gynaecologists in hospital settings (secondary care), by gynaecological oncologists in specialist units (tertiary care), or across a mixture of healthcare settings. Forty-nine studies were conducted in specialist settings (nine in mixed secondary and tertiary settings, 28 in tertiary care settings and 12 in secondary settings) and 10 studies did not report the healthcare setting.

Menopausal status and age alter the spectrum of disease (the prevalence of OC, range of histological subtypes and the range of differential diagnoses). In postmenopausal women, the prevalence of OC is higher and certain histological subtypes (EOC) are more common. In premenopausal women the prevalence of germ cell tumours is higher and the normal menstrual cycle and benign pathology such as endometriosis can result in false-positive test results. In the absence of information on menopausal status, 50 years can be used to stratify women for estimation of test accuracy to reflect this change in spectrum and risk. Across all studies reporting age (41/59 included studies), mean age varied between 37 and 65 years and age range varied between 11 and 94 years. One study restricted inclusion to premenopausal women and four studies restricted inclusion to postmenopausal women.

Testing prior to surgical investigation in this patient group in current clinical practice will have included one or more of clinical history and examination, biomarker measurement and USS. None of the studies detailed the clinical pathway of participants from presentation to the decision to test and the role of the index tests. Only three ROMA studies (Farzaneh 2014; Karlsen 2012; Ortiz-Munoz 2014), and one RMI study (Karlsen 2012) specified the presence of symptoms including 'gynaecological symptom's, pelvic pain and vaginal bleeding, pain, distension and weight loss', whilst 10 ROMA studies reported that an adnexal mass was identified following investigation with one of USS, MRI or CT.

Excluding certain tumour types changes the population spectrum as index test performance differed in different histological subtypes and at different stages of malignancy. For example, CA125 is known to have a higher sensitivity in EOC compared to other types of ovarian tumour such as stromal and germ cell tumours (Kobayashi 2012). The range of ovarian pathology reported in included studies varied. Eighteen ROMA and four RMI studies explicitly restricted inclusion to EOC, and seven ROMA studies and one RMI study explicitly excluded borderline tumours. A further 18 ROMA and three RMI studies did not report the occurrence of borderline tumours.

Characteristics of included studies are summarised in Table 3 (RMI), Table 4 (ROMA), Table 5 (LR2) and Table 6 (ADNEX).

Methodological quality of included studies

The methodological quality of all 59 included studies (71 data sets) evaluating one or more of RMI, ROMA, LR2 and ADNEX studies is summarised in Figure 3 and Figure 4. Separate figures summarise study quality by index test: RMI, ROMA, LR2 and ADNEX (Appendix 5).

Figure 3. Risk of bias and applicability concerns graph for 59 individual included studies for index tests. Review authors' judgements about each domain presented as percentages across included studies. ADNEX: Assessment of Different NEoplasias in the adneXa model; LR2: Logistic Regression 2 model; RMI I: Risk of Malignancy Index I; ROMA: Risk of Ovarian Malignancy Algorithm.

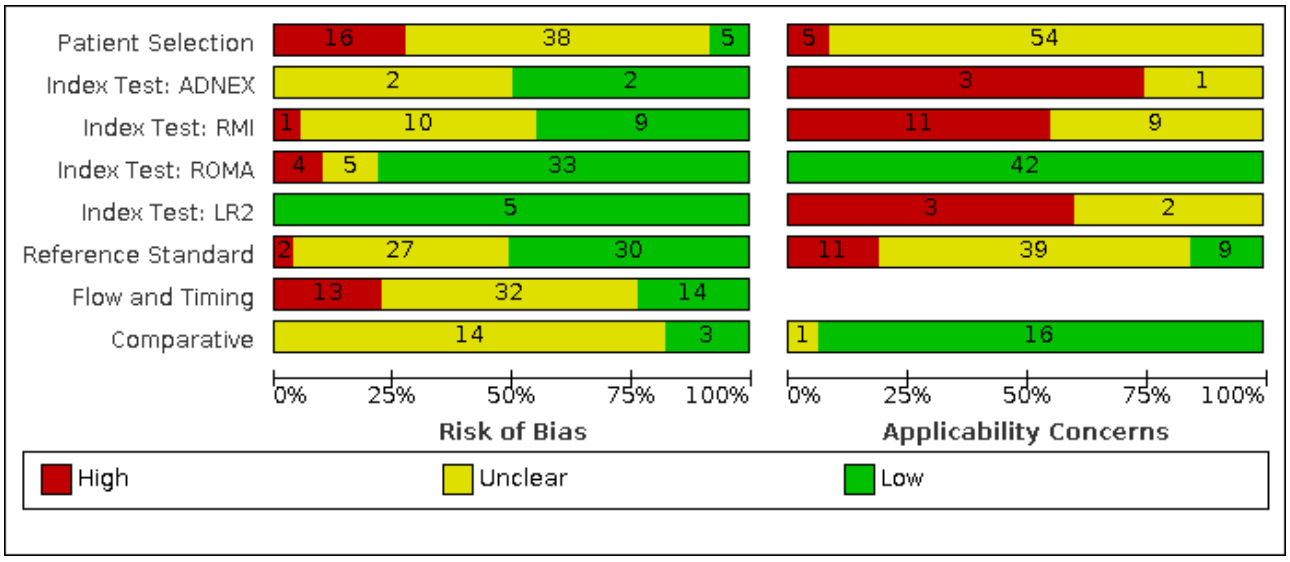


Figure 4. Risk of bias and applicability concerns figure for 59 individual included studies for index tests. Review authors' judgements about each domain for each included study. Empty cells indicate that an index test was not

evaluated by a study. ADNEX: Assessment of Different NEoplasias in the adneXa model; LR2: Logistic Regression 2 model; RMI I: Risk of Malignancy Index I; ROMA: Risk of Ovarian Malignancy Algorithm.

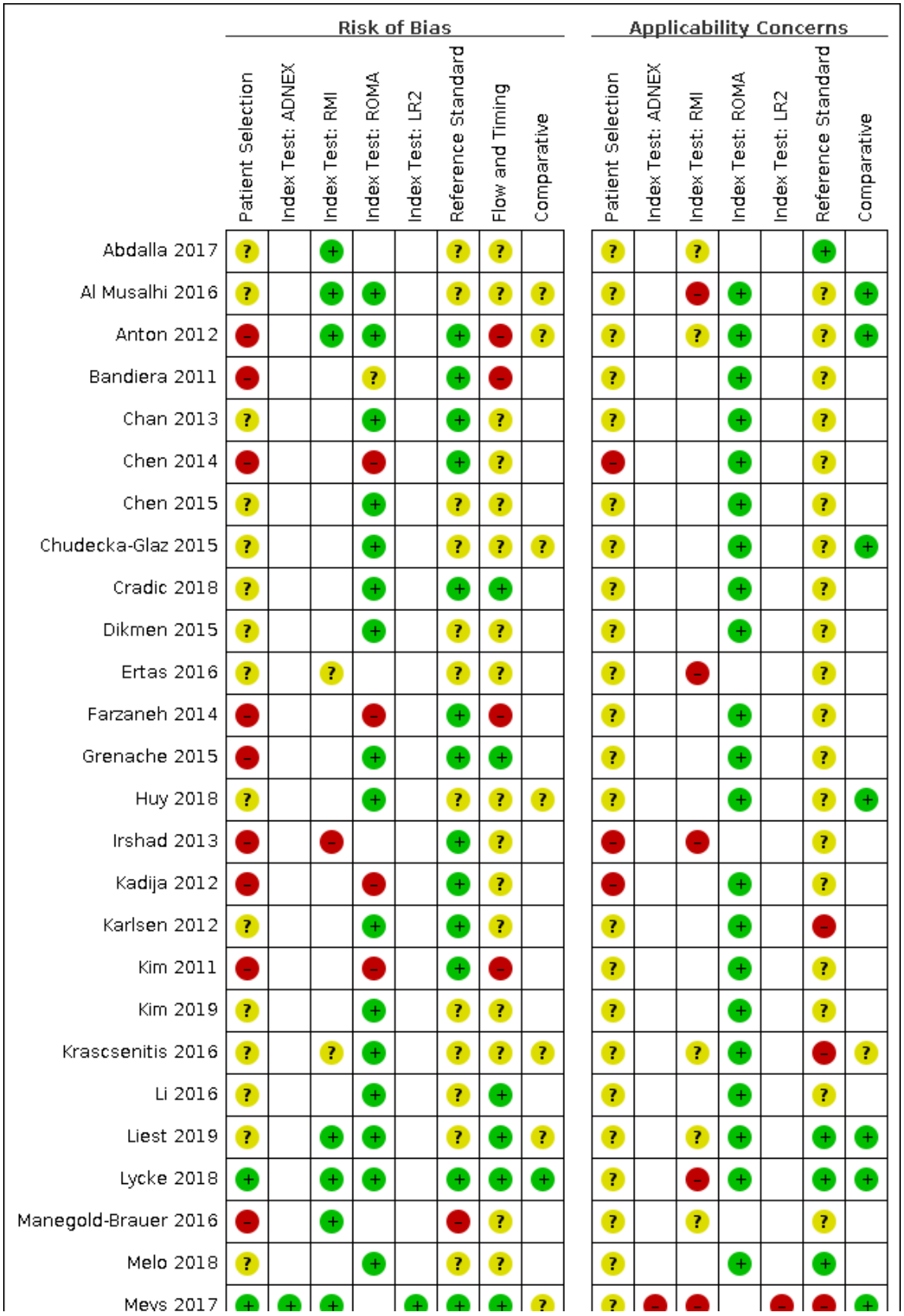


Figure 4. (Continued)

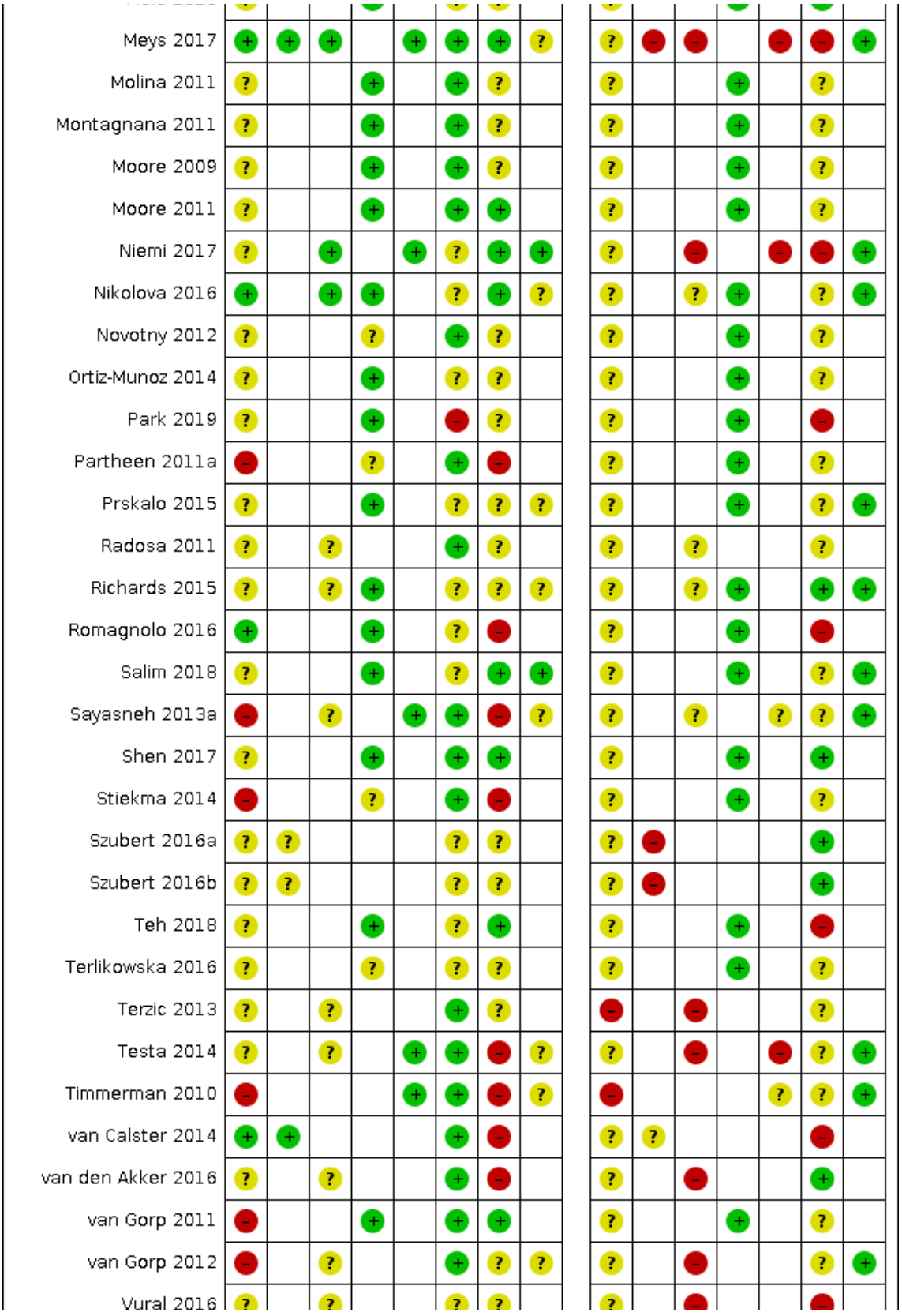
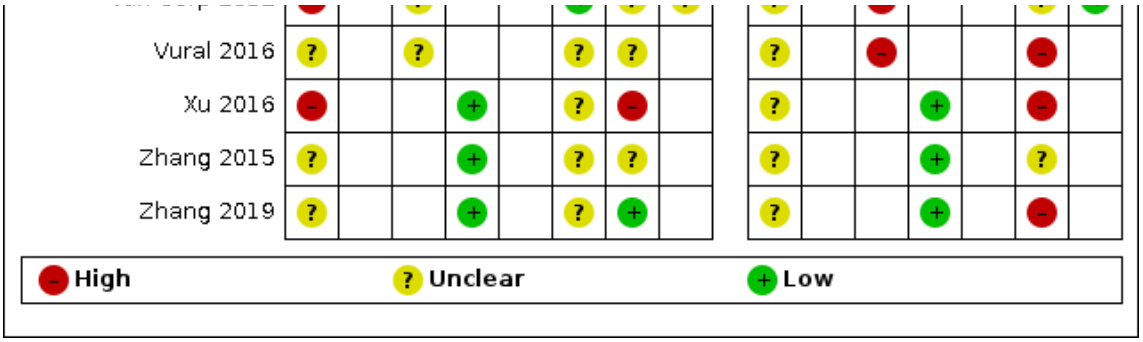


Figure 4. (Continued)



Participant selection domain

Across all included studies for the participant selection domain (Figure 3), 16/59 (27%) studies were at high risk of bias and 38/59 (64%) at unclear risk of bias. Only five studies were at low risk of bias on the basis that authors explicitly reported consecutive sampling and comprehensively listed tumour pathology identified at histology allowing a judgement to be made about selection of tumour types that might affect estimates of accuracy such as EOC and borderline tumours (Lycke 2018; Meys 2017; Nikolova 2016; Romagnolo 2016; van Calster 2014). Fifty-four of 59 (92%) studies were at high or unclear applicability concern for the participant selection domain because study participants did not obviously represent symptomatic women.

Index test domain

For the index test domain, 33/42 (79%) ROMA studies, 2/4 (50%) ADNEX studies and 9/20 (45%) RMI studies were at low risk of bias either because of the prospective nature of studies, or in the case of ROMA, the objective nature of the index test. One retrospective RMI study was at high risk of bias because RMI test results were interpreted with knowledge of the reference standard result (presence of absence of OC) (Irshad 2013). Four ROMA studies were at high risk of bias because they did not predefine the definition of the cut-off point for a positive test result (Chen 2014; Farzaneh 2014; Kadija 2012; Kim 2011). For the index test domain, applicability concern was high or unclear for all RMI, ADNEX and LR2 studies because USS was conducted by specialist sonographers or their level of specialisation was unclear.

Reference standard and target condition domain

For the reference standard domain, 30/59 (51%) studies were at low risk of bias. Twenty-seven of 59 (46%) studies were at unclear risk of bias, and two were at high risk of bias (Huy 2018; Park 2019), either because the minimum length of follow-up for index negatives was not reported at six months, or because there was concern that the reference standard outcome was ascertained with knowledge of the index test result. For the reference standard and target condition domain, applicability concern was as high or unclear in 50/59 (85%) studies because borderline tumours had been excluded from analysis or classification of borderline tumours for estimation of test accuracy was unclear.

Flow and timing domain

For the flow and timing domain, 32/59 (54%) studies were at unclear risk of bias most commonly because of no information about the interval between the index test and the reference standard. Thirteen of 59 (22%) studies were at high risk of bias because not all participants receiving an index test received a reference standard.

Findings

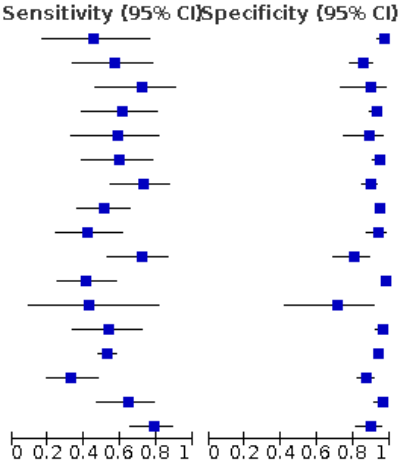
Comparison of accuracy in premenopausal and postmenopausal women

Table 2, Figure 5 (RMI), Figure 6 (ROMA), Figure 7 (LR2) and Figure 8 (ADNEX) present the accuracy of the 59 unique included studies and 71 data sets in pre- and postmenopausal women. There was a consistent difference in sensitivity (higher in postmenopausal women) and specificity (lower in postmenopausal women) across all versions of all index tests at all thresholds analysed. Subsequently, we estimated sensitivity and specificity in pre- and postmenopausal women separately.

Figure 5. Forest plot of tests: Risk of Malignancy Index I (RMI I) at thresholds of 200 and 250, separately for pre-menopausal and post-menopausal women. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

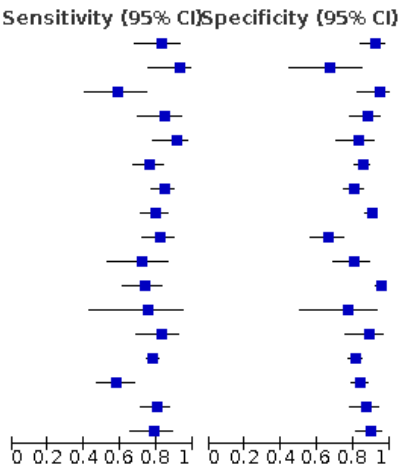
RMI I 200 premenopausal

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abdalla 2017	5	6	6	178	0.45 [0.17, 0.77]	0.97 [0.93, 0.99]
Al Musalhi 2016	12	21	9	120	0.57 [0.34, 0.78]	0.85 [0.78, 0.91]
Anton 2012	13	3	5	26	0.72 [0.47, 0.90]	0.90 [0.73, 0.98]
Ertas 2016	14	20	9	248	0.61 [0.39, 0.80]	0.93 [0.89, 0.95]
Krascsenitis 2016	10	5	7	38	0.59 [0.33, 0.82]	0.88 [0.75, 0.96]
Liest 2019	16	18	11	287	0.59 [0.39, 0.78]	0.94 [0.91, 0.96]
Lycke 2018	24	24	9	206	0.73 [0.54, 0.87]	0.90 [0.85, 0.93]
Manegold-Brauer 2016	25	35	24	546	0.51 [0.36, 0.66]	0.94 [0.92, 0.96]
Meys 2017	13	6	18	91	0.42 [0.25, 0.61]	0.94 [0.87, 0.98]
Niemi 2017	23	13	9	53	0.72 [0.53, 0.86]	0.80 [0.69, 0.89]
Radosa 2011	16	19	23	832	0.41 [0.26, 0.58]	0.98 [0.97, 0.99]
Richards 2015	3	4	4	10	0.43 [0.10, 0.82]	0.71 [0.42, 0.92]
Sayasneh 2013a	15	5	13	132	0.54 [0.34, 0.72]	0.96 [0.92, 0.99]
Testa 2014	200	59	178	917	0.53 [0.48, 0.58]	0.94 [0.92, 0.95]
van den Akker 2016	15	30	31	204	0.33 [0.20, 0.48]	0.87 [0.82, 0.91]
van Gorp 2012	25	6	14	133	0.64 [0.47, 0.79]	0.96 [0.91, 0.98]
Vural 2016	41	9	11	78	0.79 [0.65, 0.89]	0.90 [0.81, 0.95]



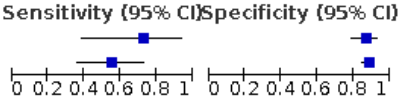
RMI I 200 postmenopausal

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abdalla 2017	34	6	7	70	0.83 [0.68, 0.93]	0.92 [0.84, 0.97]
Al Musalhi 2016	25	8	2	16	0.93 [0.76, 0.99]	0.67 [0.45, 0.84]
Anton 2012	21	2	15	35	0.58 [0.41, 0.74]	0.95 [0.82, 0.99]
Ertas 2016	37	9	7	64	0.84 [0.70, 0.93]	0.88 [0.78, 0.94]
Krascsenitis 2016	40	10	4	48	0.91 [0.78, 0.97]	0.83 [0.71, 0.91]
Liest 2019	89	46	28	260	0.76 [0.67, 0.83]	0.85 [0.80, 0.89]
Lycke 2018	112	42	21	173	0.84 [0.77, 0.90]	0.80 [0.75, 0.86]
Manegold-Brauer 2016	98	35	25	320	0.80 [0.71, 0.86]	0.90 [0.87, 0.93]
Meys 2017	69	39	15	75	0.82 [0.72, 0.90]	0.66 [0.56, 0.74]
Niemi 2017	23	13	9	53	0.72 [0.53, 0.86]	0.80 [0.69, 0.89]
Radosa 2011	49	21	18	383	0.73 [0.61, 0.83]	0.95 [0.92, 0.97]
Richards 2015	9	4	3	13	0.75 [0.43, 0.95]	0.76 [0.50, 0.93]
Sayasneh 2013a	38	5	8	39	0.83 [0.69, 0.92]	0.89 [0.75, 0.96]
Testa 2014	470	85	132	362	0.78 [0.75, 0.81]	0.81 [0.77, 0.85]
van den Akker 2016	54	48	39	249	0.58 [0.47, 0.68]	0.84 [0.79, 0.88]
van Gorp 2012	89	11	22	74	0.80 [0.72, 0.87]	0.87 [0.78, 0.93]
Vural 2016	41	9	11	78	0.79 [0.65, 0.89]	0.90 [0.81, 0.95]



RMI I 250 premenopausal

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Nikolova 2016	8	12	3	82	0.73 [0.39, 0.94]	0.87 [0.79, 0.93]
Terzic 2013	17	38	14	287	0.55 [0.36, 0.73]	0.88 [0.84, 0.92]



RMI I 250 postmenopausal

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Irshad 2013	21	3	2	10	0.91 [0.72, 0.99]	0.77 [0.46, 0.95]
Terzic 2013	59	22	15	88	0.80 [0.69, 0.88]	0.80 [0.71, 0.87]

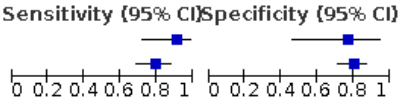
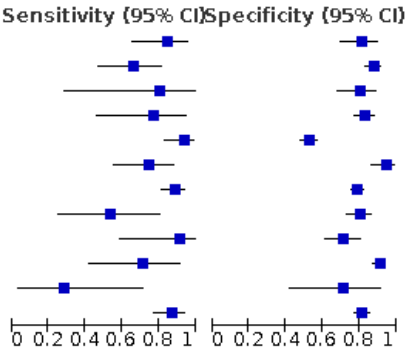


Figure 6. Forest plot of tests: Risk of Ovarian Malignancy Algorithm (ROMA) in at thresholds of 7.4 (± 2), 12.5, 13.1 (± 2), 7.4, 13.1 and 11.4 in premenopausal women, and at thresholds of 25.3 (± 2), 14.4, 27.7 (± 2), 25.3, 27.7 and 29.9 in postmenopausal women. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

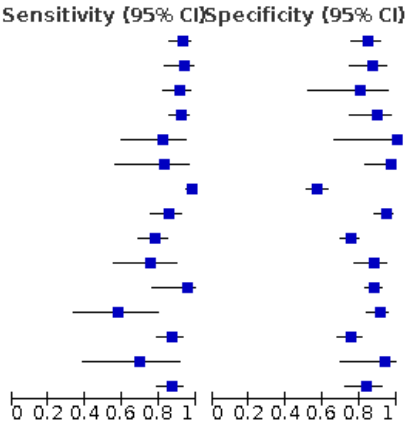
ROMA 7.4 (± 2) premenopausal

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bandiera 2011	22	13	4	56	0.85 [0.65, 0.96]	0.81 [0.70, 0.90]
Chan 2013	21	34	11	235	0.66 [0.47, 0.81]	0.87 [0.83, 0.91]
Grenache 2015	4	13	1	52	0.80 [0.28, 0.99]	0.80 [0.68, 0.89]
Huy 2018	10	37	3	180	0.77 [0.46, 0.95]	0.83 [0.77, 0.88]
Karlsen 2012	46	251	3	279	0.94 [0.83, 0.99]	0.53 [0.48, 0.57]
Kim 2011	23	4	8	67	0.74 [0.55, 0.88]	0.94 [0.86, 0.98]
Li 2016	96	136	12	501	0.89 [0.81, 0.94]	0.79 [0.75, 0.82]
Melo 2018	7	28	6	114	0.54 [0.25, 0.81]	0.80 [0.73, 0.86]
Nikolova 2016	10	27	1	67	0.91 [0.59, 1.00]	0.71 [0.61, 0.80]
Park 2019	10	29	4	282	0.71 [0.42, 0.92]	0.91 [0.87, 0.94]
Richards 2015	2	4	5	10	0.29 [0.04, 0.71]	0.71 [0.42, 0.92]
Shen 2017	60	79	9	347	0.87 [0.77, 0.94]	0.81 [0.77, 0.85]



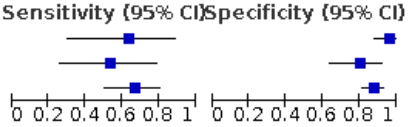
ROMA 25.3 (± 2) postmenopausal

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bandiera 2011	81	15	6	81	0.93 [0.86, 0.97]	0.84 [0.76, 0.91]
Chan 2013	46	7	3	46	0.94 [0.83, 0.99]	0.87 [0.75, 0.95]
Chen 2014	63	3	6	12	0.91 [0.82, 0.97]	0.80 [0.52, 0.96]
Chudecka-Glaz 2015	114	4	10	33	0.92 [0.86, 0.96]	0.89 [0.75, 0.97]
Farzaneh 2014	18	0	4	9	0.82 [0.60, 0.95]	1.00 [0.66, 1.00]
Huy 2018	14	1	3	29	0.82 [0.57, 0.96]	0.97 [0.83, 1.00]
Karlsen 2012	198	120	5	159	0.98 [0.94, 0.99]	0.57 [0.51, 0.63]
Li 2016	70	5	12	85	0.85 [0.76, 0.92]	0.94 [0.88, 0.98]
Liest 2019	91	76	26	230	0.78 [0.69, 0.85]	0.75 [0.70, 0.80]
Melo 2018	21	8	7	56	0.75 [0.55, 0.89]	0.88 [0.77, 0.94]
Novotny 2012	20	28	1	207	0.95 [0.76, 1.00]	0.88 [0.83, 0.92]
Park 2019	11	9	8	90	0.58 [0.33, 0.80]	0.91 [0.83, 0.96]
Partheen 2011a	81	41	12	124	0.87 [0.79, 0.93]	0.75 [0.68, 0.82]
Richards 2015	9	1	4	15	0.69 [0.39, 0.91]	0.94 [0.70, 1.00]
Shen 2017	91	9	14	47	0.87 [0.79, 0.93]	0.84 [0.72, 0.92]



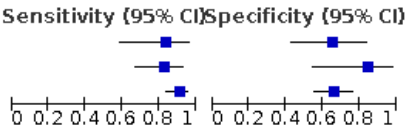
ROMA 12.5 premenopausal

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Kadija 2012	7	2	4	54	0.64 [0.31, 0.89]	0.96 [0.88, 1.00]
Montagnana 2011	8	7	7	29	0.53 [0.27, 0.79]	0.81 [0.64, 0.92]
van Gorp 2011	28	17	14	125	0.67 [0.50, 0.80]	0.88 [0.82, 0.93]



ROMA 14.4 postmenopausal

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Kadija 2012	15	8	3	15	0.83 [0.59, 0.96]	0.65 [0.43, 0.84]
Montagnana 2011	33	2	7	11	0.82 [0.67, 0.93]	0.85 [0.55, 0.98]
van Gorp 2011	108	29	11	57	0.91 [0.84, 0.95]	0.66 [0.55, 0.76]



ROMA 13.1 (± 2) premenopausal

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Al Musalhi 2016	11	14	10	127	0.52 [0.30, 0.74]	0.90 [0.84, 0.94]
Anton 2012	14	9	4	20	0.78 [0.52, 0.94]	0.69 [0.49, 0.85]
Chen 2014	48	16	6	38	0.89 [0.77, 0.96]	0.70 [0.56, 0.82]
Chen 2015	20	7	0	41	1.00 [0.83, 1.00]	0.85 [0.72, 0.94]
Chudecka-Glaz 2015	29	16	9	198	0.76 [0.60, 0.89]	0.93 [0.88, 0.96]
Cradic 2018	37	12	2	63	0.95 [0.83, 0.99]	0.84 [0.74, 0.91]
Dikmen 2015	12	10	2	73	0.86 [0.57, 0.98]	0.88 [0.79, 0.94]
Farzaneh 2014	16	7	5	40	0.76 [0.53, 0.92]	0.85 [0.72, 0.94]
Grenache 2015	3	7	2	58	0.60 [0.15, 0.95]	0.89 [0.79, 0.96]
Kadija 2012	7	2	4	54	0.64 [0.31, 0.89]	0.96 [0.88, 1.00]
Kim 2019	10	71	4	496	0.71 [0.42, 0.92]	0.87 [0.84, 0.90]
Krascsenitis 2016	12	11	5	32	0.71 [0.44, 0.90]	0.74 [0.59, 0.86]
Lycke 2018	26	44	7	186	0.79 [0.61, 0.91]	0.81 [0.75, 0.86]
Molina 2011	20	25	7	201	0.74 [0.54, 0.89]	0.89 [0.84, 0.93]

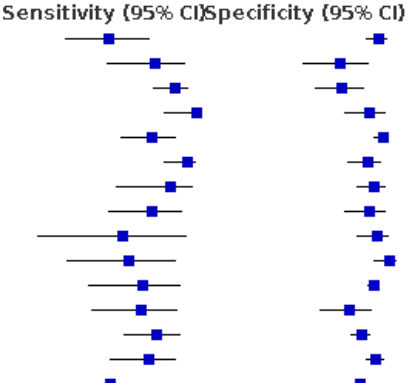
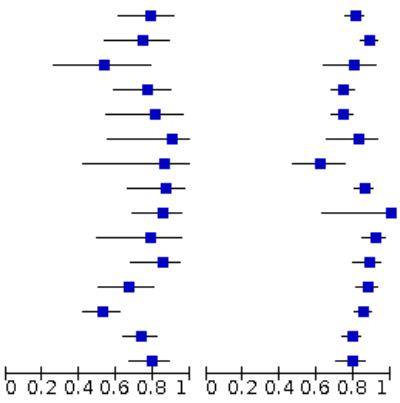


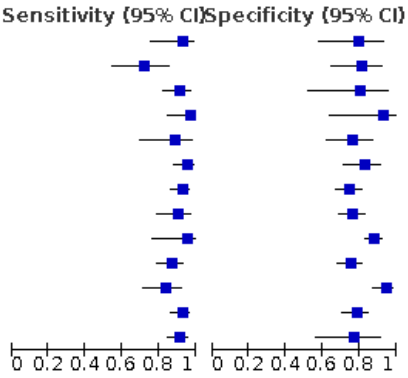
Figure 6. (Continued)

Lycke 2018	26	44	7	186	0.79 [0.61, 0.91]	0.81 [0.75, 0.86]
Molina 2011	20	25	7	201	0.74 [0.54, 0.89]	0.89 [0.84, 0.93]
Montagnana 2011	8	7	7	29	0.53 [0.27, 0.79]	0.81 [0.64, 0.92]
Moore 2009	26	51	8	151	0.76 [0.59, 0.89]	0.75 [0.68, 0.81]
Moore 2011	13	60	3	173	0.81 [0.54, 0.96]	0.74 [0.68, 0.80]
Ortiz-Munoz 2014	9	6	1	28	0.90 [0.55, 1.00]	0.82 [0.65, 0.93]
Prskalo 2015	6	19	1	31	0.86 [0.42, 1.00]	0.62 [0.47, 0.75]
Romagnolo 2016	20	30	3	186	0.87 [0.66, 0.97]	0.86 [0.81, 0.90]
Stiekma 2014	29	0	5	8	0.85 [0.69, 0.95]	1.00 [0.63, 1.00]
Teh 2018	11	7	3	81	0.79 [0.49, 0.95]	0.92 [0.84, 0.97]
Terlikowska 2016	28	10	5	77	0.85 [0.68, 0.95]	0.89 [0.80, 0.94]
van Gorp 2011	28	17	14	125	0.67 [0.50, 0.80]	0.88 [0.82, 0.93]
Xu 2016	56	38	51	226	0.52 [0.42, 0.62]	0.86 [0.81, 0.90]
Zhang 2015	70	59	25	226	0.74 [0.64, 0.82]	0.79 [0.74, 0.84]
Zhang 2019	50	24	13	91	0.79 [0.67, 0.89]	0.79 [0.71, 0.86]



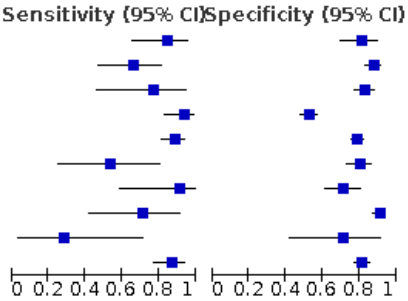
ROMA 27.7 (± 2) postmenopausal

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Al Musalhi 2016	25	5	2	19	0.93 [0.76, 0.99]	0.79 [0.58, 0.93]
Anton 2012	26	7	10	30	0.72 [0.55, 0.86]	0.81 [0.65, 0.92]
Chen 2014	63	3	6	12	0.91 [0.82, 0.97]	0.80 [0.52, 0.96]
Dikmen 2015	32	1	1	12	0.97 [0.84, 1.00]	0.92 [0.64, 1.00]
Grenache 2015	23	12	3	38	0.88 [0.70, 0.98]	0.76 [0.62, 0.87]
Molina 2011	80	10	4	49	0.95 [0.88, 0.99]	0.83 [0.71, 0.92]
Moore 2009	108	38	9	112	0.92 [0.86, 0.96]	0.75 [0.67, 0.81]
Moore 2011	46	36	5	114	0.90 [0.79, 0.97]	0.76 [0.68, 0.83]
Novotny 2012	20	28	1	207	0.95 [0.76, 1.00]	0.88 [0.83, 0.92]
Partheen 2011a	81	41	12	124	0.87 [0.79, 0.93]	0.75 [0.68, 0.82]
Romagnolo 2016	50	5	10	83	0.83 [0.71, 0.92]	0.94 [0.87, 0.98]
Salim 2018	113	30	9	108	0.93 [0.86, 0.97]	0.78 [0.70, 0.85]
Stiekma 2014	103	6	10	20	0.91 [0.84, 0.96]	0.77 [0.56, 0.91]



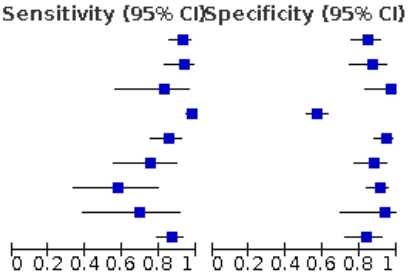
ROMA 7.4 premenopausal

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bandiera 2011	22	13	4	56	0.85 [0.65, 0.96]	0.81 [0.70, 0.90]
Chan 2013	21	34	11	235	0.66 [0.47, 0.81]	0.87 [0.83, 0.91]
Huy 2018	10	37	3	180	0.77 [0.46, 0.95]	0.83 [0.77, 0.88]
Karlsen 2012	46	251	3	279	0.94 [0.83, 0.99]	0.53 [0.48, 0.57]
Li 2016	96	136	12	501	0.89 [0.81, 0.94]	0.79 [0.75, 0.82]
Melo 2018	7	28	6	114	0.54 [0.25, 0.81]	0.80 [0.73, 0.86]
Nikolova 2016	10	27	1	67	0.91 [0.59, 1.00]	0.71 [0.61, 0.80]
Park 2019	10	29	4	282	0.71 [0.42, 0.92]	0.91 [0.87, 0.94]
Richards 2015	2	4	5	10	0.29 [0.04, 0.71]	0.71 [0.42, 0.92]
Shen 2017	60	79	9	347	0.87 [0.77, 0.94]	0.81 [0.77, 0.85]



ROMA 25.3 postmenopausal

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bandiera 2011	81	15	6	81	0.93 [0.86, 0.97]	0.84 [0.76, 0.91]
Chan 2013	46	7	3	46	0.94 [0.83, 0.99]	0.87 [0.75, 0.95]
Huy 2018	14	1	3	29	0.82 [0.57, 0.96]	0.97 [0.83, 1.00]
Karlsen 2012	198	120	5	159	0.98 [0.94, 0.99]	0.57 [0.51, 0.63]
Li 2016	70	5	12	85	0.85 [0.76, 0.92]	0.94 [0.88, 0.98]
Melo 2018	21	8	7	56	0.75 [0.55, 0.89]	0.88 [0.77, 0.94]
Park 2019	11	9	8	90	0.58 [0.33, 0.80]	0.91 [0.83, 0.96]
Richards 2015	9	1	4	15	0.69 [0.39, 0.91]	0.94 [0.70, 1.00]
Shen 2017	91	9	14	47	0.87 [0.79, 0.93]	0.84 [0.72, 0.92]



ROMA 13.1 premenopausal

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Al Musalhi 2016	11	14	10	127	0.52 [0.30, 0.74]	0.90 [0.84, 0.94]
Anton 2012	14	9	4	20	0.78 [0.52, 0.94]	0.69 [0.49, 0.85]
Dikmen 2015	12	10	2	73	0.86 [0.57, 0.98]	0.88 [0.79, 0.94]
Grenache 2015	3	7	2	58	0.60 [0.15, 0.95]	0.89 [0.79, 0.96]

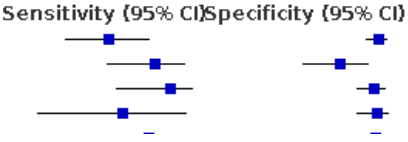
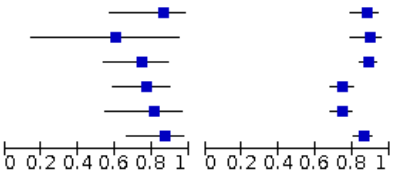


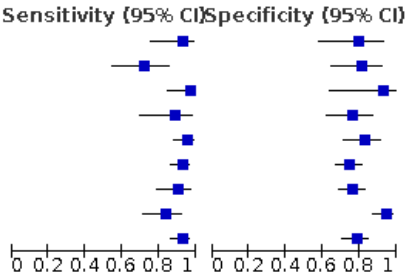
Figure 6. (Continued)

Dikmen 2015	12	10	2	73	0.86 [0.57, 0.98]	0.88 [0.79, 0.94]
Grenache 2015	3	7	2	58	0.60 [0.15, 0.95]	0.89 [0.79, 0.96]
Molina 2011	20	25	7	201	0.74 [0.54, 0.89]	0.89 [0.84, 0.93]
Moore 2009	26	51	8	151	0.76 [0.59, 0.89]	0.75 [0.68, 0.81]
Moore 2011	13	60	3	173	0.81 [0.54, 0.96]	0.74 [0.68, 0.80]
Romagnolo 2016	20	30	3	186	0.87 [0.66, 0.97]	0.86 [0.81, 0.90]



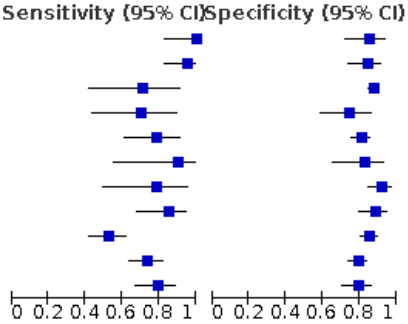
ROMA 27.7 postmenopausal

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Al Musalhi 2016	25	5	2	19	0.93 [0.76, 0.99]	0.79 [0.58, 0.93]
Anton 2012	26	7	10	30	0.72 [0.55, 0.86]	0.81 [0.65, 0.92]
Dikmen 2015	32	1	1	12	0.97 [0.84, 1.00]	0.92 [0.64, 1.00]
Grenache 2015	23	12	3	38	0.88 [0.70, 0.98]	0.76 [0.62, 0.87]
Molina 2011	80	10	4	49	0.95 [0.88, 0.99]	0.83 [0.71, 0.92]
Moore 2009	108	38	9	112	0.92 [0.86, 0.96]	0.75 [0.67, 0.81]
Moore 2011	46	36	5	114	0.90 [0.79, 0.97]	0.76 [0.68, 0.83]
Romagnolo 2016	50	5	10	83	0.83 [0.71, 0.92]	0.94 [0.87, 0.98]
Salim 2018	113	30	9	108	0.93 [0.86, 0.97]	0.78 [0.70, 0.85]



ROMA 11.4 premenopausal

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Chen 2015	20	7	0	41	1.00 [0.83, 1.00]	0.85 [0.72, 0.94]
Cradic 2018	37	12	2	63	0.95 [0.83, 0.99]	0.84 [0.74, 0.91]
Kim 2019	10	71	4	496	0.71 [0.42, 0.92]	0.87 [0.84, 0.90]
Krascsenitis 2016	12	11	5	32	0.71 [0.44, 0.90]	0.74 [0.59, 0.86]
Lycke 2018	26	44	7	186	0.79 [0.61, 0.91]	0.81 [0.75, 0.86]
Ortiz-Munoz 2014	9	6	1	28	0.90 [0.55, 1.00]	0.82 [0.65, 0.93]
Teh 2018	11	7	3	81	0.79 [0.49, 0.95]	0.92 [0.84, 0.97]
Terlikowska 2016	28	10	5	77	0.85 [0.68, 0.95]	0.89 [0.80, 0.94]
Xu 2016	56	38	51	226	0.52 [0.42, 0.62]	0.86 [0.81, 0.90]
Zhang 2015	70	59	25	226	0.74 [0.64, 0.82]	0.79 [0.74, 0.84]
Zhang 2019	50	24	13	91	0.79 [0.67, 0.89]	0.79 [0.71, 0.86]



ROMA 29.9 postmenopausal

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Chen 2015	38	7	2	15	0.95 [0.83, 0.99]	0.68 [0.45, 0.86]
Cradic 2018	34	8	3	48	0.92 [0.78, 0.98]	0.86 [0.74, 0.94]
Kim 2019	39	17	17	178	0.70 [0.56, 0.81]	0.91 [0.86, 0.95]
Krascsenitis 2016	42	17	2	41	0.95 [0.85, 0.99]	0.71 [0.57, 0.82]
Lycke 2018	113	49	0	186	1.00 [0.97, 1.00]	0.79 [0.73, 0.84]
Ortiz-Munoz 2014	27	7	2	112	0.93 [0.77, 0.99]	0.94 [0.88, 0.98]
Prskalo 2015	61	7	5	29	0.92 [0.83, 0.97]	0.81 [0.64, 0.92]
Teh 2018	13	4	0	10	1.00 [0.75, 1.00]	0.71 [0.42, 0.92]
Terlikowska 2016	55	3	8	38	0.87 [0.77, 0.94]	0.93 [0.80, 0.98]
Xu 2016	57	1	46	46	0.55 [0.45, 0.65]	0.98 [0.89, 1.00]
Zhang 2015	154	14	15	49	0.91 [0.86, 0.95]	0.78 [0.66, 0.87]
Zhang 2019	103	5	15	55	0.87 [0.80, 0.93]	0.92 [0.82, 0.97]

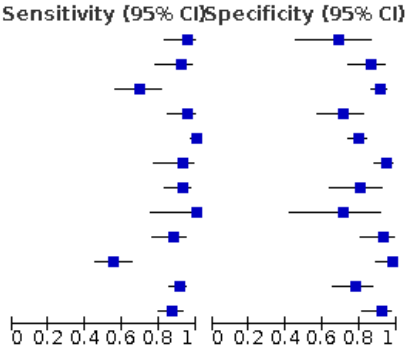


Figure 7. Forest plot of tests: Logistic Regression 2 model (LR2) separately for premenopausal and postmenopausal women. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

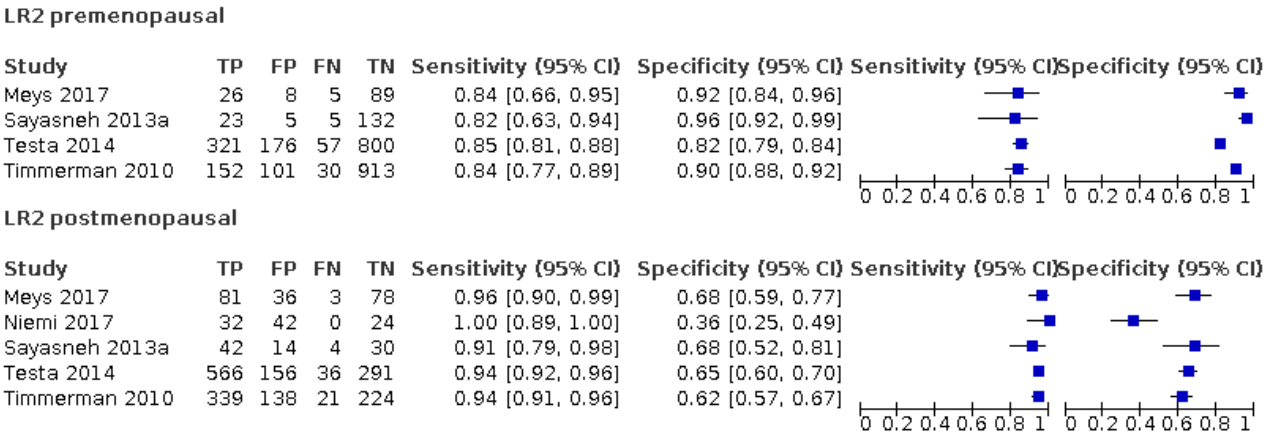
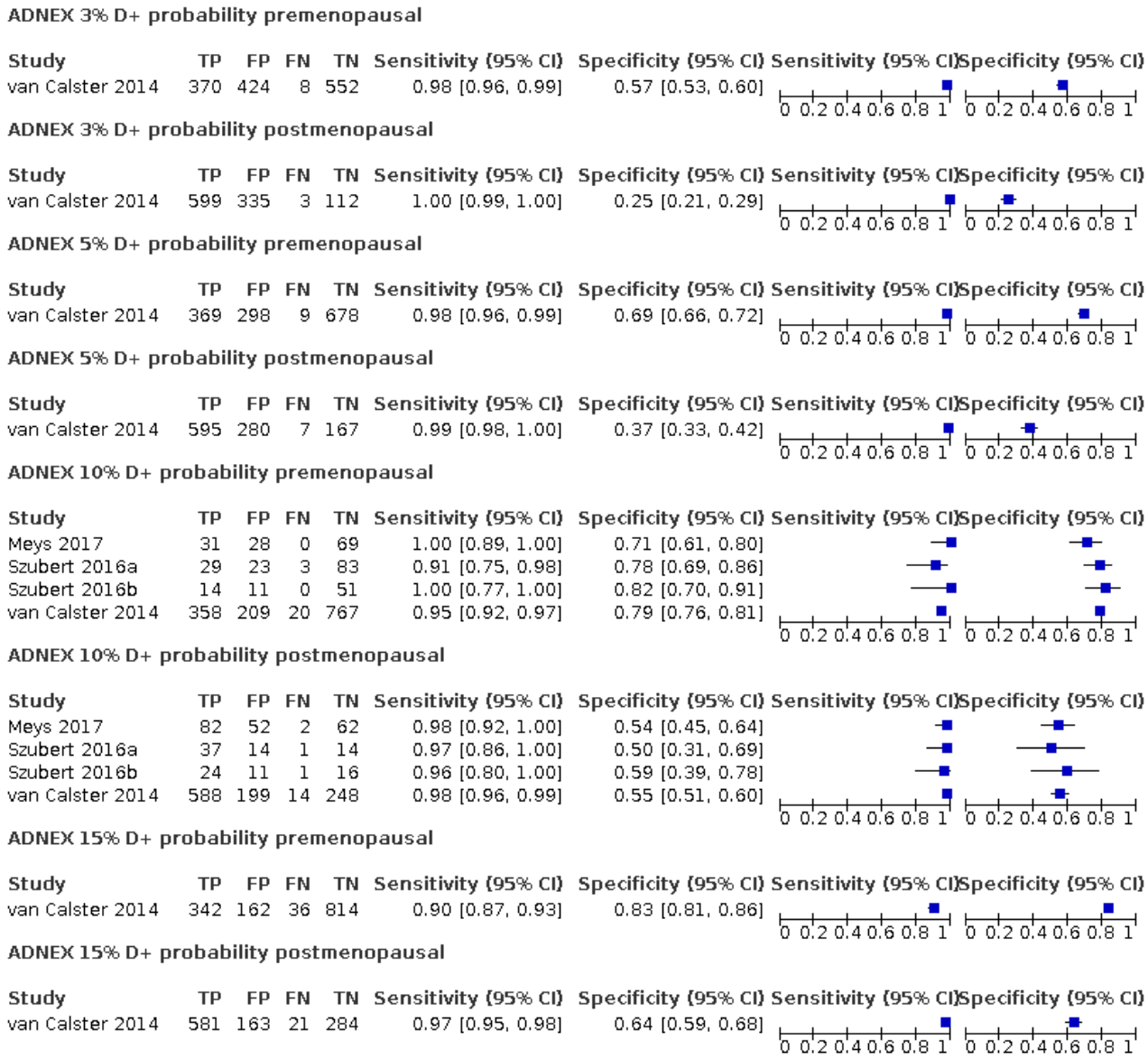


Figure 8. Forest plot of tests: Assessment of Different NEoplasias in the adneXa model (ADNEX) at thresholds of 3%, 5%, 10% and 15% disease probability separately for premenopausal and postmenopausal women. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.



Test positivity threshold

ROMA and ADNEX included studies reporting accuracy across a range of test positivity thresholds. The expected trade-off between sensitivity and specificity with changes in threshold was observed; as test positivity threshold increased, sensitivity increased and specificity decreased. For ROMA, there was no evidence of a difference in accuracy at thresholds reported by included studies.

It is of note that this pattern of test performance suggests a population selected on the basis of prior testing (i.e. representative of specialist settings). At earlier points in the testing pathway for OC, it would be expected that specificity would be lower in premenopausal women compared to postmenopausal women as a result of false-positives caused by benign conditions common in premenopausal women (ovarian cysts, endometriosis) and the normal menstrual cycle.

Accuracy of RMI, ROMA, LR2 and ADNEX in premenopausal women

RMI at a threshold of 200

Based on 17 studies, including 5233 premenopausal women, of whom 851 had a diagnosis of OC, the sensitivity of RMI at a threshold of 200 was 57.1% (95% CI 50.6% to 63.4%) and the specificity was 92.5% (95% CI 90.0% to 94.4%).

RMI at a threshold of 250

Based on two studies, including 461 premenopausal women, of whom 42 had a diagnosis of OC, the sensitivity of RMI at a threshold of 250 was 59.5% (95% CI 44.3% to 73.1%) and the specificity was 88.1% (84.6% to 90.8%)

LR2 to achieve a post-test probability of ovarian cancer of 10%

Based on four studies, including 2843 premenopausal women, of whom 619 had a diagnosis of OC, the sensitivity of LR2 was 83.2% (95% CI 78.6% to 87.0%) and the specificity was 90.4% (95% CI 84.6% to 94.1%).

ROMA

For ROMA, there was no evidence of a difference in accuracy at thresholds reported by included studies. Based on the threshold pair reported by the most studies: based on 27 studies, 4463 premenopausal women, of whom 825 had a diagnosis of OC, the sensitivity of ROMA at a threshold of 13.1 ± 2 was 77.8% (95% CI 72.5% to 82.4%) and the specificity was 84.3% (95% CI 81.3% to 86.8%).

ADNEX to achieve a post-test probability of ovarian cancer of 10%

For ADNEX, accuracy was reported at a threshold to achieve a post-test probability of OC of 3% (one study), 5% (one study), 10% (four studies) and 15% (one study). Based on four studies, including 1696 premenopausal women, of whom 455 had a diagnosis of OC, the sensitivity of ADNEX to achieve a post-test probability of OC of 10% was 94.9% (95% CI 92.5% to 96.6%) and the specificity was 78.2% (95% CI 75.8% to 80.4%).

Accuracy of RMI, ROMA, LR2 and ADNEX in postmenopausal women

RMI at a threshold of 200

Based on 17 studies, including 4369 postmenopausal women, of whom 1664 had a diagnosis of OC, the sensitivity of RMI at a threshold of 200 was 78.7% (95% CI 74.3% to 82.5%) and the specificity was 85.5% (95% CI 81.3% to 88.9%).

RMI at a threshold of 250

Based on two studies, including 220 postmenopausal women, of whom 97 had a diagnosis of OC, the sensitivity of RMI at a threshold of 250 was 82.5% (95% CI 73.6% to 88.8%) and the specificity was 79.7% (95% CI 71.6% to 85.9%).

LR2 to achieve a post-test probability of ovarian cancer of 10%

Based on five studies, including 2157 postmenopausal women, of whom 1124 had a diagnosis of OC, the sensitivity of LR2 was 94.5% (95% CI 92.8% to 95.7%) and the specificity was 60.5% (95% CI 49.3% to 70.7%).

ROMA

For ROMA, there was no evidence of a difference in accuracy at thresholds reported by the included studies. Based on the threshold pair reported by the most studies: based on 13 studies, including 2002 postmenopausal women, of whom 852 had a diagnosis of OC, the sensitivity of ROMA at a threshold of 27.7 ± 2 was 90.4% (95% CI 87.4% to 92.7%) and the specificity was 81.3% (95% CI 76.9% to 85.0%).

ADNEX to achieve a post-test probability of ovarian cancer of 10%

For ADNEX, accuracy was reported at a threshold to achieve a post-test probability of OC of 3% (one study), 5% (one study), 10% (four studies) and 15% (one study). Based on four studies, including 1365 postmenopausal women, of whom 749 had a diagnosis of OC, the sensitivity of ADNEX to achieve a post-test probability of OC of 10% was 97.6% (95% CI 96.2% to 98.5%) and the specificity was 55.2% (95% CI 51.2% to 59.1%).

HSROC (between study) comparison of RMI, ROMA, LR2 and ADNEX

To maximise data for comparison, studies were included regardless of the test positivity threshold used and we undertook an indirect comparison of index (Table 7) tests by fitting HSROC curves for premenopausal women (Figure 9) and postmenopausal women (Figure 10) separately. RMI was chosen as the baseline comparator as this is the test combination currently in routine clinical use in the UK. In premenopausal women, ADNEX and LR2 but not ROMA demonstrated superior accuracy compared to RMI (relative Diagnostic Odds Ratio (rDOR): ADNEX: 4.70, 95% CI 1.45 to 15.20; P = 0.014; LR2: 2.19, 95% CI 1.18 to 4.06; P = 0.0108; ROMA: 1.19, 95% CI 0.69 to 2.07; P = 0.5202). In postmenopausal women only ROMA demonstrated superior overall accuracy compared to RMI (rDOR 1.75, 95% CI 1.23 to 2.5; P = 0.0024) (Table 7).

Figure 9. Summary ROC plot of tests (pre-menopausal women): RMI I, ROMA, LR2 and ADNEX 10% D+ probability. ADNEX: Assessment of Different NEoplasias in the adneXa model; LR2: Logistic Regression 2 model; RMI I: Risk of Malignancy Index I; ROC: receiver operating characteristic; ROMA: Risk of Ovarian Malignancy Algorithm.

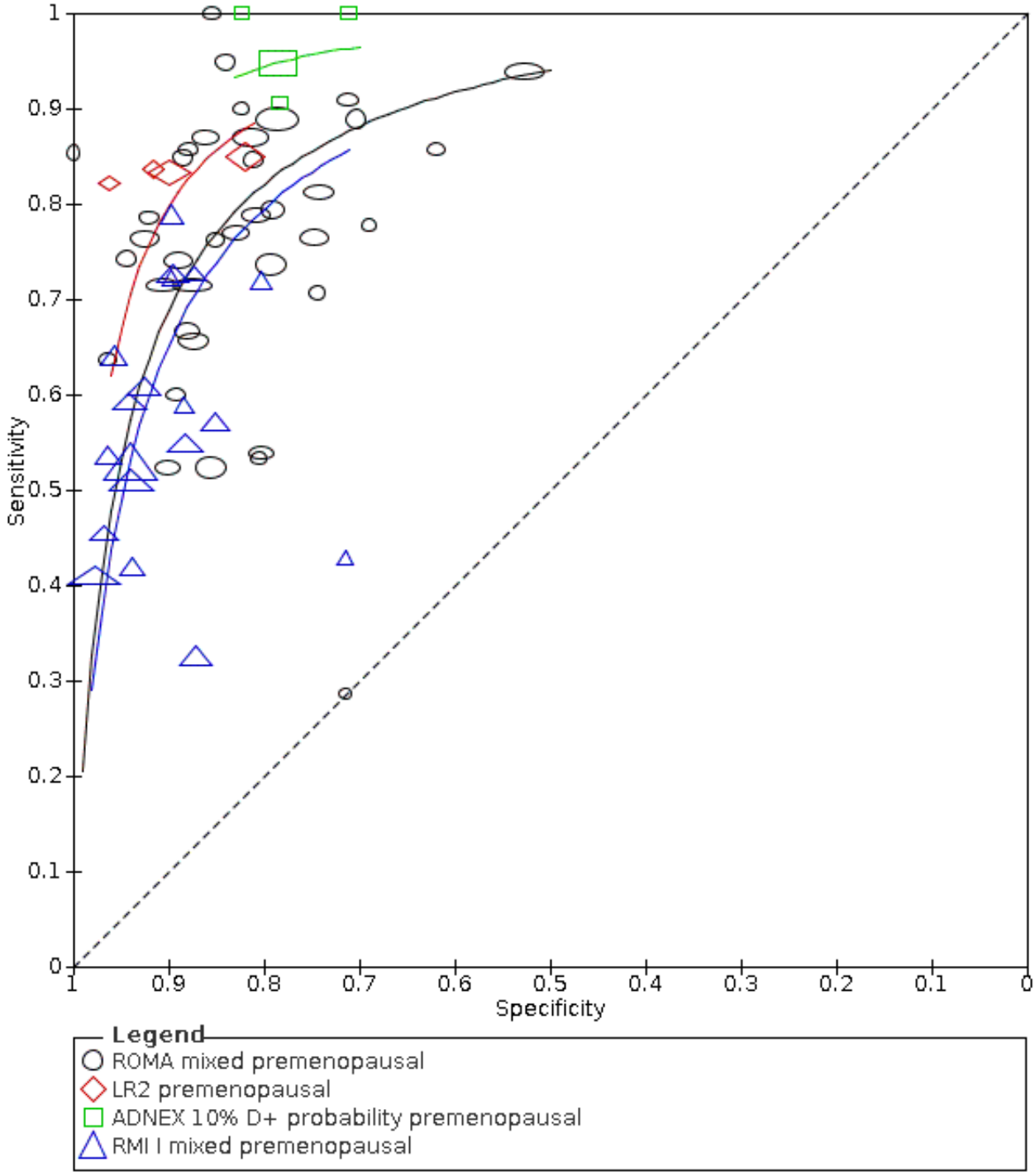
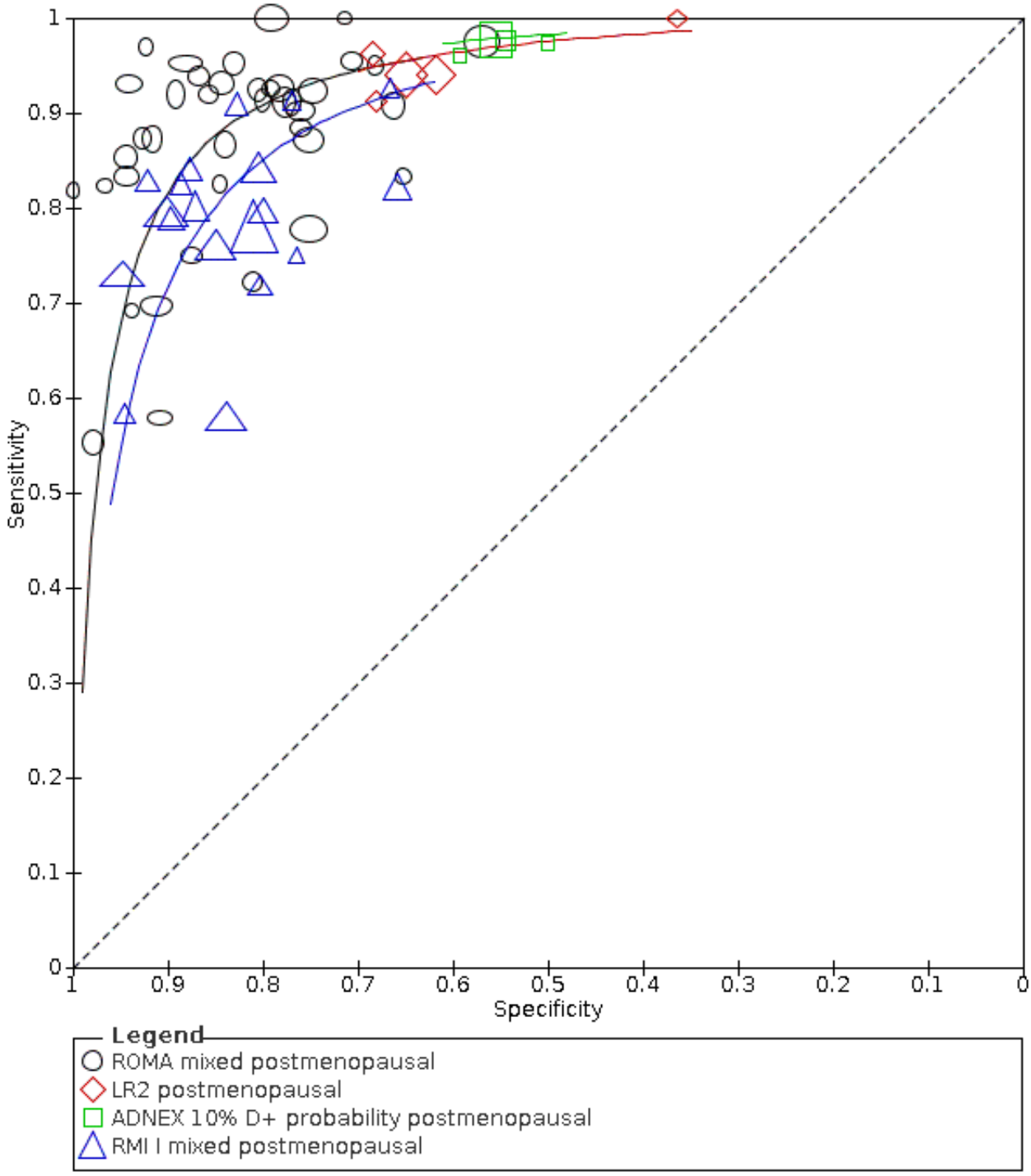


Figure 10. Summary ROC plot of tests (post-menopausal women): RMI I, ROMA, LR2 and ADNEX 10% D+ probability. ADNEX: Assessment of Different NEoplasias in the adneXa model; LR2: Logistic Regression 2 model; RMI I: Risk of Malignancy Index I; ROC: receiver operating characteristic; ROMA: Risk of Ovarian Malignancy Algorithm.



Differences in sensitivity between tests was estimated at fixed specificities of 80% and 90% (Table 7). These specificity thresholds were chosen in keeping with clinical consensus about an acceptable false-positive rate which is reflected in previous research and RCOG guidelines (RCOG 2016). It should be noted that the estimate of sensitivity for ADNEX in pre- and postmenopausal women at a fixed specificity of 90% is extrapolating beyond the data contributed by included ADNEX studies.

In premenopausal women at a fixed specificity of 80%, RMI has an estimated average sensitivity of 79.4% (95% CI 69.5% to 86.7%). The average difference in sensitivity of ROMA compared to RMI at a fixed specificity of 80% is compatible with chance (2.6% percentage points, 95% CI −5.5 to 10.7), but there was an increase in average sensitivity with LR2 and ADNEX (LR2: 9.6 percentage points higher, 95% CI 2.2 to 17.0; ADNEX: 14.9 percentage points higher, 95% CI 5.4 to 24.5).

In postmenopausal women at a fixed specificity of 80%, RMI has an average sensitivity of 85.1% (95% CI 80.9% to 88.5%). ROMA, LR2 and ADNEX demonstrated an increase in average sensitivity compared to RMI (ROMA: 5.8 percentage points, 95% CI 21.1 to 9.6; LR2: 5.7 percentage points, 95% CI 0.7 to 10.7; ADNEX: 8.3 percentage points, 95% CI 1.5 to 15.1).

Bivariate (between study) comparison of RMI, ROMA, LR2 and ADNEX

For decision-making purposes, the consequences of false-negatives (driven by sensitivity) and false-positives (driven by specificity) will not necessarily be considered equivalent and expressing accuracy in terms of overall discrimination misses this important distinction. In making recommendations for practice it is therefore useful to present test performance illustrating the trade-off between sensitivity and specificity at specific operating thresholds. Table 8 illustrates a comparison of tests at fixed thresholds in premenopausal women and Table 9 presents a comparison of tests at fixed thresholds in postmenopausal women: ROMA at a threshold of 13.1 (± 2) in premenopausal women (27/42 ROMA studies) and at a threshold of 27.7(± 2) in postmenopausal women (13/42 ROMA studies); LR2 at a post-test probability of 10% (4/4 studies in premenopausal women and 5/5 studies in postmenopausal women) and ADNEX at a post-test probability of 10% (4/4 studies in pre- and postmenopausal women) compared to RMI at a threshold of 200 (17/19 studies in pre- and postmenopausal women). For ROMA and ADNEX, the threshold pair reported by the most studies was chosen for this analysis.

Premenopausal women

In premenopausal women, RMI at a threshold of 200 (17 studies, 5233 participants, 851 cases of OC) had a sensitivity of 57.2% (95% CI 50.3 to 63.8) and a specificity of 92.5 (95% CI 90.3 to 94.2). Compared to RMI: ROMA at a threshold of 13.1 (± 2) (27 studies, 4463 participants, 825 cases of OC), demonstrated an increase in sensitivity of 20.2 percentage points (95% CI 12.2 to 28.3) but a decrease in specificity of –8.2 percentage points (95% CI –11.7 to –4.7), LR2 at a threshold to achieve a post-test probability of OC of 10% (4 studies, 2843 participants, 619 cases of OC), demonstrated an increase in sensitivity of 26.2 percentage points (95% CI 16.2 to 36.2) but with comparable specificity –2.1 percentage points (95% CI –7.2 to +2.9), ADNEX at a threshold to achieve a post-test probability of OC of 10% (4 studies, 1696 participants, 455 cases of OC), demonstrated an increase in sensitivity of 38.3 percentage points (95% CI 30.9 to 45.8) but a decrease in specificity of –14.8 percentage points (95% CI –24.0 to –5.5). In summary, in premenopausal women, ROMA, ADNEX and LR2 all demonstrated a higher sensitivity compared to RMI at a threshold of 200. In addition ADNEX appeared to demonstrate a marginally higher sensitivity compared to ROMA. LR2 had comparable specificity to RMI at a threshold of 200 whilst for ROMA and ADNEX specificity was lower.

Postmenopausal women

In postmenopausal women, RMI at a threshold of 200 (17 studies, 4369 participants, 1664 cases of OC) had a sensitivity of 78.4% (95% CI 74.6 to 81.7) and a specificity of 85.4% (95% CI 82.0 to 88.2). Compared to RMI: ROMA at a threshold of 27.7 (± 2) (13 studies, 2002 participants, 852 cases of OC), demonstrated an increase in sensitivity of 11.9 percentage points (95% CI 7.6 to 16.3) but a comparable specificity of –3.9 percentage points (95% CI –9.4 to 1.5), LR2 at a threshold to achieve a post-test probability of OC of 10% (5 studies, 2157 participants, 1124 cases of OC), demonstrated an increase in sensitivity of 16.4 percentage points (95% CI 12.3 to 20.5) but a decrease in specificity of –24.8 percentage points (95% CI –35.1 to –14.5), ADNEX at a threshold to achieve a post-test probability of OC of 10% (4 studies, 1365 participants, 749 cases of OC), demonstrated an increase in sensitivity of 19.2 percentage points (95% CI 15.4 to 23.1) but a decrease in specificity of –30.4 percentage points (95% CI –42.9 to –17.9). In summary, in postmenopausal women, ROMA, ADNEX and LR2 all demonstrated a higher sensitivity compared to RMI I at a threshold of 200. ROMA demonstrated a comparable specificity to RMI whilst for LR2 and ADNEX specificity was lower compared to RMI.

Investigation of the effect of classification of borderline tumours on estimates of test accuracy

In current clinical practice borderline ovarian tumours undergo similar surgical management to invasive malignant tumours. Included studies did not consistently include borderline ovarian tumours with malignant tumours for the purposes of estimating test accuracy. Exclusion of borderline tumours when estimating test accuracy in primary studies would be expected to result in overestimation of sensitivity, as they are a source of false-negative test results. In premenopausal women (38 ROMA studies; 19 RMI studies) and postmenopausal women (40 ROMA studies), there were sufficient data, when utilising all test positivity thresholds at a fixed specificity of 80%, to allow comparison of sensitivity estimated by studies where borderline tumours were classified as positive (grouped with malignant tumours) with studies excluding borderline tumours from analysis or where the classification of borderline tumours for analysis was unclear.

In postmenopausal women, for ROMA, there was a decrease in sensitivity of 6.4 percentage points (95% CI 1.2 to 11.5) for studies grouping borderline tumours with malignant compared to studies that excluded borderline tumours or where categorisation of borderline tumours for analysis was unclear (Table 10).

DISCUSSION

Summary of main results

To our knowledge, our systematic review is the first to compare the accuracy of ROMA, RMI and ADNEX in separately in premenopausal and postmenopausal women. Previous reviews have mostly evaluated combination tests (ROMA, RMI or LR2) in isolation and none have evaluated ADNEX. The most recent systematic review undertaking meta-analysis using hierarchical models was based on searches up to 2015 (Meys 2016). Estimates of sensitivity and specificity in premenopausal women (sensitivity 63%, specificity 93%) and postmenopausal women (sensitivity 79%, specificity 86%) were higher, but of a similar magnitude to those in this review.

Accuracy in premenopausal compared to postmenopausal women

We observed a consistent difference in sensitivity (higher in postmenopausal women) and specificity (lower in postmenopausal women) across all versions of all index tests at all thresholds analysed greater than could be expected by chance. This finding has important implications for research and practice: the utility of tests for diagnosing OC should be considered separately in premenopausal and postmenopausal women.

Comparison of the accuracy of RMI, ROMA, LR2 and ADNEX

In the UK, women with a suspected adnexal mass and with either an abnormal CA125 or USS are referred for investigation to secondary care where RMI is performed. Therefore, we investigated the performance of ROMA, LR2 and ADNEX relative to RMI. In pre- and postmenopausal women, RMI has lower sensitivity compared to ROMA, LR2 and ADNEX.

Premenopausal women

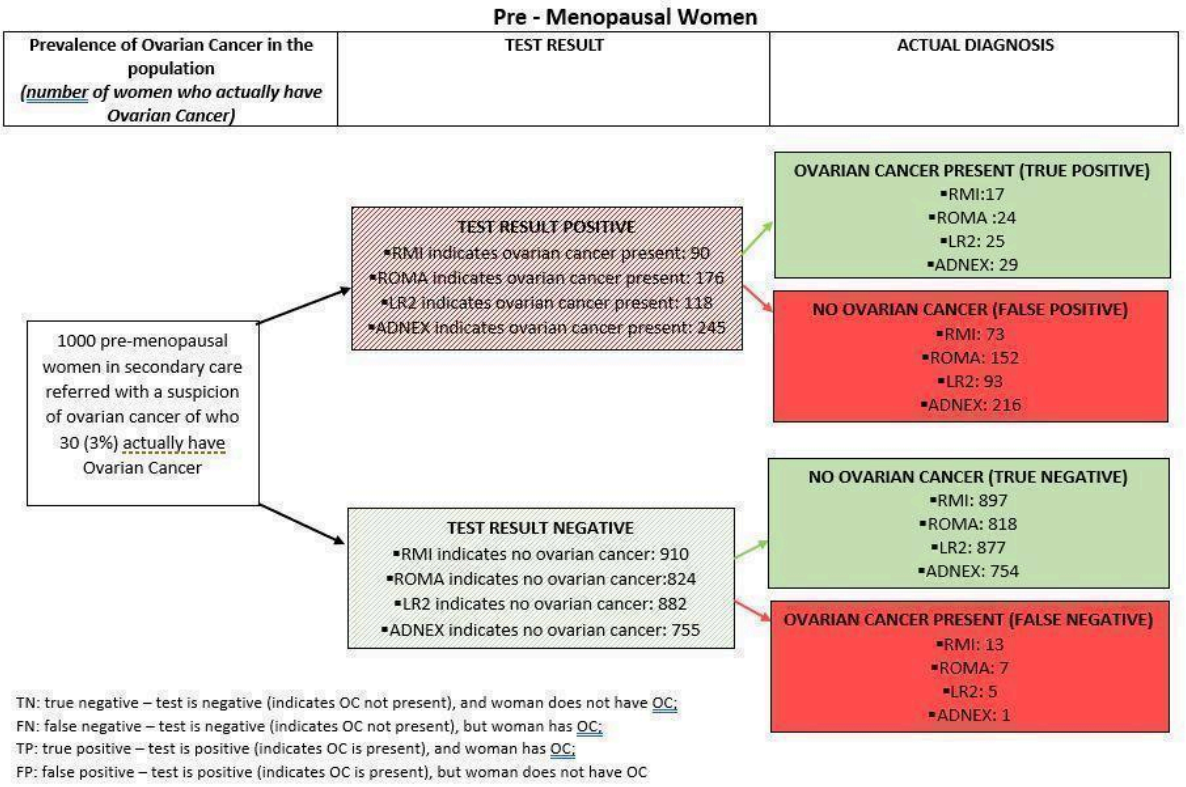
In premenopausal women, ROMA at a threshold of 13.1 (± 2), LR2 at a threshold to achieve a post-test probability of OC of 10% (post-test probability 10%) and ADNEX (post-test probability 10%) demonstrated a higher sensitivity compared to RMI (ROMA: 77.4%, 95% CI 72.7% to 81.5%; LR2: 83.3%, 95% CI 74.7% to 89.5%; ADNEX: 95.5%, 95% CI 91.0% to 97.8%; RMI: 57.2%, 95% CI 50.3% to 63.8%). The specificity of ROMA and ADNEX were lower in premenopausal women compared to RMI (ROMA: 84.3%, 95% CI 81.2% to 87.0%; ADNEX: 77.8%, 95% CI 67.4% to 85.5%; RMI: 92.5%, 95% CI 90.3% to 94.2%); the specificity of LR2 was comparable to RMI (90.4%, 95% CI 84.6% to 94.1%).

Based on our analysis, in a clinical setting with a pretest probability of OC of 3% (NICE 2017) in premenopausal women, for every 1000 premenopausal women tested:

- consequences of a positive test result:
  - an estimated 90 will have an RMI result indicating OC is present and of these 73 (81%) will not have OC;
  - an estimated 176 will have a ROMA result indicating OC is present and of these 152 (86%) will not have OC;
  - an estimated 118 will have an LR2 result indicating OC is present and of these 93 (79%) will not have OC;
  - an estimated 245 will have an ADNEX result indicating OC is present and of these 216 (88%) will not have OC;
- consequences of a negative test result:
  - of the 910 people with an RMI result indicating that OC is not present, 13 (1%) will actually have OC;
  - of the 824 people with a ROMA result indicating that OC is not present, 7 (0.8%) will actually have OC;
  - of the 882 people with an LR2 result indicating that OC is not present, 5 (0.6%) will actually have OC;
  - of the 755 people with an ADNEX result indicating that OC is not present, 1 (0.1%) will actually have OC.

See Figure 11.

Figure 11. Illustration of the consequences of testing a hypothetical cohort of premenopausal women referred from primary care (estimated prevalence of ovarian cancer 3%). ADNEX: Assessment of Different NEoplasias in the adneXa model; LR2: Logistic Regression 2 model; RMI I: Risk of Malignancy Index I; ROC: receiver operating characteristic; ROMA: Risk of Ovarian Malignancy Algorithm.



Postmenopausal women

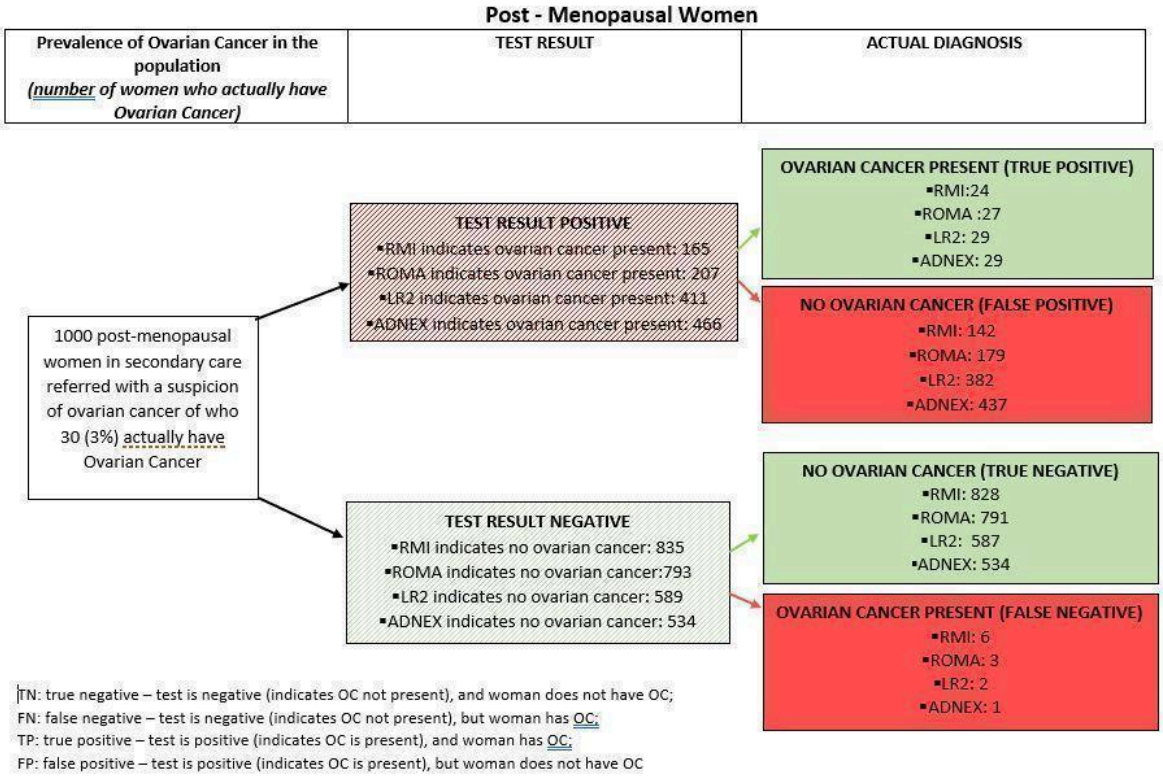
In postmenopausal women, ROMA at a threshold of 27.7 (± 2), LR2 (post-test probability 10%) and ADNEX (post-test probability 10%) demonstrated a higher sensitivity compared to RMI (ROMA: 90.3%, 95% CI 87.5% to 92.6%; LR2: 94.8%, 95% CI 92.3% to 96.6%; ADNEX: 97.6%, 95% CI 95.6% to 98.7%; RMI 78.4%, 95% CI 74.6% to 81.7%). Specificity of ROMA at a threshold of 27.7 (± 2) was comparable to RMI (ROMA: 81.5%, 95% CI 76.5% to 85.5%; RMI: 85.4%, 95% CI 82.0% to 88.2%), whereas for LR2 (post-test probability 10%) and ADNEX (post-test probability 10%), specificity was lower (LR2: 60.6%, 95% CI 50.5% to 69.9%; ADNEX: 55.0%, 95% CI 42.8% to 66.6%).

Based on our analysis, in a clinical setting with a pretest probability of OC of 3% in postmenopausal women, for every 1000 postmenopausal women tested:

- consequences of a positive test result:
  - an estimated 165 will have an RMI result indicating OC is present and of these 142 (86%) will not have OC;
  - an estimated 207 will have a ROMA result indicating OC is present and of these 179 (86%) will not have OC;
  - an estimated 411 will have an LR2 result indicating OC is present and of these 382 (93%) will not have OC;
  - an estimated 466 will have an ADNEX result indicating OC and of these 437 (94%) will not have OC;
- consequences of a negative test result:
  - of the 835 people with an RMI result indicating that OC is not present, 6 (0.7%) will actually have OC;
  - of the 793 people with a ROMA result indicating that OC is not present, 3 (0.4%) will actually have OC;
  - of the 492 people with an LR2 result indicating that OC is not present, 2 (0.4 %) will actually have OC;
  - of the 534 people with an ADNEX result indicating that OC is not present, 1 (0.1%) will actually have OC.

See Figure 12.

Figure 12. Illustration of the consequences of testing a hypothetical cohort of postmenopausal women referred from primary care (estimated prevalence of ovarian cancer 3%). ADNEX: Assessment of Different NEoplasias in the adneXa model; LR2: Logistic Regression 2 model; RMI I: Risk of Malignancy Index I; ROC: receiver operating characteristic; ROMA: Risk of Ovarian Malignancy Algorithm.



Considerations other than accuracy when deciding on ROMA, LR2 or ADNEX as alternative tests to RMI will include the relative costs and the feasibility of introducing ROMA or ADNEX. The adoption of ROMA does not rely on availability of expertise in USS, but would require investment in laboratory facilities for processing of HE4 tests. In addition, a decision is likely to be influenced by factors such as baseline risk (prevalence) of OC, which will be dependent on healthcare setting and menopausal status, and the adverse consequences of unnecessary investigation and treatment, for example, loss of fertility.

Strengths and weaknesses of the review

Strengths

This is the first review of test combinations for the diagnosis of OC to include and compare all tests currently used in clinical practice. Although literature searches were completed in 2019, this review remains the most up-to-date comprehensive review to our knowledge. We used sensitive search strategies to capture relevant literature regardless of country of publication, publication status (published or unpublished), language or clinical setting (primary care or specialist care (secondary and tertiary). Novel features of this review include systematic investigation of the effects of menopausal status and classification of borderline tumours on estimates of test accuracy and statistical comparison of tests relevant to clinical practice at the time of writing. We attempted to mitigate against heterogeneity by attempting to restrict our analysis to primary tumours of adnexal origin and where this was not possible or unclear in studies reporting mixed primary, recurrent and metastatic disease, this was reflected in downgrading of quality assessment.

Weaknesses

Due to time and resource constraints, we were unable to consider including non-English Language studies. The impact of this omission on study findings is unknown. We acknowledge a major limitation of this review is the search date, which at the time of writing is 2.5 years old. We cannot rule out the possibility that inclusion of more-recent studies will have changed our summary estimates of accuracy for each of the four included index tests. The potential impact on estimates of test accuracy of not including more recently published studies is likely to be less for RMI (19 included studies) and ROMA (40 included studies) compared to ADNEX (four included studies) and LR2 (five included studies). LR2 has been superseded by ADNEX as the multivariable USS model of choice in clinical practice; this clinical situation is reflected by the fact that in the intervening period between our 2015 and 2019 searches, only an additional two LR2 studies were identified for inclusion in this review. In recognition of the relatively small number of ADNEX studies included in our review, we performed

a scoping search for primary studies published since our search cut-off date of June 2019. The search found three studies, two single-centre studies (Chen 2019; Nam 2021), and one multicentre study (van Calster 2014). Only one study reported sensitivity and specificity separately in pre- and postmenopausal women (Nam 2021). Sensitivity and specificity were both 83% in premenopausal women and sensitivity was 100% and specificity was 76% in postmenopausal women at a threshold to achieve a post-test probability of OC of 10%. These estimates are in line with those from studies included in this review and we consider it unlikely that inclusion of this single eligible additional ADNEX study would alter the overall conclusions of this review regarding the relative performance of tests.

We also recognise the limitation on our review methods of the pragmatic decision to reduce the number of bibliographic databases searched for the review update (between 2015 and 2019). Although we developed the 2019 search strategy iteratively, testing the sensitivity of the search strategy using articles we had already identified as potentially eligible, we cannot rule out the possibility that eligible studies may have been missed.

The major limitation of our review is deficiencies in included studies. Lack of data and poor reporting in included studies precluded quality assessment and investigation of potential important sources of heterogeneity in test accuracy estimates. These included clinical setting (primary versus specialist), target condition (primary versus recurrent and metastatic disease), and cancer histological subtype and stage. Included studies varied with respect to the range of ovarian pathology included with some restricting inclusion to EOC whilst in others metastatic disease to the ovaries could not be disaggregated from primary OC for the purposes of analysis. A lack of distinction between pre- and postmenopausal women when evaluating test accuracy continues to be a major limitation of research in this area. Thirty-seven of 59 included studies were conducted in specialist gynaecological oncology centres in women scheduled for surgery. The method of presentation of these women was documented in only four included studies.

Applicability of findings to the review question

This review aimed to answer the question of the accuracy of imaging and biomarkers for women *with symptoms suspicious for* OC. In the UK, NICE and the RCOG recommend women with suspicious symptoms presenting in primary care should receive additional investigations with biomarkers and USS to determine further management (NICE 2011; RCOG 2016). The American College of Obstetrics and Gynaecology recommends TVS as the initial test of choice *if physical examination suggests the presence of an adnexal mass* (ACOG 2016).

The presence of suspicious symptoms is therefore a trigger for further investigation. Most included studies were at high or unclear applicability to the review question on the basis that women were either asymptomatic, or it was unclear if they were symptomatic, at the point of index test use. Further, we did not identify any studies of the accuracy of test combinations to diagnose OC in a generalist setting. Most included studies had a prevalence of OC that was in keeping with tertiary hospitals. Test accuracy estimates from this review are therefore unlikely to be applicable to generalist settings, where the prevalence of OC is lower and the spectrum of the tested population more heterogeneous.

With the exception of one study (Karlsen 2012), all included women had a confirmed adnexal mass at the point of testing. Karlsen 2012 had the lowest estimated specificity (53%) and one of the highest estimates of sensitivity (94%) (Figure 6). Early in the OC testing pathway it would be expected that test specificity would be lower, particularly in premenopausal women, reflecting a more diverse population in terms of comorbidity (e.g. endometriosis and functional benign tumours), and normal physiological processes such as the menstrual cycle, which are causes of false-positive test results and a lower test specificity. Thus in generalist settings, the relationship between sensitivity and specificity and menopausal status observed in this review may be reversed. The implication is that estimates of the accuracy of index tests in this review are likely to be applicable to women selected on the basis of prior tests in specialist settings (secondary and tertiary care), but are unlikely to be applicable to women without a confirmed adnexal mass (i.e. in primary care settings).

All studies of index tests with an USS component (RMI, LR2 and ADNEX) were at high or unclear risk of bias in the index test domain on the basis that sonographers were specialists or their level of skill was not reported. Therefore, we cannot assume that the performance of RMI, LR2 or ADNEX could be replicated by non-specialist sonographers as would be the case for investigations initiated in primary care or secondary care settings.

A further concern regarding the applicability of this review's findings is that in most studies, borderline tumours were either excluded or it was unclear how they were classified for estimation of test accuracy (excluded, classified as malignant or classified as benign). Borderline ovarian tumours account for an estimated 15% of ovarian tumours (Skirnisdottir 2008). In current clinical practice, borderline ovarian tumours undergo similar surgical management to invasive malignant tumours. We observed a decrease in sensitivity of 6.4 percentage points (95% CI 1.2 to 11.5) in ROMA studies of postmenopausal women grouping borderline tumours with malignant compared to studies where borderline tumours were excluded, or where categorisation of borderline tumours for analysis was unclear (Table 10). Exclusion of borderline tumours in studies in this review is therefore likely to have resulted in overestimation of sensitivity.

AUTHORS' CONCLUSIONS

Implications for practice

This review has demonstrated that menopausal status is associated with changes in disease spectrum, which is reflected in differences in test performance for women presenting with an adnexal mass. The implications of this finding for practice is that the utility of tests for diagnosing ovarian cancer (OC) should be considered separately in premenopausal and postmenopausal women.

Furthermore, current guidelines recommending the Risk of Malignancy Index (RMI) as a diagnostic or triage test in pre- and postmenopausal women in secondary care settings should be reviewed.

The Logistic Regression Model 2 (LR2) has been superseded by the Assessment of Different NEoplasias in the adneXa model (ADNEX) in clinical practice. The strength with which we can draw conclusions about the relative accuracy of Risk of Ovarian Malignancy Algorithm (ROMA) or ADNEX, as replacements to RMI, is undermined by

the relatively small number of included ADNEX studies. However, our scoping for more-recent ADNEX studies resulted in accuracy estimates within the range of present included studies. In spite of relatively wide confidence intervals for estimates of accuracy for ADNEX, we can still conclude that:

- for premenopausal women presenting to specialist settings with an adnexal mass suspicious for OC, ROMA and ADNEX both offer higher sensitivities compared to RMI, but at the expense of a decrease in specificity;
- for postmenopausal women, ROMA and ADNEX both offer higher sensitivities compared to RMI, but at the expense of a decrease in specificity for ADNEX.

The decision about which test (ROMA or ADNEX) should replace RMI will depend in part on how healthcare systems view the trade-off between sensitivity (false-negative diagnoses) and specificity (false-positive diagnoses). Inclusion of a larger number of ADNEX studies will improve precision and may reveal a distinction between the specificity of ADNEX and ROMA in premenopausal women.

The choice of which combination test (ROMA or ADNEX) should replace RMI in practice in secondary care will also require consideration of the relative costs and the feasibility of introducing the test. ADNEX offers a polynomial probability of histology, which is valuable information for counselling patients on treatment options. However, implementing tests based on USS models will require training in specialist USS skills and quality assurance processes, similar to those introduced for nuchal scans in early pregnancy. Implementing USS through dedicated 'pelvic mass clinics' may represent a method for achieving this. Implementing testing with ROMA will require investment in laboratory processes.

The implications of our findings for women presenting in generalist settings, and early in the diagnostic pathway in secondary care, is less clear. Participants in included studies had a confirmed adnexal mass and the presence of symptoms at the time of testing was mostly not reported. Prevalence of OC in premenopausal women in included studies was upwards of 9% and in postmenopausal women 40%. Included participants are therefore likely to represent a highly selected referred population, rather than a population in whom referral is being considered. The comparative accuracy of tests observed here may also not be stable when transferred to non-specialist settings.

Implications for research

Most studies in this review were conducted in specialist centres and the prevalence of OC in both pre- and postmenopausal women was typical of tertiary healthcare settings, ranging from 8% to 81% across included studies. No studies were identified in populations with a prevalence of OC typical of that seen at the point of first referral to hospital (e.g. rapid access clinics) or in community settings. Clinical setting has significant implications for the performance of diagnostic tests and the cost-benefit impact on a healthcare system. Research is urgently needed to evaluate tests for diagnosis of OC in community settings. Future studies performed earlier in the OC diagnostic pathway should also take care to report aspects of setting that will have a bearing on test performance such as healthcare setting (e.g. primary care or rapid access hospital clinic); presenting signs and symptoms and details of test conduct such as the skill of those eliciting symptoms; signs and conducting and interpreting imaging tests. In populations such as these that are more heterogeneous the use of rigorous clinical follow-up as a reference standard in index test negative cases should be pursued. Importantly, higher reporting standards of diagnostic test accuracy studies are required. This is a common and major limitation to systematic review of diagnostic test accuracy studies, as previously noted (Nagar 2021).

Primary studies should in future clearly report the occurrence of tumours found to be borderline at histology. Separate classification of these tumour types will ensure test accuracy research can be used flexibility, as knowledge advances about the malignant potential of such tumours and their most effective management.

ACKNOWLEDGEMENTS

We thank Jo Morrison for clinical and editorial advice. Gail Quinn, Clare Jess, Tracey Harrison, Jo Platt for their contribution to the editorial process. We thank Chris Partlett for help with data extraction *and April Coombe for helpful advice in responding to peer review comments about the search strategy update.*

This project was supported by the National Institute for Health Research (NIHR), via Cochrane Infrastructure funding to the Cochrane Gynaecological, Neuro-oncology and Orphan Cancer Group. The views and opinions expressed herein are those of the authors and do not necessarily reflect those of the Systematic Reviews Programme, NIHR, National Health Service or the Department of Health.

REFERENCES

References to studies included in this review

Abdalla 2017 *{published data only}*

Abdalla N, Piorkowski R, Pazura M, Cendrowski K, Sawicki W. Combined use of ultrasound and CA125 to predict malignancy in a patient with oviduct transitional carcinoma: case report. *BJOG: an International Journal of Obstetrics and Gynaecology* 2017;**124**(Suppl 1):169.

Al Musalhi 2016 *{published data only}*

Al Musalhi K, Al Kindi M, Al Aisary F, Ramadhan F, Al Rawahi T, Al Hatali K, et al. Evaluation of HE4, CA-125, risk of ovarian malignancy algorithm (ROMA) and risk of malignancy index (RMI) in the preoperative assessment of patients with adnexal mass. *Oman Medical Journal* 2016;**31**(5):336-44.

Anton 2012 *{published data only}*

Anton C, Carvalho FM, Oliveira EI, Maciel GA, Baracat EC, Carvalho JP. A comparison of CA125, HE4, risk ovarian malignancy algorithm (ROMA), and risk malignancy index (RMI) for the classification of ovarian masses. *Clinics (Sao Paulo, Brazil)* 2012;**67**(5):437-41.

Bandiera 2011 *{published data only}*

Bandiera E, Romani C, Specchia C, Zanotti L, Galli C, Ruggeri G, et al. Serum human epididymis protein 4 and risk for ovarian malignancy algorithm as new diagnostic and prognostic tools for epithelial ovarian cancer management. *Cancer Epidemiology, Biomarkers & Prevention* 2011;**20**(12):2496-506.

Chan 2013 *{published data only}*

* Chan KK, Chen CA, Nam JH, Ochiai K, Wilailak S, Choon AT, et al. The use of HE4 in the prediction of ovarian cancer in Asian women with a pelvic mass. *Gynecologic Oncology* 2013;**128**(2):239-44.

Wilailak S, Chan KK, Chen CA, Nam JH, Ochiai K, Aw TC, et al. Distinguishing benign from malignant pelvic mass utilizing an algorithm with HE4, menopausal status, and ultrasound findings. *Journal of Gynecologic Oncology* 2015;**26**(1):46-53. [Secondary publication]

Chen 2014 *{published data only}*

Chen WT, Gao X, Han XD, Zheng H, Guo L, Lu RQ. HE4 as a serum biomarker for ROMA prediction and prognosis of epithelial ovarian cancer. *Asian Pacific Journal of Cancer Prevention* 2014;**15**(1):101-5.

Chen 2015 *{published data only}*

Chen X, Zhou H, Chen R, He J, Wang Y, Huang L, et al. Development of a multimarker assay for differential diagnosis of benign and malignant pelvic masses. *Clinica Chimica Acta* 2015;**440**:57-63.

Chudecka-Glaz 2015 *{published data only}*

Chudecka-Glaz AM. ROMA, an algorithm for ovarian cancer. *Clinica Chimica Acta* 2015;**440**:143-51.

Cradic 2018 *{published data only}*

Cradic KW, Lasho MA, Algeciras-Schimnich A. Validation of the cut-points recommended for ROMA using the Roche Elecsys CA125 and HE4 assays. *Annals of Clinical & Laboratory Science* 2018;**48**(1):90-3.

Dikmen 2015 *{published data only}*

Dikmen ZG, Colak A, Dogan P, Tuncer S, Akbiyik F. Diagnostic performances of CA125, HE4, and ROMA index in ovarian cancer. *European Journal of Gynaecological Oncology* 2015;**36**(4):457-62.

Ertas 2016 *{published data only}*

Ertas S, Vural F, Vural F, Tufekci EC, Ertas AC, Kose G, et al. Predictive value of malignancy risk indices for ovarian masses in premenopausal and postmenopausal women. *Asian Pacific Journal of Cancer Prevention* 2016;**17**(4):2177-83.

Farzaneh 2014 *{published data only}*

Farzaneh F, Honarvar Z, Yaraghi M, Yaseri M, Arab M, Hosseini M, et al. Preoperative evaluation of risk of ovarian malignancy algorithm index in prediction of malignancy of adnexal masses. *Iranian Red Crescent Medical Journal* 2014;**16**(6):e17185.

Grenache 2015 *{published data only}*

Grenache DG, Heichman KA, Werner TL, Vucetic Z. Clinical performance of two multi-marker blood tests for predicting malignancy in women with an adnexal mass. *Clinica Chimica Acta* 2015;**438**:358-63.

Huy 2018 *{published data only}*

Huy NV, van Khoa V, Tam LM, Vinh TQ, Tung NS, Thanh CN, et al. Standard and optimal cut-off values of serum ca-125, HE4 and ROMA in preoperative prediction of ovarian cancer in Vietnam. *Gynecologic Oncology Reports* 2018;**25**:110-4.

Irshad 2013 *{published data only}*

Irshad F, Irshad M, Naz M, Asim Ikram M. Accuracy of "risk of malignancy index" in the preoperative diagnosis of ovarian malignancy in post menopausal women. *Rawal Medical Journal* 2013;**38**(3):266-70.

Kadija 2012 *{published data only}*

Kadija S, Stefanovic A, Jeremic K, Radojevic MM, Nikolic L, Markovic I, et al. The utility of human epididymal protein 4, cancer antigen 125, and risk for malignancy algorithm in ovarian cancer and endometriosis. *International Journal of Gynecological Cancer* 2012;**22**(2):238-44.

Karlsen 2012 *{published data only}*

Karlsen MA, Sandhu N, Hogdall C, Christensen IJ, Nedergaard L, Lundvall L, et al. Evaluation of HE4, CA125, risk of ovarian malignancy algorithm (ROMA) and risk of malignancy index (RMI) as diagnostic tools of epithelial ovarian cancer in patients with a pelvic mass. *Gynecologic Oncology* 2012;**127**(2):379-83.

Kim 2011 *{published data only}*

Kim YM, Whang DH, Park J, Kim SH, Lee SW, Park HA, et al. Evaluation of the accuracy of serum Human Epididymis protein

4 in combination with CA125 for detecting ovarian cancer: a prospective case-control study in a Korean population. *Clinical Chemistry & Laboratory Medicine* 2011;**49**(3):527-34.

Kim 2019 *{published data only}*

Kim B, Park Y, Ahn HJ, Lee KA, Chung JE, Han SW. Diagnostic performance of CA 125, HE4, and risk of Ovarian Malignancy Algorithm for ovarian cancer. *Journal of Clinical Laboratory Analysis* 2019;**33**(1):e22624.

Krascsenitis 2016 *{published data only}*

Krascsenits G, Balazs B, Dudnyikova A, Purcsi K, Orosz E, Pete I. Investigating the predictive value of RMI and ROMA indices in patients with ovarian tumors of uncertain dignity. *Magyar Onkologia* 2016;**60**(4):320-7.

Li 2016 *{published data only}*

Li L, Wan J, Cai G, Yuan L, Liang J, Song J, et al. Value of serum human epididymis secretory protein 4 as a marker for differential diagnosis of malignant and benign gynecological diseases of patients in southern China. *Clinica Chimica Acta* 2016;**459**:170-6.

Liest 2019 *{published data only}*

Liest AL, Omran AS, Mikiver R, Rosenberg P, Uppugunduri S. RMI and ROMA are equally effective in discriminating between benign and malignant gynecological tumors: a prospective population-based study. *Acta Obstetricia et Gynecologica Scandinavica* 2019;**98**(1):24-33.

Lycke 2018 *{published data only}*

Lycke M, Kristjansdottir B, Sundfeldt K. A multicenter clinical trial validating the performance of HE4, CA125, risk of ovarian malignancy algorithm and risk of malignancy index. *Gynecologic Oncology* 2018;**151**(1):159-65.

Manegold-Brauer 2016 *{published data only}*

Manegold-Brauer G, Buechel J, Knipprath-Meszaros A, Schoetzau A, Hacker NF, Tercanli S, et al. Improved detection rate of ovarian cancer using a 2-step triage model of the risk of malignancy index and expert sonography in an outpatient screening setting. *International Journal of Gynecological Cancer* 2016;**26**(6):1062-9.

Melo 2018 *{published data only}*

Melo A, Verissimo R, Farinha M, Martins NN, Martins FN. Discriminative value of CA-125, HE4, risk of Malignancy Index II (RMI-II) and risk of malignancy algorithm (ROMA) in the differential diagnosis of pelvic masses: conclusions from a referral centre in Portugal. *Journal of Obstetrics & Gynaecology* 2018;**38**(8):1140-5.

Meys 2017 *{published data only}*

Meys EM, Jeelof LS, Achten NM, Slangen BF, Lambrechts S, Kruitwagen RF, et al. Estimating risk of malignancy in adnexal masses: external validation of the ADNEX model and comparison with other frequently used ultrasound methods. *Ultrasound in Obstetrics & Gynecology* 2017;**49**(6):784-92.

Molina 2011 *{published data only}*

Molina R, Escudero JM, Auge JM, Filella X, Foj L, Torne A, et al. HE4 a novel tumour marker for ovarian cancer: comparison with CA 125 and ROMA algorithm in patients with gynaecological diseases. *Tumour Biology* 2011;**32**(6):1087-95. [1010-4283]

Montagnana 2011 *{published data only}*

Montagnana M, Danese E, Ruzzenente O, Bresciani V, Nuzzo T, Gelati M, et al. The ROMA (Risk of Ovarian Malignancy Algorithm) for estimating the risk of epithelial ovarian cancer in women presenting with pelvic mass: is it really useful? *Clinical Chemistry & Laboratory Medicine* 2011;**49**(3):521-5.

Moore 2009 *{published data only}*

Moore RG, McMeekin DS, Brown AK, DiSilvestro P, Miller MC, Allard WJ, et al. A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecologic Oncology* 2009;**112**(1):40-6.

Moore 2011 *{published data only}*

Moore RG, Miller MC, Disilvestro P, Landrum LM, Gajewski W, Ball JJ, et al. Evaluation of the diagnostic accuracy of the risk of ovarian malignancy algorithm in women with a pelvic mass. *Obstetrics & Gynecology* 2011;**118**(2 Pt 1):280-8.

Niemi 2017 *{published data only}*

Niemi RJ, Saarelainen SK, Luukkaala TH, Maenpaa JU. Reliability of preoperative evaluation of postmenopausal ovarian tumors. *Journal of Ovarian Research* 2017;**10**(1):15.

Nikolova 2016 *{published data only}*

Nikolova T, Zivadinovic R, Evtimovska N, Klisarovska V, Stanojevic M, Georgievska J, et al. Diagnostic performance of Human Epididymis protein4 compared to a combination of biophysical and biochemical markers to differentiate ovarian endometriosis from epithelial ovarian cancer in premenopausal women. *Journal of Obstetrics and Gynaecology Research* 2017;**43**(12):1870-9.

Novotny 2012 *{published data only}*

Novotny Z, Presl J, Kucera R, Topolcan O, Vrzalova J, Fuchsova R, et al. HE4 and ROMA index in Czech postmenopausal women. *Anticancer Research* 2012;**32**(9):4137-40.

Ortiz-Munoz 2014 *{published data only}*

Ortiz-Munoz B, Aznar-Oroval E, Garcia Garcia A, Covisa Peris A, Perez Ballestero P, Sanchez Yepes M, et al. HE4, Ca125 and ROMA algorithm for differential diagnosis between benign gynaecological diseases and ovarian cancer. *Tumour Biology* 2014;**35**(7):7249-58.

Park 2019 *{published data only}*

Park H, Shin J, Lee DW, Kim MJ, Lee HN. Diagnostic accuracy of the risk of Ovarian Malignancy Algorithm in clinical practice at a single hospital in Korea. *Annals of Laboratory Medicine* 2019;**39**(3):252-62.

Partheen 2011a *{published data only}*

Partheen K, Kristjansdottir B, Sundfeldt K. Evaluation of ovarian cancer biomarkers HE4 and CA-125 in women presenting with a

suspicious cystic ovarian mass. *Journal of Gynecologic Oncology* 2011;**22**(4):244-52.

Prskalo 2015 *{published data only}*

Prskalo ZS, Gace M, Dobrijevic S, Mayer L. Benefits human epididymis protein (HE4) compared to traditional used tumor markers in gynecological oncology. *Libri Oncologici* 2015;**43**(1-3):9-14.

Radosa 2011 *{published data only}*

Radosa MP, Camara O, Vorwergk J, Diebolder H, Winzer H, Mothes A, et al. Preoperative multimodal strategies for risk assessment of adnexal masses: analysis of 1362 cases in a gynecologic cancer center. *International Journal of Gynecological Cancer* 2011;**21**(6):1056-62.

Richards 2015 *{published data only}*

Richards A, Herbst U, Pather S, Saidi S, Tejada-Berges T, Williams P, et al. HE4, CA125, the Risk of Malignancy Algorithm and the Risk of Malignancy Index and complex pelvic masses – a prospective comparison in the pre-operative evaluation of pelvic masses in an Australian population. *Australian & New Zealand Journal of Obstetrics & Gynaecology* 2015;**55**(5):493-7.

Romagnolo 2016 *{published data only}*

Romagnolo C, Leon A, Fabricio AS, Taborelli M, Polesel J, Del Pup L, et al. HE4, CA125 and risk of ovarian malignancy algorithm (ROMA) as diagnostic tools for ovarian cancer in patients with a pelvic mass: an Italian multicenter study. *Gynecologic Oncology* 2016;**141**(2):303-11.

Salim 2018 *{published data only}*

Salim E, Zubairi AM, Danish SH, Ali U. Diagnostic accuracy of Risk of Ovarian Malignancy Algorithm (ROMA) in post-menopausal patients with ovarian mass. *Journal of the College of Physicians and Surgeons Pakistan* 2018;**28**(6):440-4.

Sayasneh 2013a *{published data only}*

Sayasneh A, Kaijser J, Preisler J, Johnson S, Stalder C, Husicka R, et al. A multicenter prospective external validation of the diagnostic performance of IOTA simple descriptors and rules to characterize ovarian masses. *Gynecologic Oncology* 2013;**130**(1):140-6. [Sayasneh 2013]

* Sayasneh A, Wynants L, Preisler J, Kaijser J, Johnson S, Stalder C, et al. Multicentre external validation of IOTA prediction models and RMI by operators with varied training. *British Journal of Cancer* 2013;**108**(12):2448-54.

Shen 2017 *{published data only}*

Shen F, Lu S, Peng Y, Yang F, Chen Y, Lin Y, et al. Performance of ROMA based on Architect CA 125 II and HE4 values in Chinese women presenting with a pelvic mass: a multicenter prospective study. *Clinica Chimica Acta* 2017;**471**:119-25.

Stiekma 2014 *{published data only}*

Stiekema A, Lok CA, Kenter GG, van Driel WJ, Vincent AD, Korse CM. A predictive model combining human epididymal protein 4 and radiologic features for the diagnosis of ovarian cancer. *Gynecologic Oncology* 2014;**132**(3):573-7.

Szubert 2016a *{published data only}*

Szubert S, Wojtowicz A, Moszynski R, Zywica P, Dyczkowski K, Stachowiak A, et al. External validation of the IOTA ADNEX model performed by two independent gynecologic centers. *Gynecologic Oncology* 2016;**142**(3):490-5.

Szubert 2016b *{published data only}*

Szubert S, Wojtowicz A, Moszynski R, Zywica P, Dyczkowski K, Stachowiak A, et al. External validation of the IOTA ADNEX model performed by two independent gynecologic centers. *Gynecologic Oncology* 2016;**142**(3):490-5.

Teh 2018 *{published data only}*

Teh BH, Yong SL, Sim W, Lau KB, Suharjono HN. Evaluation in the predictive value of serum human epididymal protein 4 (HE4), cancer antigen 125 (CA 125) and a combination of both in detecting ovarian malignancy. *Hormone Molecular Biology and Clinical Investigation* 2018;**35**:1.

Terlikowska 2016 *{published data only}*

Terlikowska KM, Dobrzycka B, Witkowska AM, Mackowiak-Matejczyk B, Sledziewski TK, Kinalski M, et al. Preoperative HE4, CA125 and ROMA in the differential diagnosis of benign and malignant adnexal masses. *Journal of Ovarian Research* 2016;**9**(1):43.

Terzic 2013 *{published data only}*

Terzic M, Dotlic J, Likic I, Brndusic N, Pilic I, Ladjevic N, et al. Risk of malignancy index validity assessment in premenopausal and postmenopausal women with adnexal tumors. *Taiwanese Journal of Obstetrics & Gynecology* 2013;**52**(2):253-7.

Testa 2014 *{published data only}*

Testa A, Kaijser J, Wynants L, Fischerova D, van Holsbeke C, Franchi D, et al. Strategies to diagnose ovarian cancer: new evidence from phase 3 of the multicentre international IOTA study. *British Journal of Cancer* 2014;**111**(4):680-8.

Timmerman 2010 *{published data only}*

Timmerman D, Ameye L, Fischerova D, Epstein E, Melis GB, Guerriero S, et al. Simple ultrasound rules to distinguish between benign and malignant adnexal masses before surgery: prospective validation by IOTA group. *BMJ* 2010;**341**:c6839.

van Calster 2014 *{published data only}*

van Calster B, van Hoorde K, Valentin L, Testa AC, Fischerova D, van Holsbeke C, et al. Evaluating the risk of ovarian cancer before surgery using the ADNEX model to differentiate between benign, borderline, early and advanced stage invasive, and secondary metastatic tumours: prospective multicentre diagnostic study. *BMJ* 2014;**349**:g5920.

van den Akker 2016 *{published data only}*

van den Akker PA, Zusterzeel PL, Aalders AL, Snijders MP, Samlal RA, Vollebergh JH, et al. Use of risk of malignancy index to indicate frozen section analysis in the surgical care of women with ovarian tumors. *International Journal of Gynaecology & Obstetrics* 2016;**133**(3):355-8.

van Gorp 2011 *{published data only}*

van Gorp T, Cadron I, Despierre E, Daemen A, Leunen K, Amant F, et al. HE4 and CA125 as a diagnostic test in ovarian cancer: prospective validation of the Risk of Ovarian Malignancy Algorithm. *British Journal of Cancer* 2011;**104**(5):863-70.

van Gorp 2012 *{published data only}*

Kaijser J, van Gorp T, Smet ME, van Holsbeke C, Sayasneh A, Epstein E, et al. Are serum HE4 or ROMA scores useful to experienced examiners for improving characterization of adnexal masses after transvaginal ultrasonography? *Ultrasound in Obstetrics & Gynecology* 2014;**43**(1):89-97. [Secondary publication Kaijser 2014]

* van Gorp T, Veldman J, van Calster B, Cadron I, Leunen K, Amant F, et al. Subjective assessment by ultrasound is superior to the risk of malignancy index (RMI) or the risk of ovarian malignancy algorithm (ROMA) in discriminating benign from malignant adnexal masses. European Journal of Cancer 2012;**48**(11):1649-56.

Vural 2016 *{published data only}*

Vural F, Aka N, Ertas S, Kose G, Tufekci EC. The ovarian cancers in geriatric population: the validity of inflammatory markers, malignancy risk indices 1, 2, 3, 4, and CA-125 levels in malignancy discrimination of adnexal masses. *European Journal of Gynaecological Oncology* 2016;**37**(6):846-51.

Xu 2016 *{published data only}*

Xu Y, Zhong R, He J, Ding R, Lin H, Deng Y, et al. Modification of cut-off values for HE4, CA125 and the ROMA algorithm for early-stage epithelial ovarian cancer detection: results from 1021 cases in South China. *Clinical Biochemistry* 2016;**49**(1-2):32-40.

Zhang 2015 *{published data only}*

Zhang P, Wang C, Cheng L, Zhang P, Guo L, Liu W, et al. Comparison of HE4, CA125, and ROMA diagnostic accuracy: a prospective and multicenter study for Chinese women with epithelial ovarian cancer. *Medicine* 2015;**94**(52):e2402.

Zhang 2019 *{published data only}*

Zhang L, Chen Y, Wang K. Comparison of CA125, HE4, and ROMA index for ovarian cancer diagnosis. *Current Problems in Cancer* 2019;**43**(2):135-44.

Additional references

ACOG 2016

ACOG. Evaluation and management of adnexal masses. *Obstetrics Gynecology* 2016;**128**(5):e210-26.

American Cancer Society 2016

American Cancer Society. Cancer facts and figures 2016. Atlanta (GA). www.cancer.org/acs/groups/content/@research/documents/document/acspc-047079.pdf (accessed prior to 29 June 2022).

Anderson 2009

Anderson GL, McIntosh M, Wu L, Barnett M, Goodman G, Thorpe JD, et al. Assessing lead time of selected ovarian cancer biomarkers: a nested case-control study. *Journal of the National Cancer Institute* 2010;**102**(1):26-38.

Atkurk 2011

Akturk E, Karaca RE, Alanbay I, Dede M, Karasahin E, Yenen MC, et al. Comparison of four malignancy risk indices in the detection of malignant ovarian masses. *Journal of Gynecologic Oncology* 2011;**22**(3):177-82.

Bankhead 2005

Bankhead CR, Kehoe ST, Austoker J. Symptoms associated with diagnosis of ovarian cancer: a systematic review. *BJOG: an International Journal of Obstetrics and Gynaecology* 2005;**112**(7):857-65.

Bray 2018

Bray F, Ferlay J, Soerjomataram I, Siegal RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a Cancer Journal for Clinicians* 2018;**68**:394-424.

Cancer Research UK 2017

Cancer Research UK. Ovarian cancer statistics 2016–2018. www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer#heading-Zero (accessed prior to 29 June 2022).

Chacon 2019

Chacón E, Dasí J, Caballero C, Alcázar JL. Risk of ovarian malignancy algorithm versus risk malignancy index-I for preoperative assessment of adnexal masses: a systematic review and meta-analysis. *Gynecologic and Obstetric Investigation* 2019;**84**(6):591-8.

Chen 2019

Chen H, Qian L, Jiang M, Du Q, Yuan F, Feng W. Performance of IOTA ADNEX model in evaluating adnexal masses in a gynecological oncology center in China. *Ultrasound in Obstetrics and Gynaecology* 2019;**54**(6):815-22.

Cho 2009

Cho KR, Shih I. Ovarian cancer. *Annual Review of Pathology Mechanisms of Disease* 2009;**4**:287-313.

Chu 2006

Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology* 2006;**59**(12):1331-2.

Deeks 2005

Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of Clinical Epidemiology* 2005;**58**(9):882-93.

Di Legge 2012

Di Legge A, Testa AC, Ameye L, van Calster B, Lissoni AA, Leone FP, et al. Lesion size affects diagnostic performance of IOTA logistic regression models, IOTA simple rules and risk of malignancy index in discriminating between benign

and malignant adnexal masses. *Ultrasound in Obstetrics & Gynecology* 2012;**40**:345-54. [DOI: org/10.1002/uog.11167]

Dodge 2012

Dodge JE, Covens AL, Lacchetti C, Elit LM, Le T, Devries-Aboud M, et al, and the Gynaecology Cancer Disease Site Group. Preoperative identification of a suspicious adnexal mass: a systematic review and meta-analysis. *Gynecologic Oncology* 2012;**126**:157-66.

Drapkin 2005

Drapkin R, von Horsten HH, Lin Y, Mok SC, Crum CP, Welch WR, et al. Human Epididymis protein 4 (HE4) is a secreted glycoprotein that is overexpressed by serous and endometrioid ovarian carcinomas. *Cancer Research* 2005;**65**(6):2162-9.

Duklewski 2009

Duklewski K, Aronson A. Ovarian cysts. www.mdguidelines.com/ovarian-cyst-benign (accessed prior to 29 June 2022).

Ellis-Brookes 2012

Elliss-Brookes L, McPhail S, Ives A, Greenslade M, Shelton J, Hiom S, et al. Routes to diagnosis for cancer – determining the patient journey using multiple routine data sets. *British Journal of Cancer* 2012;**107**:1220-6.

Fakhar 2018

Fakhar H, Rezaei Tavirini M, Zali H, Darkhaneh R, Leili E, Aghazadeh M. Comparison of serum Human Epididymis protein (HE4), carbohydrate antigen 125 (CA125) and Risk of Ovarian Malignancy Algorithm (ROMA) as markers in ovarian cancer: a systematic review and a meta-analysis. *Indian Journal of Gynecologic Oncology* 2018;**16**(10):174-5.

Ferraro 2013

Ferraro S, Braga F, Lanzoni M, Boracchi P, Biganzoli EM, Panteghini M. Serum human epididymis protein 4 vs carbohydrate antigen 125 for ovarian cancer diagnosis: a systematic review. *Journal of Clinical Pathology* 2013;**66**(4):273-81.

Fitch 2002

Fitch M, Deane K, Howell D, Gray RE. Women's experiences with ovarian cancer: reflections on being diagnosed. *Canadian Oncology Nursing Journal (Revue Canadienne de Nursing Oncologique)* 2002;**12**(3):152-68.

Geomini 2009

Geomini P, Kruitwagen R, Bremer GL, Cnossen J, Mol BW. The accuracy of risk scores in predicting ovarian malignancy: a systematic review. *Obstetrics and Gynecology* 2009;**113**(2 Pt 1):384-94.

Hamilton 2009

Hamilton W, Peters TJ, Bankhead C, Sharp D. Risk of ovarian cancer in women with symptoms in primary care: population based case-control study. *BMJ* 2009;**339**:b2998.

Hellstorm 2003

Hellström I, Raycraft J, Hayden-Ledbetter M, Ledbetter JA, Schummer M, McIntosh M, et al. The HE4 (WFDC2) protein is a biomarker for ovarian carcinoma. *Cancer Research* 2003;**63**(13):3695-700.

Holcomb 2011

Holcomb K, Vucetic Z, Miller MC, Knapp RC. Human epididymis protein 4 offers superior specificity in the differentiation of benign and malignant adnexal masses in premenopausal women. *American Journal of Obstetrics and Gynecology* 2011;**205**(4):358.e1-6.

Jacobs 1989

Jacobs I, Bast RC Jr. The CA 125 tumour-associated antigen: a review of the literature. *Human Reproduction* 1989;**4**(1):1-12.

Jacobs 1990

Jacobs I, Oram D, Fairbanks J, Turner J, Frost C, Grudzinskas JG. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. *British Journal of Obstetrics and Gynaecology* 1990;**97**(10):922-9.

Kaijser 2014

Kaijser J, Sayasneh A, van Hoorde K, Ghaem-Maghami S, Bourne T, Timmerman D, et al. Presurgical diagnosis of adnexal tumours using mathematical models and scoring systems: a systematic review and meta-analysis. *Human Reproduction Update* 2014;**20**(3):449-62.

Kobayashi 2012

Kobayashi E, Ueda Y, Matsuzaki S, Yokoyama T, Kimura T, Yoshino K, et al. Biomarkers for screening, diagnosis, and monitoring of ovarian cancer. *Cancer Epidemiology, Biomarkers & Prevention* 2012;**21**(11):1902-12.

Lancaster 2015

Lancaster JM, Powell CB, Chen LM, Richardson DL. Society of Gynecologic Oncology statement on risk assessment for inherited gynecologic cancer predispositions. SGO Clinical Practice Committee [published erratum appears in Gynecol Oncol 2015;138:765]. *Gynecologic Oncology* 2015;**136**:3-7.

Ledermann 2013

Ledermann JA, Raja FA, Fotopoulou C, Gonzalez-Martin A, Colombo N, Sessa C, on behalf of the ESMO Guidelines Working Group. Newly diagnosed and relapsed epithelial ovarian carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* 2013;**24**(Suppl 6):vi24-32.

Li 2012

Li F, Tie R, Chang K, Wang F, Deng S, Lu W, et al. Does risk for ovarian malignancy algorithm excel human epididymis protein 4 and CA125 in predicting epithelial ovarian cancer: a meta-analysis. *BMC Cancer* 2012;**12**:258.

Meys 2016

Meys EM, Kaijser J, Kruitwagen RF, Slangen BF, van Calster B, Aertgeerts B, et al. Subjective assessment versus ultrasound

models to diagnose ovarian cancer: a systematic review and meta-analysis. *European Journal of Cancer* 2016;**58**:17e29.

Moore 2012

Moore RG, Miller MC, Steinhoff MM, Skates SJ, Lu KH, Lambert-Messerlian G, et al. Serum HE4 levels are less frequently elevated than CA-125 in women with benign gynecologic disorders. *American Journal of Obstetrics and Gynecology* 2012;**206**(4):351 e1-8.

Moss 2005

Moss EL, Hollingworth J, Reynolds TM. The role of CA125 in clinical practice. *Journal of Clinical Pathology* 2005;**58**(3):308-12.

Nagar 2021

Nagar H, Wietek N, Goodall RJ, Hughes W, Schmidt-Hansen M, Morrison J. Sentinel node biopsy for diagnosis of lymph node involvement in endometrial cancer. *Cochrane Database of Systematic Reviews* 2021, Issue 6. Art. No: CD013021. [DOI: 10.1002/14651858.CD013021.pub2]

Nam 2021

Nam G, Lee SR, Jeong K, Moon HS, Chae HD. Assessment of different NEoplasias in the adneXa model for differentiation of benign and malignant adnexal masses in Korean women. *Obstetrics and Gynaecology Science* 2021;**64**(3):293-9.

NICE 2011

National Institute for Health and Care Excellence. CG 122: ovarian cancer: the recognition and initial management of ovarian cancer. www.nice.org.uk/guidance/cg122/chapter/guidance (accessed prior to 29 June 2022).

NICE 2017

National Institute for Health and Care Excellence. Suspected cancer: recognition and referral. www.nice.org.uk/guidance/ng12 (accessed 22 May 2019).

Nunes 2014

Nunes N, Ambler G, Foo X, Naftalin J, Widschwendter M. Use of IOTA simple rules for diagnosis of ovarian cancer: meta-analysis. *Ultrasound in Obstetrics & Gynecology* 2014;**44**:503-14.

Office for National Statistics 2016

Office for National Statistics. Office for National Statistics (ONS) 2016-based life expectancies and population projections. www.ons.gov.uk/releases/nationalpopulationprojections2016basedstatisticalbulletin (accessed December 2017).

Posadas 2004

Posadas EM, Davidson B, Kohn EC. Proteomics and ovarian cancer: implications for diagnosis and treatment: a critical review of the recent literature. *Current Opinion in Oncology* 2004;**16**(5):478-84.

Rai 2015

Rai N, Nevin J, Downey G, Abedin P, Balogun M, Kehoe S, et al. Outcomes following implementation of symptom triggered diagnostic testing for ovarian cancer. *European Journal of Obstetrics, Gynecology and Reproductive Biology* 2015;**187**:64-9.

RCOG 2011

Royal College of Obstetricians and Gynaecologists. Guideline no 62: management of suspected ovarian masses in premenopausal women. www.rcog.org.uk/globalassets/documents/guidelines/gtg_62.pdf (accessed prior to 29 June 2022).

RCOG 2016

Royal College of Obstetricians and Gynaecologists. Guideline no 34: the management of ovarian cysts in post menopausal women. www.rcog.org.uk/en/guidelines-research-services/guidelines/gtg34/ (accessed prior to 29 June 2022).

Reitsma 2005

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 2005;**58**(10):982-90.

Review Manager 2014 [Computer program]

Nordic Cochrane Centre, The Cochrane Collaboration Review Manager (RevMan). Version 5.4. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Rutter 2001

Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine* 2001;**20**(19):2865-84.

SAS 2015 [Computer program]

SAS Institute Inc SAS. Version 9.4. Cary (NC): SAS Institute Inc, 2015. Available at www.sas.com.

Shepherd 2000

Shepherd JE. Current strategies for prevention, detection, and treatment of ovarian cancer. *Journal of the American Pharmaceutical Association* 2000;**40**(3):392-401.

Shih 2004

Shih I, Kurman RJ. Ovarian tumorigenesis: a proposed model based on morphological and molecular genetic analysis. *American Journal of Pathology* 2004;**164**:1511-8.

Skirnisdottir 2008

Skírnisdóttir I, Garmo H, Wilander E, Holmberg L. Borderline ovarian tumors in Sweden 1960-2005: trends in incidence and age at diagnosis compared to ovarian cancer. *International Journal of Cancer* 2008;**123**(8):1897-901.

Smittenaar 2016

Smittenaar CR, Petersen KA, Stewart K, Moitt N. Cancer incidence and mortality projections in the UK until 2035. *British Journal of Cancer* 2016;**115**(9):1147-55.

StataCorp 2019 [Computer program]

StataCorp. Version 16. College Station, TX, USA: StataCorp, 2019. Available at www.stata.com.

Stukan 2015

Stukan M, Dudziak M, Ratajczak K, Grabowski JP. Usefulness of diagnostic indices comprising clinical, sonographic, and

biomarker data for discriminating benign from malignant ovarian masses. *Ultrasound Medicine* 2015;**34**:207-17.

Takwoingi 2010 [Computer program]

MetaDAS: a SAS macro for metaanalysis of diagnostic accuracy studies. User Guide Version 1.3. Takwoingi Y, Deeks JJ, Version 1.3. Available from: srdta.cochrane.org/, 2010.

Takwoingi 2017

Takwoingi Y, Guo B, Riley RD, Deeks JJ. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Statistical Methods in Medical Research* 2017;**26**(4):1896-911.

Timmerman 2016

Timmerman D, van Calster B, Tests A, Savelli L, Fischerova D, Froyman W, et al. Predicting the risk of malignancy in adnexal masses based on the simple rules from the International Ovarian Tumor Analysis group. *American Journal of Obstetrics and Gynaecology* 2016;**214**(4):424-37.

Walsh 2011

Walsh T, Casadei S, Lee MK, Pennil CC, Nord AS, Thornton AM, et al. Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 2011;**108**(44):18032-7.

Wang 2014

Wang J, Gao J, Yao H, Wu Z, Wang M, Qi J. Diagnostic accuracy of serum HE4, CA125 and ROMA in patients with ovarian cancer: a meta-analysis. *Tumour Biology* 2014;**35**:6127-38.

Whiting 2011

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529-36.

Wiegand 2010

Wiegand KC, Shah SP, Al-Agha OM, Zhao Y, Tse K, Zeng T, et al. ARID1A mutations in endometriosis-associated ovarian carcinomas. *New England Journal of Medicine* 2010;**363**(16):1532-43.

Wolf 2019

Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine* 2019;**170**:51-8.

References to other published versions of this review

Rai 2015b

Rai N, Champaneria R, Snell K, Mallett S, Bayliss SE, Neal RD, et al. Symptoms, ultrasound imaging and biochemical markers alone or in combination for the diagnosis of ovarian cancer in women with symptoms suspicious of ovarian cancer. *Cochrane Database of Systematic Reviews* 2015, Issue 12. Art. No: CD011964. [DOI: 10.1002/14651858.CD011964]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

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

Abdalla 2017

Study characteristics	
Patient Sampling	Country: Poland Centres: single Study design: within-person comparison Recruitment: prospective Method of patient selection: unclear Inappropriate exclusions: presence of fibroids > 5 cm were excluded
Patient characteristics and setting	Clinical setting: mixed Study entry criteria: patients scheduled to undergo surgery for adnexal tumours Sample size: 312 Age range: 18–85 years

Abdalla 2017 *(Continued)*

	Mean age: not reported Percentage postmenopausal (n): 37.5% (117)
Index tests	Test: RMI Prior test: ultrasound and measurement of tumour markers CA125 and HE4 Threshold for test positivity predefined: yes Threshold for test positivity: 200 Type of ultrasound (TAS, TVS, or both): both Operator experience of sonographer (generalist, specialist or trainee): not reported Type of technology or manufacturer of biomarker test: ultrasound performed with ultrasound apparatus Philips iU22. CA125 and HE4 measured via electrochemiluminescence immunoassay performed using a Cobas 8000 e602 apparatus
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 260, borderline 7, malignant 45, metastatic and others not reported
Flow and timing	
Comparative	N/A
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			

Abdalla 2017 *(Continued)*

C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound		
Are there concerns that the included patients and setting do not match the review question?		Unclear
DOMAIN 2: Index Test (ADNEX)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Unclear

Abdalla 2017 *(Continued)*

DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ROMA)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (LR2)
Were the index test results interpreted without knowledge of the results of the reference standard?

Abdalla 2017 *(Continued)*

If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			

Abdalla 2017 *(Continued)*

For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	
Could the conduct of the comparative studies have introduced bias?	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	

Al Musalhi 2016

Study characteristics	
Patient Sampling	Country: Oman Centres: single Study design: within-person comparison Recruitment: prospective method of patient selection: convenience Inappropriate exclusions: none
Patient characteristics and setting	Clinical setting: mixed Study entry criteria: patients with an ovarian mass Sample size: 213 Age range: not reported Mean age: not reported Percentage postmenopausal (n): 24% (51)
Index tests	Test: RMI I and ROMA Prior test: presume USS Threshold for test positivity predefined: yes Threshold for test positivity: ROMA: premenopausal 13.1, postmenopausal 27.7, RMI I: 200 Type of ultrasound (TAS, TVS or both): TVS Operator experience of sonographer (generalist, specialist or trainee): specialised gynaecologist
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 165, borderline 7, malignant 48, metastatic and others not reported Target condition: OC/EOC (44% EOC)
Flow and timing	

Al Musalhi 2016 *(Continued)*

Comparative
Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			

Al Musalhi 2016 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		High
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		

Al Musalhi 2016 *(Continued)*

DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 3: Reference Standard		
Is the reference standards likely to correctly classify the target condition?	Unclear	

Al Musalhi 2016 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	Unclear		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	Unclear		
Could the conduct of the comparative studies have introduced bias?		Unclear risk	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?			Low concern

Anton 2012

Study characteristics	
Patient Sampling	Country: Brazil Centres: single Study design: within-person comparison Recruitment: prospective cross-sectional study Method of patient selection: convenience

Anton 2012 *(Continued)*

	Inappropriate exclusions (all, stage, all age, included comorbidities such as infertility or endometriosis): none Comments (if applicable): N/A
Patient characteristics and setting	Clinical setting: secondary Study entry criteria: women referred with pelvic masses diagnosed by USS or CT or MRI undergoing surgery or image-guided biopsy when they presented with signs of carcinomatosis Sample size: 120 Age range: not reported Mean age: benign 50.7 years, BOT 56.4 years, malignant 54.7 years Median age: benign 51 years, BOT 58 years, malignant 54 years Percentage postmenopausal (n): 60.8% (73) Comments: 2 participants were excluded as 1 had leiomyoma and 1 mesothelioma instead of ovarian mass on histology
Index tests	Combination RMI, ROMA Prior test: unclear Threshold for test positivity predefined: yes ROMA, yes RMI Threshold for test positivity: ROMA premenopausal ≥ 13.1%, postmenopausal ≥ 22.7%. RMI cut-off 200 Type of ultrasound (TAS, TVS or both): mixed modalities of imaging, parameters identical to the sonographic parameters for RMI were used from the other imaging modalities. Operator experience of sonographer (generalist, specialist or trainee): unclear Type of technology or manufacturer of biomarker test: CA125 (Cobas and Roche), HE4 (EIA)
Target condition and reference standard(s)	Only surgical patients included Follow-up: none Duration of follow-up: N/A Histology: benign 66, borderline 17, malignant 30, metastatic and others 7 Staging: early not reported, late not reported
Flow and timing	
Comparative	ROMA vs RMI
Notes	
Methodological quality	

Anton 2012 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
A) Includes all ages regardless of menopausal status or justify restrictions	Unclear		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		High risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (RMI)			

Anton 2012 *(Continued)*

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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Unclear
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	

Anton 2012 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
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?		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		

Anton 2012 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?	Yes		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	Unclear		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	Unclear		
Could the conduct of the comparative studies have introduced bias?		Unclear risk	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?			Low concern

Bandiera 2011

Study characteristics	
Patient Sampling	Country: USA Centres: single Study design: within-person comparison Recruitment: unclear Method of patient selection: convenience Inappropriate exclusions: BOT excluded; non-EOC excluded Comments (if applicable): N/A
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: not reported Sample size: 278

Bandiera 2011 *(Continued)*

	Age range: 25–89 years Mean age: premenopausal: benign 41.5 years, malignant 44.7 years; postmenopausal: benign 64.0 years, malignant 66.3 years Median age: not reported Percentage postmenopausal (n): 65.8% (183) Comments: pre- and postmenopausal women were balanced in cohorts
Index tests	Combination Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 7.4, postmenopausal 25.3 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: CA125 and HE4 (CMIA)
Target condition and reference standard(s)	Only surgical patients included Follow-up: none Duration of follow-up: N/A Histology: benign 165, borderline excluded, malignant 113, metastatic and others excluded ? Staging: early 33, late 80, unstaged 1
Flow and timing	
Comparative	N/A
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		

Bandiera 2011 *(Continued)*

B) Includes all stages and types of ovarian cancer	No	
C) Includes comorbidities such as infertility and endometriosis	Yes	
Could the selection of patients have introduced bias?		High risk
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated		
B) Prior test in primary care: self-reported symptoms		
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound		
Are there concerns that the included patients and setting do not match the review question?		Unclear
DOMAIN 2: Index Test (ADNEX)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		

Bandiera 2011 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ACOG)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ROMA)	
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?	Yes
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk
A) Was ultrasound performed in all patients by non-specialised sonographers	

Bandiera 2011 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
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?		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Unclear	
Were all patients included in the analysis?	No	

Bandiera 2011 *(Continued)*

Could the patient flow have introduced bias?	High risk
DOMAIN 5: Comparative	
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?	
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	
Could the conduct of the comparative studies have introduced bias?	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	

Chan 2013

Study characteristics	
Patient Sampling	Country: Asia-pacific Centres: multicentre (6; Hong Kong, Japan, Korea, Taiwan, Thailand, Philippines) Study design: within-person comparison Recruitment: prospective cross-sectional study Method of patient selection: consecutive Inappropriate exclusions (all stages, all ages, included comorbidities such as infertility or endometriosis): none Comments (if applicable): N/A
Patient characteristics and setting	Clinical setting: unclear Study entry criteria: women aged > 18 years with adnexal mass diagnosed by any imaging method (USS, CT or MRI) Sample size: 414 Age range: not reported Mean age: not reported Median age: not reported Percentage postmenopausal (n): 26% (108) Comments: N/A

Chan 2013 *(Continued)*

Index tests	Combination vs biomarker Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: ROMA combined 0; premenopausal 7.4, postmenopausal 25.3 Type of ultrasound (TAS, TVS or both): both Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: ARCHITECT
Target condition and reference standard(s)	Only surgical patients included Histology: benign 322, borderline 16, malignant 74, metastatic and others 3 (unclear metastatic/others) Staging: early 23, late 38, unstaged 4
Flow and timing	
Comparative	N/A
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
A) Includes all ages regardless of menopausal status or justify restrictions	Unclear		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			

Chan 2013 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?	Unclear
DOMAIN 2: Index Test (ADNEX)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (RMI)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ACOG)	

Chan 2013 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		

Chan 2013 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
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?		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?		

Chan 2013 *(Continued)*

For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	
Could the conduct of the comparative studies have introduced bias?	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	

Chen 2014

Study characteristics	
Patient Sampling	Country: China Centres: single Study design: non-comparative Recruitment: retrospective cross-sectional study Method of patient selection: convenience Inappropriate exclusions (all stages, all ages, included comorbidities such as infertility or endometriosis): women with non-EOC excluded
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: women with EOC and benign lesions Sample size: 192 Age range: not reported Mean age: not reported Median age: not reported Percentage postmenopausal (n): 43.75% (84)
Index tests	Combination ROMA Prior test: unclear Threshold for test positivity predefined: no Threshold for test positivity: cut-off at 75% specificity; premenopausal 12.2%, postmenopausal 25.8% Type of ultrasound (TAS, TVS, or both): N/A Operator experience of sonographer (generalist, specialist or trainee): Type of technology or manufacturer of biomarker test
Target condition and reference standard(s)	Only surgical patients included

Chen 2014 *(Continued)*

	Histology: benign 69, borderline not reported, malignant 123, metastatic and others not reported Staging: early not reported, late not reported, unstaged not reported
Flow and timing	
Comparative	N/A
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
A) Includes all ages regardless of menopausal status or justify restrictions	Unclear		
B) Includes all stages and types of ovarian cancer	No		
C) Includes comorbidities such as infertility and endometriosis	Unclear		
Could the selection of patients have introduced bias?		High risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			

Chen 2014 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (RMI)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?

Chen 2014 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		High risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		

Chen 2014 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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?			
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?			
Could the conduct of the comparative studies have introduced bias?			
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?			

Chen 2015

Study characteristics	
Patient Sampling	Country: China Centres: single Study design: within-person comparison Recruitment: unclear Method of patient selection: convenience Inappropriate exclusions (all stages, all ages, included comorbidities such as infertility or endometriosis): unclear
Patient characteristics and setting	Clinical setting: unclear Study entry criteria: women with pelvic masses scheduled for surgery Sample size: 232 Age range: 17–81 years Mean age: benign 33 years, malignant 53 years Median age: not reported Percentage postmenopausal (n): not reported
Index tests	Combination ROMA Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 11.4, postmenopausal 29.9 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: ECLIA
Target condition and reference standard(s)	Only surgical patients included Histology: benign 70, borderline not reported, malignant 60, metastatic and others not reported Staging: early not reported, late not reported, unstaged not reported
Flow and timing	
Comparative	N/A
Notes	
Methodological quality	

Chen 2015 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Unclear		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (RMI)			

Chen 2015 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ACOG)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ROMA)	
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

Chen 2015 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
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	
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		

Chen 2015 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?		
Could the conduct of the comparative studies have introduced bias?		
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		

Chudecka-Glaz 2015

Study characteristics	
Patient Sampling	Country: Poland Centres: single Study design: within-person comparison Recruitment: prospective Method of patient selection: consecutive Inappropriate exclusions: none reported
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: women presenting with ovarian tumour, ovarian cyst or ascites (suspected OC) Sample size: 413

Chudecka-Glaz 2015 *(Continued)*

	Age range: OC 24–90 years; benign 18–88 years Mean age: not reported Median age: OC 59.7 years; benign 35 years Percentage postmenopausal (n): 61% (251)
Index tests	Test: ROMA and ROMA-P Prior test: not reported Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 14.1, postmenopausal 25 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: HE4: the Roche Elecsys assay on a Cobas e601 apparatus; CA125: ARCHITECT CA125 II assay on an ARCHITECT
Target condition and reference standard(s)	Only surgical patients included Histology: benign (n) 251, borderline (n) not reported, malignant (n) 162, metastatic and others not reported
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			

Chudecka-Glaz 2015 *(Continued)*

B) Prior test in primary care: self-reported symptoms	
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound	
Are there concerns that the included patients and setting do not match the review question?	Unclear
DOMAIN 2: Index Test (ADNEX)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (RMI)	
DOMAIN 2: Index Test (ACOG)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	

Chudecka-Glaz 2015 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		

Chudecka-Glaz 2015 *(Continued)*

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		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?	Yes		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	Yes		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	Unclear		
Could the conduct of the comparative studies have introduced bias?		Unclear risk	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?			Low concern

Cradic 2018

Study characteristics	
Patient Sampling	Country: USA Centres: single

Cradic 2018 *(Continued)*

	Study design: non-comparative Recruitment: retrospective Method of patient selection: unclear Inappropriate exclusions: not reported; age group not stated
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: women with EOC or benign ovarian lesions Sample size: 207 Age range: not reported Mean age: not reported Percentage postmenopausal (n): 45% (93)
Index tests	Test: ROMA Prior test: not reported Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 11.4, postmenopausal 29.9 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: not reported
Target condition and reference standard(s)	Only surgical patients included Histology: benign (n) 131, borderline (n) not reported, malignant (n) 76, metastatic and others none reported Target condition: EOC
Flow and timing	
Comparative	
Notes	

Methodological quality

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

Cradic 2018 *(Continued)*

A) Includes all ages regardless of menopausal status or justify restrictions	Unclear	
B) Includes all stages and types of ovarian cancer	Unclear	
C) Includes comorbidities such as infertility and endometriosis	Unclear	
Could the selection of patients have introduced bias?		Unclear risk
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated		
B) Prior test in primary care: self-reported symptoms		
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound		
Are there concerns that the included patients and setting do not match the review question?		Unclear
DOMAIN 2: Index Test (ADNEX)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		

Cradic 2018 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ACOG)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ROMA)	
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
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes

Cradic 2018 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
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	
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Yes	

Cradic 2018 *(Continued)*

Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Low risk
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?		
Could the conduct of the comparative studies have introduced bias?		
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		

Dikmen 2015

Study characteristics	
Patient Sampling	Country: Turkey Centres: unclear Study design: non-comparative Recruitment: unclear Method of patient selection: unclear Inappropriate exclusions: not reported
Patient characteristics and setting	Clinical setting: unclear Study entry criteria: women were 'preoperative' Sample size: 143 Age range: not reported Mean age: benign 42 (SD 10) years, malignant 56 (SD 14) years Percentage postmenopausal (n): 32% (46)
Index tests	Test: ROMA Prior test: unclear

Dikmen 2015 *(Continued)*

	Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 13.1, postmenopausal 27.7 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: not reported; stated, "CA125 and HE4 analysed in parallel using a specific system"
Target condition and reference standard(s)	Only surgical patients included Histology (n): 100%; benign 96, borderline not reported, malignant 47, metastatic and others not reported Follow-up: none
Flow and timing	
Comparative	N/A
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
A) Includes all ages regardless of menopausal status or justify restrictions	Unclear		
B) Includes all stages and types of ovarian cancer	Unclear		
C) Includes comorbidities such as infertility and endometriosis	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear

Dikmen 2015 *(Continued)*

DOMAIN 2: Index Test (ADNEX)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (RMI)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?

Dikmen 2015 *(Continued)*

If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		

Dikmen 2015 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?			

Dikmen 2015 *(Continued)*

Could the conduct of the comparative studies have introduced bias?	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	

Ertas 2016

Study characteristics	
Patient Sampling	Country: Turkey Centres: single Study design: non-comparative Recruitment: retrospective Method of patient selection: unclear Inappropriate exclusions: none reported
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: women with adnexal masses that underwent surgery and with complete data available Sample size: 408 Age range: 14–87 years Mean age: OC 54.4 (SD 13.6) years; benign 40.8 (SD 13.8) years Percentage postmenopausal (n): 71.4% (117)
Index tests	Test: RMI I Prior test: not reported Threshold for test positivity predefined: yes Threshold for test positivity: 200 Type of ultrasound (TAS, TVS or both): both Operator experience of sonographer (generalist, specialist or trainee): specialist (expert radiologist) Type of technology or manufacturer of biomarker test: CA125: Architect Abbott i2000sr CMIA)): ultrasound: TVS and TAS using a Mindray DC7 ultrasound device with 5 Mhz convex abdominal and 8 Mhz vaginal probes.
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 341, borderline 12, malignant 55, metastatic and others not reported
Flow and timing	

Ertas 2016 *(Continued)*

Comparative	N/A		
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Unclear		
B) Includes all stages and types of ovarian cancer	Unclear		
C) Includes comorbidities such as infertility and endometriosis	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			

Ertas 2016 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
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?	Yes	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		High
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		

Ertas 2016 *(Continued)*

DOMAIN 2: Index Test (ROMA)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (LR2)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 3: Reference Standard	
Is the reference standards likely to correctly classify the target condition?	Yes

Ertas 2016 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?			
Could the conduct of the comparative studies have introduced bias?			
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?			

Farzaneh 2014

Study characteristics	
Patient Sampling	Country: Iran Centres: single Study design: within-person comparison Recruitment: prospective cross-sectional study Method of patient selection: convenience

Farzaneh 2014 *(Continued)*

	Inappropriate exclusions (all stages, all ages, included comorbidities such as infertility or endometriosis): excluded non-EOC
Patient characteristics and setting	Clinical setting: secondary Study entry criteria: women with adnexal mass undergoing surgery and having attained menarche 12 months before presenting with adnexal mass Sample size: 99 Age range: 17–79 years Mean age: benign 39 years, EOC 51 years Median age: not reported Percentage postmenopausal (n): 31.3% (31)
Index tests	Combination ROMA Prior test: unclear Threshold for test positivity predefined: no Threshold for test positivity: best cut-off as determined by Youdon index all 18.3, premenopausal 11.5, postmenopausal 25.5 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: CA125 (Abbott), HE4 (EIA) Comments: blood samples were collected 30 minutes before the operation
Target condition and reference standard(s)	Only surgical patients included Histology: benign 56, borderline not reported, malignant 43, metastatic and others not reported Staging: early 12, late 31, unstaged 0
Flow and timing	
Comparative	N/A
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		

Farzaneh 2014 *(Continued)*

Did the study avoid inappropriate exclusions?	No	
A) Includes all ages regardless of menopausal status or justify restrictions	Yes	
B) Includes all stages and types of ovarian cancer	No	
C) Includes comorbidities such as infertility and endometriosis	Yes	
Could the selection of patients have introduced bias?		High risk
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated		
B) Prior test in primary care: self-reported symptoms		
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound		
Are there concerns that the included patients and setting do not match the review question?		Unclear
DOMAIN 2: Index Test (ADNEX)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		

Farzaneh 2014 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ACOG)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ROMA)	
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?	No
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes

Farzaneh 2014 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		High risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
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?		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Yes	

Farzaneh 2014 *(Continued)*

Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	No	
Could the patient flow have introduced bias?		High risk
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?		
Could the conduct of the comparative studies have introduced bias?		
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		

Grenache 2015

Study characteristics	
Patient Sampling	Country: USA Centres: multicentre Study design: within-person comparison Recruitment: retrospective cross-sectional study Method of patient selection: convenience Inappropriate exclusions (all stages, all ages, included comorbidities such as infertility or endometriosis): none Comments (if applicable): N/A
Patient characteristics and setting	Clinical setting: unclear Study entry criteria: women with abnormal adnexal mass detected on physical examination and imaging of ultrasound, CT or MRI) followed by surgery Sample size: 146 Age range: 18–89 years Mean age: 52 years Percentage postmenopausal (n): 52% (76)

Grenache 2015 *(Continued)*

	Comments: benign samples (90) were randomly collected from cohort of ICRA diagnosis of benign disease and all 6 malignant samples from the same cohort were included. Samples (50) were randomly collected from cohort of ICRA diagnosis of malignancy (25 from the confirmed benign group and 25 from the confirmed malignant group). The sampling tried to mimic prevalence of malignancy in women undergoing surgery (21%)
Index tests	Combination ROMA Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: ROMA premenopausal ≥ 1.31, postmenopausal ≥ 2.77 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: MVI-Quest, HE4 and CA125 (Abbot) Comments: laboratory personnel were blinded to all clinical information. All blood samples were collected < 30 days prior to surgery except 1 (50 days)
Target condition and reference standard(s)	Only surgical patients included Histology: benign 115, borderline 7, malignant 19, metastatic and others 5 (3 mets) Staging: early 18, late 14, unstaged 4
Flow and timing	
Comparative	N/A
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		

Grenache 2015 *(Continued)*

C) Includes comorbidities such as infertility and endometriosis	Yes
Could the selection of patients have introduced bias?	High risk
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated	
B) Prior test in primary care: self-reported symptoms	
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound	
Are there concerns that the included patients and setting do not match the review question?	Unclear
DOMAIN 2: Index Test (ADNEX)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (RMI)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	

Grenache 2015 *(Continued)*

A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		

Grenache 2015 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (LR2)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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?			
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
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		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Grenache 2015 *(Continued)*

For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?	
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	
Could the conduct of the comparative studies have introduced bias?	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	

Huy 2018

Study characteristics	
Patient Sampling	Country: Vietnam Centres: single Study design: non-comparative Recruitment: retrospective Method of patient selection: unclear Inappropriate exclusions: unclear about borderline cases
Patient characteristics and setting	Clinical setting: mixed Study entry criteria: women with sufficient personal information, clinical symptoms, data on serum CA125 and serum HE4 levels, and postoperative pathological findings Sample size: 277 Age range: not reported Mean age: not reported Percentage postmenopausal (n): 17% (47)
Index tests	Test: ROMA Prior test: not reported Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 7.4, postmenopausal 25.3 Type of ultrasound (TAS, TVS or both): N/A

Huy 2018 *(Continued)*

	Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: CA125 and HE4 measured using Elecsys 2010 system immunoassay (Elecsys, 2010) and ARCHITECT i1000SR system, respectively (ARCHITECT System User Manual, 2009).
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 247, borderline not reported, malignant 30, metastatic and others none Target condition: EOC
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			

Huy 2018 *(Continued)*

If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (RMI)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?

Huy 2018 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		

Huy 2018 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?	Yes		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	Yes		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	Unclear		
Could the conduct of the comparative studies have introduced bias?		Unclear risk	

Huy 2018 *(Continued)*

Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	Low concern

Irshad 2013

Study characteristics	
Patient Sampling	Country: Pakistan Centres: single Study design: non-comparative Recruitment: unclear Method of patient selection: convenience Inappropriate exclusions (all stages, all ages, included comorbidities such as infertility or endometriosis): unclear (? excludes premenopausal women)
Patient characteristics and setting	Clinical setting: secondary Study entry criteria: unclear Sample size: 36 Age range: 50–70 years Mean age: 58 (SD 5.88) years Percentage postmenopausal (n): not reported Comments: inclusion criteria not reported. Women with postmenopausal bleeding and family history of breast cancer and OC were excluded.
Index tests	Combination RMI I Prior test: unclear Threshold for test positivity predefined: no Threshold for test positivity: > 250 Type of ultrasound (TAS, TVS or both): unclear Operator experience of sonographer (generalist, specialist or trainee): unclear Type of technology or manufacturer of biomarker test: not reported
Target condition and reference standard(s)	Only surgical patients included Histology: benign 12, borderline not reported, malignant 24, metastatic and others not reported Staging: early not reported, late not reported, unstaged not reported

Irshad 2013 *(Continued)*

Flow and timing			
Comparative	N/A		
Notes			

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
A) Includes all ages regardless of menopausal status or justify restrictions	Unclear		
B) Includes all stages and types of ovarian cancer	Unclear		
C) Includes comorbidities such as infertility and endometriosis	Unclear		
Could the selection of patients have introduced bias?		High risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			

Irshad 2013 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		High risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		High
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		

Irshad 2013 *(Continued)*

DOMAIN 2: Index Test (ROMA)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (LR2)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 3: Reference Standard	
Is the reference standards likely to correctly classify the target condition?	Yes

Irshad 2013 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?		
Could the conduct of the comparative studies have introduced bias?		
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		

Kadija 2012

Study characteristics	
Patient Sampling	Country: Serbia Centres: single Study design: within-person comparison Recruitment: prospective cross-sectional study Method of patient selection: convenience

Kadija 2012 *(Continued)*

	Inappropriate exclusions (all stages, all ages, included comorbidities such as infertility or endometriosis): unclear
Patient characteristics and setting	Clinical setting: secondary Study entry criteria: women diagnosed with adnexal mass scheduled to undergo surgery Sample size: 108 Age range: not reported Mean age: not reported Median age: not reported Percentage postmenopausal (n): 40% (41) Comments: metastasis to ovaries from 4 malignancies excluded
Index tests	Combination ROMA Prior test: unclear Threshold for test positivity predefined: no Threshold for test positivity: premenopausal < 12.5%, postmenopausal < 14.4% Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: CA125 – Immulite 2000 (Siemens) HE4 (Fujirebio) Comments: pathologists and surgeons were blinded to the index test results.
Target condition and reference standard(s)	Only surgical patients included Histology: benign 79, borderline 5, malignant 24, metastatic and others 4 (excluded) Staging: early 9 (only invasive), late 15 (only invasive), unstaged not reported
Flow and timing	
Comparative	N/A
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		

Kadija 2012 *(Continued)*

Was a case-control design avoided?	Yes	
Did the study avoid inappropriate exclusions?	Unclear	
A) Includes all ages regardless of menopausal status or justify restrictions	Unclear	
B) Includes all stages and types of ovarian cancer	Unclear	
C) Includes comorbidities such as infertility and endometriosis	Yes	
Could the selection of patients have introduced bias?		High risk
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated		
B) Prior test in primary care: self-reported symptoms		
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound		
Are there concerns that the included patients and setting do not match the review question?		High
DOMAIN 2: Index Test (ADNEX)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		

Kadija 2012 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ACOG)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ROMA)	
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?	No
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes

Kadija 2012 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		High risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
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?		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear

Kadija 2012 *(Continued)*

DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?		
Could the conduct of the comparative studies have introduced bias?		
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		

Karlsen 2012

Study characteristics	
Patient Sampling	Country: Denmark Centres: single Study design: within-person comparison Recruitment: prospective cross-sectional study Method of patient selection: convenience Inappropriate exclusions (all stages, all ages, included comorbidities such as infertility or endometriosis): none Comments (if applicable): women examined as per fast track guidelines
Patient characteristics and setting	Clinical setting: secondary Study entry criteria: women admitted to surgery for pelvic mass or pelvic pain potentially caused by malignant disease or endometriosis

Karlsen 2012 *(Continued)*

	Sample size: 1218 Age range: 16–90 years Mean age: not reported Median age: 51 years Percentage postmenopausal (n): 51% (621) Comments: 69 non-OCs? metastatic
Index tests	Combination ROMA Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 13.1, postmenopausal 27.7 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: CMIA Comments: blood samples collected 2 weeks prior to surgery
Target condition and reference standard(s)	Only surgical patients included Histology: benign 809, borderline 79, malignant 261, metastatic and others 69 Staging: early 64 (only for EOC), late 188 (only for EOC), unstaged 0
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		

Karlsen 2012 *(Continued)*

C) Includes comorbidities such as infertility and endometriosis	Yes
Could the selection of patients have introduced bias?	Unclear risk
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated	
B) Prior test in primary care: self-reported symptoms	
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound	
Are there concerns that the included patients and setting do not match the review question?	Unclear
DOMAIN 2: Index Test (ADNEX)	
DOMAIN 2: Index Test (RMI)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ACOG)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	

Karlsen 2012 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		

Karlsen 2012 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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?			
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?			
Could the conduct of the comparative studies have introduced bias?			
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?			

Kim 2011

Study characteristics	
Patient Sampling	Country: South Korea Centres: single Study design: within-person comparison Recruitment: unclear Method of patient selection: convenience Inappropriate exclusions (all stages, all ages, included comorbidities such as infertility or endometriosis): only EOC included Comments (if applicable): none
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: women diagnosed with adnexal mass on the first visit to the gynaecological oncology clinic and underwent surgery Sample size: 159 Age range: 14–73 years Mean age: benign 35.7 (SD 11.8) years, OC 51.7 (SD 11.7) years Median age: not reported Percentage postmenopausal (n): 68% (108) Comments: none
Index tests	Combination ROMA Prior test: unclear Threshold for test positivity predefined: no Threshold for test positivity: premenopausal 7.6%, postmenopausal 10.9% Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: CA125 and HE4 both automated immunochemiluminescence assay
Target condition and reference standard(s)	Only surgical patients included Follow-up: none Duration of follow-up: N/A Histology: benign 81, borderline 10, malignant 68, metastatic and others 2 Staging: early 29, late 49
Flow and timing	

Kim 2011 *(Continued)*

Comparative	N/A
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	No		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		High risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			

Kim 2011 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (RMI)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?

Kim 2011 *(Continued)*

DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		High risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 3: Reference Standard		
Is the reference standards likely to correctly classify the target condition?	Yes	

Kim 2011 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	No	
Could the patient flow have introduced bias?		High risk
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?		
Could the conduct of the comparative studies have introduced bias?		
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		

Kim 2019

Study characteristics	
Patient Sampling	Country: Korea Centres: single Study design: non-comparative Recruitment: retrospective Method of patient selection: unclear

Kim 2019 *(Continued)*

	Inappropriate exclusions: unclear (presume BOT excluded as retrospective)
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: women with suspected gynaecological disease Sample size: 832 Age range: not reported Mean age: not reported Median age: benign 45.0 (IQR 36.0–51.0) years; OC: 64.0 (IQR 50.9–77.0) years Percentage postmenopausal (n): 30% (251)
Index tests	Test: ROMA Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 11.4, postmenopausal 29.9 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: CA125 and HE4 tests performed with a Cobas E 602 immunoassay analyser using Elecsys CA125 II and Elecsys HE4 test reagents (Roche Diagnostics GmbH, Mannheim, Germany)
Target condition and reference standard(s)	Histology: 563 (68%) Follow-up: not reported Histology (n): benign 762, borderline not reported, malignant 70, metastatic 3, others 3 stromal tumour, 3 germ cell tumour
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		

Kim 2019 *(Continued)*

Did the study avoid inappropriate exclusions?	Unclear	
A) Includes all ages regardless of menopausal status or justify restrictions		
B) Includes all stages and types of ovarian cancer	Unclear	
C) Includes comorbidities such as infertility and endometriosis	Unclear	
Could the selection of patients have introduced bias?		Unclear risk
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated		
B) Prior test in primary care: self-reported symptoms		
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound		
Are there concerns that the included patients and setting do not match the review question?		Unclear
DOMAIN 2: Index Test (ADNEX)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		

Kim 2019 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ACOG)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ROMA)	
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
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes

Kim 2019 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 3: Reference Standard		
Is the reference standards likely to correctly classify the target condition?	Unclear	
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear	
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	

Kim 2019 *(Continued)*

Did all patients receive the same reference standard?	Unclear
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	Unclear risk
DOMAIN 5: Comparative	
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?	
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	
Could the conduct of the comparative studies have introduced bias?	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	

Krascsenitis 2016

Study characteristics	
Patient Sampling	Country: Hungary Centres: single Study design: within-person comparison Recruitment: prospective Method of patient selection: unclear Inappropriate exclusions: not reported
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: women diagnosed with an ovarian tumour of unknown significance admitted for surgery. Sample size: 162 Age range: not reported Mean age: 55 years Percentage postmenopausal (n): 63% (102)
Index tests	Test: ROMA and RMI I Prior test: not reported

Krascsenitis 2016 *(Continued)*

	Threshold for test positivity predefined: yes Threshold for test positivity: RMI I 200; ROMA premenopausal 11.4, postmenopausal 29.9 Type of ultrasound (TAS, TVS or both): not reported Operator experience of sonographer (generalist, specialist or trainee): not reported Type of technology or manufacturer of biomarker test: not reported
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 101, borderline 11, malignant 34, metastatic and others 16
Flow and timing	
Comparative	RMI I vs ROMA
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			

Krascsenitis 2016 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
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?	Yes	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Unclear
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		

Krascsenitis 2016 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		

Krascsenitis 2016 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?	Yes		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	Unclear		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	Unclear		

Krascsenitis 2016 *(Continued)*

Could the conduct of the comparative studies have introduced bias?	Unclear risk
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	Unclear

Li 2016

Study characteristics	
Patient Sampling	Country: China Centres: single Study design: non-comparative Recruitment: retrospective Method of patient selection: unclear Inappropriate exclusions: none reported
Patient characteristics and setting	Clinical setting: unclear Study entry criteria: women diagnosed with gynaecological diseases. Histological diagnosis verified by 2 different pathologists Sample size: 916 Age range: 18–82 years Mean age: not reported Median age: 50 years Percentage postmenopausal (n): 19% (172)
Index tests	Test: ROMA Prior test: ultrasound, CT scan, PET-CT scan or MRI histological diagnosis verified by 2 different pathologists Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 7.4, postmenopausal 25.3 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: tested by the ARCHITECT CA125 II assay and ARCHITECT HE4 assay (Abbott Diagnostics, Abbott Park, IL)
Target condition and reference standard(s)	Only surgical patients included

Li 2016 *(Continued)*

	Histology (n): benign 726, borderline not reported, malignant 190, metastatic and others 0		
Flow and timing			
Comparative			
Notes			

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			

Li 2016 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (RMI)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers

Li 2016 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		

Li 2016 *(Continued)*

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		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
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		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?			
Could the conduct of the comparative studies have introduced bias?			
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?			

Liest 2019

Study characteristics	
Patient Sampling	Country: Sweden Centres: multicentre

Liest 2019 *(Continued)*

	Study design: within-person comparison Recruitment: prospective Method of patient selection: convenience (enrolled by gynaecologists) Inappropriate exclusions: none reported but age group not specified
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: women aged ≥ 18 years with a pelvic mass of probable ovarian origin and scheduled for surgery Sample size: 784 Age range: not reported Mean age: not reported Percentage postmenopausal (n): 81% (117)
Index tests	Test: ROMA and RMI Prior test: USS Threshold for test positivity predefined: yes Threshold for test positivity: ROMA: premenopausal 11, postmenopausal 25; RMI I 200 Type of ultrasound (TAS, TVS or both): unclear Operator experience of sonographer (generalist, specialist or trainee): unclear Type of technology or manufacturer of biomarker test: both CA125 and HE4 measured by an electrochemiluminescence immunoassay on the automated cobas e602 module (Roche Diagnostics, Mannheim, Germany)
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 611, borderline not reported, malignant 144 (including borderline), metastatic and others 29 Target condition: EOC
Flow and timing	
Comparative	ROMA vs RMI
Notes	

Methodological quality

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

Liest 2019 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Unclear	
Was a case-control design avoided?	Yes	
Did the study avoid inappropriate exclusions?	Unclear	
A) Includes all ages regardless of menopausal status or justify restrictions	Unclear	
B) Includes all stages and types of ovarian cancer	Unclear	
C) Includes comorbidities such as infertility and endometriosis	Unclear	
Could the selection of patients have introduced bias?		Unclear risk
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated		
B) Prior test in primary care: self-reported symptoms		
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound		
Are there concerns that the included patients and setting do not match the review question?		Unclear
DOMAIN 2: Index Test (ADNEX)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
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	

Liest 2019 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Unclear
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	

Liest 2019 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
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	
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern

Liest 2019 *(Continued)*

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	
Could the patient flow have introduced bias?		Low risk
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	Yes	
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	Unclear	
Could the conduct of the comparative studies have introduced bias?		Unclear risk
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		Low concern

Lycke 2018

Study characteristics	
Patient Sampling	Country: Sweden Centres: multicentre Study design: within-person comparison Recruitment: prospective Method of patient selection: consecutive Inappropriate exclusions: none reported
Patient characteristics and setting	Clinical setting: mixed Study entry criteria: women aged > 18 years planned for a surgical procedure for a symptomatic or suspected malignant ovarian cyst or pelvic tumour Sample size: 638 Age range: not reported

Lycke 2018 *(Continued)*

	Mean age: benign 50.76 years, BOT 55.58 years, EOC 62.67 years Percentage postmenopausal (n): 55% (348)
Index tests	Test: ROMA and RMI I Prior test: unclear but assume history, examination and ultrasound Threshold for test positivity predefined: yes Threshold for test positivity: yes ROMA: premenopausal 11.4, postmenopausal 29.9 RMI: 200 Type of ultrasound (TAS, TVS or both): unclear Operator experience of sonographer (generalist, specialist or trainee): gynaecology specialist or trainee Type of technology or manufacturer of biomarker test: Elecsys HE4 and Elecsys CA125 II with the electrochemilluminescence (ECLIA) technique (Cobas 8000, Roche Diagnostics Scandinavia, Stockholm, Sweden)
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 445, borderline 31, malignant 162, metastatic and others 0 Follow-up: none
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Low risk	

Lycke 2018 *(Continued)*

A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated	
B) Prior test in primary care: self-reported symptoms	
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound	
Are there concerns that the included patients and setting do not match the review question?	Unclear
DOMAIN 2: Index Test (ADNEX)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (RMI)	
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
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes
Could the conduct or interpretation of the index test have introduced bias?	Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers	

Lycke 2018 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		High
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern

Lycke 2018 *(Continued)*

DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
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	
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
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	
Could the patient flow have introduced bias?		Low risk
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for par-	Yes	

Lycke 2018 *(Continued)*

ticipants receiving one or other index test or testing strategy the same?		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	Yes	
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	Yes	
Could the conduct of the comparative studies have introduced bias?	Low risk	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		Low concern

Manegold-Brauer 2016

Study characteristics	
Patient Sampling	Country: Switzerland Centres: single Study design: non-comparative Recruitment: retrospective Method of patient selection: convenience Inappropriate exclusions: none
Patient characteristics and setting	Clinical setting: secondary Study entry criteria: women who had an USS examination for an adnexal mass in a general gynaecological outpatient setting with histology and CA125 results available Sample size: 1108 Age range: not reported Mean age: not reported Median age: 48 years Percentage postmenopausal (n): 43% (478)
Index tests	Test: RMI I Prior test: not reported Threshold for test positivity predefined: yes Threshold for test positivity: 200 Type of ultrasound (TAS, TVS or both): not reported Operator experience of sonographer (generalist, specialist or trainee): trainee

Manegold-Brauer 2016 *(Continued)*

	Type of technology or manufacturer of biomarker test: USS performed with high-resolution machines (GE Voluson 730 Expert, GE Voulson E8, Phillips HDI 5000, Phillips IU22).
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 936, borderline 33, malignant 118, metastatic and others 17
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		High risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			

Manegold-Brauer 2016 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Unclear
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		

Manegold-Brauer 2016 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ROMA)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (LR2)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers

Manegold-Brauer 2016 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?			
Could the conduct of the comparative studies have introduced bias?			
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?			

Melo 2018

Study characteristics	
Patient Sampling	Country: Portugal Centres: single Study design: within-person comparison Recruitment: retrospective Method of patient selection: unclear Inappropriate exclusions: not reported; age group not specified
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: women with adnexal neoplasia submitted to surgical treatment, with a histological diagnosis and in which RO-MA had been determined Sample size: 247 Age range: not reported Mean age: not reported Percentage postmenopausal (n): 37% (92)
Index tests	Test: ROMA Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 7.4, postmenopausal 25.3 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: CA125 and HE4 were measured on the ARCHITECT i2000SRrVR, a fully automated immunoassay analyser (Abbott Laboratories, Abbott Park, IL)
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 206, borderline 7, malignant 34, metastatic and others none reported
Flow and timing	
Comparative	
Notes	
Methodological quality	

Melo 2018 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
A) Includes all ages regardless of menopausal status or justify restrictions	Unclear		
B) Includes all stages and types of ovarian cancer	Unclear		
C) Includes comorbidities such as infertility and endometriosis	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (RMI)			

Melo 2018 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ACOG)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ROMA)	
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

Melo 2018 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
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	
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		

Melo 2018 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?		
Could the conduct of the comparative studies have introduced bias?		
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		

Meys 2017

Study characteristics	
Patient Sampling	Country: the Netherlands Centres: single Study design: within-person comparison Recruitment: prospective Method of patient selection: consecutive Inappropriate exclusions: none
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: women with adnexal pathology Sample size: 326 Age range: not reported

Meys 2017 *(Continued)*

	Mean age: not reported Median age: benign 53.2 (IQR 16.1–87.2) years, malignant 67.7 (IQR 32.3–87) years Percentage postmenopausal (n): 61% (198)
Index tests	Test: ADNEX, LR2 and RMI I Prior test: not reported Threshold for test positivity predefined: Threshold for test positivity: ADNEX 10%, LR2 10%, RMI I 200 Type of ultrasound (TAS, TVS or both): both Operator experience of sonographer (generalist, specialist or trainee): experienced gynaecologist Type of technology or manufacturer of biomarker test: transvaginal or transrectal grey-scale and colour Doppler ultrasound examination, using a Voluson E8 (GE Healthcare Ultrasound, Milwaukee, WI, USA) ultrasound machine along with TAS for large mass or suspected malignancy was performed.
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 211, borderline 27, malignant 115, metastatic and others 14 Target condition: OC/EOC (84% EOC)
Flow and timing	
Comparative	ADNEX vs RMI I vs LR2
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Low risk	

Meys 2017 *(Continued)*

A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
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		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 2: Index Test (RMI)			
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		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
A) Was ultrasound performed in all patients by non-specialised sonographers			

Meys 2017 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High
DOMAIN 2: Index Test (ACOG)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ROMA)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	

Meys 2017 *(Continued)*

DOMAIN 2: Index Test (LR2)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		High
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	
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		High
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	
Could the patient flow have introduced bias?		Low risk
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for par-		

Meys 2017 *(Continued)*

ticipants receiving one or other index test or testing strategy the same?		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	Yes	
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	Unclear	
Could the conduct of the comparative studies have introduced bias?	Unclear risk	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		Low concern

Molina 2011

Study characteristics	
Patient Sampling	Country: Spain Centres: single Study design: within-person comparison Recruitment: retrospective cross-sectional study Method of patient selection: convenience Inappropriate exclusions (all stages, all ages, included comorbidities such as infertility or endometriosis): unclear
Patient characteristics and setting	Clinical setting: unclear Study entry criteria: not reported Sample size: 396 Age range: 17–90 years Mean age: not reported Median age: benign gynaecological disease 40 (SD 0.8) years; gynaecological cancer 61 (SD 1.2) years Percentage postmenopausal (n): 34% (143) Comment: patient spectrum included OC, benign gynaecological disease (ovarian cyst, myomas, endometriosis, endometrial polyps)
Index tests	Combination ROMA Prior test: unclear Threshold for test positivity predefined: yes

Molina 2011 *(Continued)*

	Threshold for test positivity: ROMA: premenopausal ≥ 13.1, post-menopausal ≥ 27.7 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: CMIA
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 285 *benign gynaecological disease with 137 ovarian cysts, borderline not reported, malignant 111, metastatic and others 11 others (? Mets) Staging: early 19, late 92, unstaged 0
Flow and timing	
Comparative	N/A
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Unclear		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			

Molina 2011 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (RMI)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?

Molina 2011 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		

Molina 2011 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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?			
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?			

Molina 2011 *(Continued)*

Could the conduct of the comparative studies have introduced bias?	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	

Montagnana 2011

Study characteristics	
Patient Sampling	Country: Italy Centres: single Study design: within-person comparison Recruitment: unclear Method of patient selection: convenience Inappropriate exclusions: non-EOC excluded Comments (if applicable): N/A
Patient characteristics and setting	Clinical setting: secondary Study entry criteria: women with pelvic mass scheduled to have radical surgery Sample size: 104 Age range: not reported Mean age: EOC 56.9 (SD 14.4) years, benign 42 (SD 15.5) years Median age: not reported Percentage postmenopausal (n): 51% (53) Comments: only women undergoing radical surgery were included
Index tests	Combination ROMA Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal ≥ 12.5, postmenopausal ≥ 14.4 Interval between index test and reference standard: 1 day Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: CA125 (ECLIA), HE4 (RIA)

Montagnana 2011 *(Continued)*

Target condition and reference standard(s)	Only surgical patients included Follow-up: none Duration of follow-up: N/A Histology (n): benign 49, borderline – ? excluded, malignant 55, metastatic and others? excluded Staging: early 15, late 40, unstaged 0
Flow and timing	
Comparative	N/A
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
A) Includes all ages regardless of menopausal status or justify restrictions	Unclear		
B) Includes all stages and types of ovarian cancer	Unclear		
C) Includes comorbidities such as infertility and endometriosis	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			

Montagnana 2011 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (RMI)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?

Montagnana 2011 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		

Montagnana 2011 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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?			
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
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?	Unclear		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?			
Could the conduct of the comparative studies have introduced bias?			

Montagnana 2011 *(Continued)*

Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	

Moore 2009

Study characteristics	
Patient Sampling	Country: USA Centres: multicentre Study design: non-comparative Recruitment: prospective cross-sectional study Method of patient selection: convenience Inappropriate exclusions (all stages, all ages, included comorbidities such as infertility or endometriosis): none Comments (if applicable): N/A
Patient characteristics and setting	Clinical setting: unclear Study entry criteria: women with ovarian cyst scheduled to undergo surgery Sample size: 513 Age range: 18–87 years Mean age: 54 years Median age: not reported Percentage postmenopausal (n): 29% (150) Comments: 12 centres; aged < 48 years premenopausal, aged > 55 years postmenopausal; FSH values used to categorise women into premenopausal and postmenopausal if last menstrual period was unknown
Index tests	Combination ROMA Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: specificity of 75%, premenopausal ≥ 13.1%, postmenopausal ≥ 27.7 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: CA125 (Abbott), HE4 (EIA) Comments: laboratory testing was blinded to histology

Moore 2009 *(Continued)*

Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 352, borderline 22, malignant 143, metastatic and others 14 Staging: early 93 (3 BOT) (only EOC and BOT); late 93 (3 BOT) (only EOC and BOT); unstaged 14 (10 BOT) Comments: histological evaluations were blinded to laboratory testing
Flow and timing	
Comparative	N/A
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			

Moore 2009 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (RMI)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?

Moore 2009 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		

Moore 2009 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
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?		
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?	Unclear risk	
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?		
Could the conduct of the comparative studies have introduced bias?		

Moore 2009 *(Continued)*

Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	

Moore 2011

Study characteristics	
Patient Sampling	Country: USA Centres: multicentre Study design: non-comparative Recruitment: prospective cross-sectional study Method of patient selection: convenience Inappropriate exclusions (all stages, all ages, included comorbidities such as infertility or endometriosis): none
Patient characteristics and setting	Clinical setting: mixed Study entry criteria: women with ovarian cyst scheduled to undergo surgery Sample size: 472 Age range: 18–89 years Mean age: 50.3 years Median age: not reported Percentage postmenopausal (n): 46% (217) Comments: 13 centres, 7 general, 6 speciality; aged < 48 years premenopausal, aged > 55 years postmenopausal, aged 48–55 years FSH values used to categorise women into premenopausal and postmenopausal with unknown last menstrual period
Index tests	Combination ROMA Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: specificity of 75%, premenopausal ≥ 13.1%, postmenopausal ≥ 27.7 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: CA125 (Abbott), HE4 (EIA) Comments: blood sample collected < 30 days prior to surgery
Target condition and reference standard(s)	Only surgical patients included

Moore 2011 *(Continued)*

	Histology (n): benign 383, borderline 19, malignant 68, metastatic and others 2 Staging: early 12 (only for EOC), late 34 (only for EOC), unstaged not reported
Flow and timing	
Comparative	N/A
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			

Moore 2011 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (RMI)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?

Moore 2011 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		

Moore 2011 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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?			
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
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		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?			
Could the conduct of the comparative studies have introduced bias?			
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?			

Niemi 2017

Study characteristics	
Patient Sampling	Country: Finland Centres: single Study design: within-person comparison Recruitment: retrospective Method of patient selection: unclear Inappropriate exclusions: overtly benign or malignant-looking tumours like unilocular simple ovarian cysts and tumours associated with marked ascites (depth of the greatest pool over 10 cm) were excluded
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: women aged > 50 years presenting with an abnormal adnexal mass(es) Sample size: 98 Age range: 50–84 years Mean age: not reported Median age: 61 years Percentage postmenopausal (n): 100%
Index tests	Test: RMI I and LR2 Prior test: not reported Threshold for test positivity predefined: yes Threshold for test positivity: RMI I 200; LR2 10, 25 and 43 Type of ultrasound (TAS, TVS or both): TVS Operator experience of sonographer (generalist, specialist or trainee): experienced gynaecologist Type of technology or manufacturer of biomarker test: not reported
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 66, borderline 7, malignant 23, metastatic and others 2 Target condition: OC/EOC (EOC 78%)
Flow and timing	
Comparative	RMI I vs LR2
Notes	
Methodological quality	

Niemi 2017 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (RMI)			

Niemi 2017 *(Continued)*

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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		High
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		

Niemi 2017 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (LR2)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		High
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	
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		

Niemi 2017 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			High
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		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	Yes		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	Yes		
Could the conduct of the comparative studies have introduced bias?		Low risk	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?			Low concern

Nikolova 2016

Study characteristics	
Patient Sampling	Country: Macedonia Centres: single Study design: within-person comparison Recruitment: prospective Method of patient selection: consecutive Inappropriate exclusions: none reported
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: premenopausal women aged ≥ 18 years with USS confirming an ovarian cyst/mass and scheduled for surgical intervention

Nikolova 2016 *(Continued)*

	Sample size: 105 Age range: OC 30–50 years, benign 18–50 years Mean age: malignant 42.46 (SD 8.21) years, benign 36.90 (SD 10.12) years Percentage postmenopausal (n): 0%
Index tests	Test: ROMA and RMI I Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: ROMA premenopausal 7.4, RMI 250 Type of ultrasound (TAS, TVS or both): TVS Operator experience of sonographer (generalist, specialist or trainee): not reported Type of technology or manufacturer of biomarker test: USS was performed using a Voluson E8, 4–9 MHz RIC5-9D vaginal transducer. Sera samples were analysed using Architect CA125 II and Architect HE4 reagents on an Abbott Platform
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 94, borderline not reported, malignant 11, metastatic and others not reported Target condition: EOC only
Flow and timing	
Comparative	ROMA vs RMI I (250)
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Low risk	

Nikolova 2016 *(Continued)*

A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated		
B) Prior test in primary care: self-reported symptoms		
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound		
Are there concerns that the included patients and setting do not match the review question?		Unclear
DOMAIN 2: Index Test (ADNEX)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		

Nikolova 2016 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Unclear
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern

Nikolova 2016 *(Continued)*

DOMAIN 2: Index Test (LR2)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
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		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for par-			

Nikolova 2016 *(Continued)*

ticipants receiving one or other index test or testing strategy the same?		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	Yes	
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	Unclear	
Could the conduct of the comparative studies have introduced bias?	Unclear risk	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		Low concern

Novotny 2012

Study characteristics	
Patient Sampling	Country: Czech Republic Centres: single Study design: within-person comparison Recruitment: unclear Method of patient selection: convenience Inappropriate exclusions: premenopausal women excluded Comments (if applicable): N/A
Patient characteristics and setting	Clinical setting: secondary Study entry criteria: women with pelvic abnormalities Sample size: 256 Age range: 47–93 years Mean age: benign 65.28 years, malignant 64.37 years Median age: benign 64 years, malignant 63 years Percentage postmenopausal (n): 100% (256)
Index tests	Combination ROMA Prior test: unclear Threshold for test positivity predefined: no Threshold for test positivity: premenopausal 26.3, postmenopausal 37.7 Type of ultrasound (TAS, TVS or both): N/A

Novotny 2012 *(Continued)*

	Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: Architect
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 256, borderline not reported, malignant 21, metastatic and others not reported Staging: early not reported, late not reported, unstaged not reported
Flow and timing	
Comparative	N/A
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
A) Includes all ages regardless of menopausal status or justify restrictions	Unclear		
B) Includes all stages and types of ovarian cancer	Unclear		
C) Includes comorbidities such as infertility and endometriosis	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			

Novotny 2012 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (RMI)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?

Novotny 2012 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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?	Unclear	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		

Novotny 2012 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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?			
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?			
Could the conduct of the comparative studies have introduced bias?			

Novotny 2012 *(Continued)*

Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	

Ortiz-Munoz 2014

Study characteristics	
Patient Sampling	Country: Spain Centres: single Study design: within-person comparison Recruitment: retrospective cross-sectional study Method of patient selection: convenience Inappropriate exclusions (all stages, all ages, included comorbidities such as infertility or endometriosis): unclear Comments (if applicable): N/A
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: women with gynaecological symptoms, diagnosed with primary OC Sample size: 148 Age range: not reported Mean age: not reported Median age: benign premenopausal 39.5 (SD 8.4) years, postmenopausal 56 (SD 11.5) years; malignant premenopausal 40.5 (SD 5.8) years, postmenopausal 57 (SD 9.4) years Percentage postmenopausal (n): 70% (104)
Index tests	Combination ROMA Prior test: symptoms Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 11.4, postmenopausal 29.9 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: LIA Comments: all blood tests performed 1 day prior to surgery
Target condition and reference standard(s)	22 benign cases were considered benign? on follow-up but duration of follow-up not detailed.

Ortiz-Munoz 2014 *(Continued)*

	Histology (n): benign 119, borderline not reported, malignant 29, metastatic and others not reported Staging: early 6, late 23, unstaged 0
Flow and timing	22 women diagnosed with simple ovarian cysts by TVS, unclear if they were based on follow-up, or duration of follow-up.
Comparative	N/A
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
A) Includes all ages regardless of menopausal status or justify restrictions	Unclear		
B) Includes all stages and types of ovarian cancer	Unclear		
C) Includes comorbidities such as infertility and endometriosis	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			

Ortiz-Munoz 2014 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (RMI)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?

Ortiz-Munoz 2014 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		

Ortiz-Munoz 2014 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?			
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
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?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?			
Could the conduct of the comparative studies have introduced bias?			
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?			

Park 2019

Study characteristics	
Patient Sampling	Country: Korea Centres: single Study design: non-comparative Recruitment: retrospective Method of patient selection: consecutive Inappropriate exclusions: 2 cases of non-EOC excluded from analysis
Patient characteristics and setting	Clinical setting: secondary Study entry criteria: women aged > 18 years for whom gynaecologists had requested HE4, CA125 and ROMA tests to evaluate a pelvic mass; 2 groups of participants considered: • malignant cases: 309 participants with available pathological examination reports of a biopsy • benign cases: 134 participants with imaging studies with minimum 4 weeks' follow-up and without biopsy Sample size: 433 (biopsy 309, follow-up 134) Age range: not reported Median age: EOC 52.3 (SD 6.1) years; benign 43.0 (SD 21) years, BOT 47.8 (SD 12.9) years Percentage postmenopausal (n): biopsy: 26% (81) Follow-up: minimum 28 weeks; median 29 weeks
Index tests	Test: ROMA Prior test: USS, CT or MRI Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 7.4, postmenopausal 25.3 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: HE4 and CA125 measured using the ARCHITECT HE4 assay (Product Number: B2P540) and the CA125 II assay (Product Number: B2K450) (Abbott Diagnostics, Abbott Park, IL, USA).
Target condition and reference standard(s)	Histology: 309 (69%) Follow-up: 134 (31%) Duration of follow-up: median 29 weeks (minimum 4 weeks)

Park 2019 *(Continued)*

	Histology (n): benign 406, borderline 15, malignant EOC 18 (4%), non-EOC 2 (< 1%), metastatic and others 2 (< 1%)
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			

Park 2019 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (RMI)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers

Park 2019 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		

Park 2019 *(Continued)*

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?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?			
Could the conduct of the comparative studies have introduced bias?			
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?			

Partheen 2011a

Study characteristics	
Patient Sampling	Country: Sweden Centres: single

Partheen 2011a *(Continued)*

	Study design: within-person comparison Recruitment: prospective cross-sectional study Method of patient selection: convenience Inappropriate exclusions (all stages, all ages, included comorbidities such as infertility or endometriosis): solid and mass were excluded, non-EOC tumours were excluded.
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: women with complex cystic mass and suspicious of malignancy undergoing surgery Sample size: 374 Age range: not reported Mean age: not reported Median age: not reported Percentage postmenopausal (n): 73.7% (276) Comments: women aged > 56 years were considered postmenopausal; women aged < 47 to 56 years were considered menopausal if > 12 months of amenorrhoea
Index tests	Combination ROMA Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: specificity fixed at 75% premenopausal 17.3%, postmenopausal 26.0% Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): unclear Type of technology or manufacturer of biomarker test: HE4 (EIA), CA125 (Abbott)
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 215, borderline 45, malignant 108, metastatic and others 6 others (? Mets) Staging: early 57, late 57, unstaged 0 Comments: women with final histology reporting the tumour was non-ovarian were excluded: BOT excluded for analysis for ROMA
Flow and timing	
Comparative	N/A
Notes	
Methodological quality	

Partheen 2011a *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
A) Includes all ages regardless of menopausal status or justify restrictions	Unclear		
B) Includes all stages and types of ovarian cancer	Unclear		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		High risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (RMI)			

Partheen 2011a *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ACOG)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ROMA)	
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?	Unclear

Partheen 2011a *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Unclear	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
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?		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		

Partheen 2011a *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear
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	
Could the patient flow have introduced bias?		High risk
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?		
Could the conduct of the comparative studies have introduced bias?		
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		

Prskalo 2015

Study characteristics	
Patient Sampling	Country: Croatia Centres: single Study design: non-comparative Recruitment: retrospective Method of patient selection: unclear Inappropriate exclusions: none reported
Patient characteristics and setting	Clinical setting: mixed Study entry criteria: women with suspected adnexal mass on a TVS scheduled for elective surgery Sample size: 159

Prskalo 2015 *(Continued)*

	Age range: not reported Mean age: premenopausal 36.9 (SD 8.9) years, postmenopausal 60.2 (SD 9.6) years Percentage postmenopausal (n): 64% (102)
Index tests	Test: ROMA Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 11.7, postmenopausal 29.9 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: HE4 and CA125 measured by electrochemiluminescence immunoassay on the Cobas e411 analyser (Hitachi, Tokyo, Japan; Roche, Mannheim, Germany)
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 105, borderline 11, malignant 43, metastatic and others none
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			

Prskalo 2015 *(Continued)*

B) Prior test in primary care: self-reported symptoms	
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound	
Are there concerns that the included patients and setting do not match the review question?	Unclear
DOMAIN 2: Index Test (ADNEX)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (RMI)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	

Prskalo 2015 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		

Prskalo 2015 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
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	
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?	Yes	

Prskalo 2015 *(Continued)*

For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	Yes
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	Unclear
Could the conduct of the comparative studies have introduced bias?	Unclear risk
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	Low concern

Radosa 2011

Study characteristics	
Patient Sampling	Patient sampling Country: Germany Centres: single Study design: within-person comparison Recruitment: unclear Method of patient selection: consecutive Inappropriate exclusions (all stages, all ages, included comorbidities such as infertility or endometriosis): none Comments: level 2 sonographers performed or supervised USS
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: women with adnexal mass who subsequently underwent surgery were selected Sample size: not reported Age range: not reported Mean age: 43.3 years Median age: not reported Percentage postmenopausal (n): 32% (442) Comments: N/A
Index tests	Combination RMI Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: RMI > 200 Type of ultrasound (TAS, TVS or both): both

Radosa 2011 *(Continued)*

	Operator experience of sonographer (generalist, specialist or trainee): specialist level 2 Type of technology or manufacturer of biomarker test: CLIA
Target condition and reference standard(s)	Only surgical patients included Follow-up: none Duration of follow-up: N/A Histology (n): benign 1260, borderline 19, malignant 79, metastatic and others 4 Staging: early 11 (OC), late 68 (OC), unstaged borderline not reported
Flow and timing	
Comparative	N/A
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
A) Includes all ages regardless of menopausal status or justify restrictions	Unclear		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			

Radosa 2011 *(Continued)*

If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
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?	Yes	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Unclear
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		

Radosa 2011 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ROMA)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (LR2)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?

Radosa 2011 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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?			
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?			
Could the conduct of the comparative studies have introduced bias?			

Radosa 2011 *(Continued)*

Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	

Richards 2015

Study characteristics	
Patient Sampling	Country: Australia Centres: single Study design: within-person comparison Recruitment: prospective Method of patient selection: unclear Inappropriate exclusions: none
Patient characteristics and setting	Clinical setting: mixed Study entry criteria: women undergoing surgery for a complex pelvic mass, presumed to be arising from the ovary Sample size: 50 Age range: not reported Mean age: not reported Median age: 60 years Percentage postmenopausal (n): 58% (29)
Index tests	Test: RMI I and ROMA Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: RMI I 200; ROMA: premenopausal 7.4, postmenopausal 25.3 Type of ultrasound (TAS, TVS or both): not reported Operator experience of sonographer (generalist, specialist or trainee): not reported Type of technology or manufacturer of biomarker test: the tumour markers were determined by the use of chemiluminescent enzyme immunoassay on an ARCHITECT analyser (Abbott Diagnostics, North Ryde, NSW, Australia)
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 30, borderline 4, malignant 16, metastatic and others not reported Target condition: EOC

Richards 2015 *(Continued)*

Flow and timing			
Comparative	ROMA vs RMI I		
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			

Richards 2015 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
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?	Yes	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Unclear
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		

Richards 2015 *(Continued)*

DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 3: Reference Standard		
Is the reference standards likely to correctly classify the target condition?	Yes	

Richards 2015 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?	Yes		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	Unclear		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	Unclear		
Could the conduct of the comparative studies have introduced bias?		Unclear risk	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?			Low concern

Romagnolo 2016

Study characteristics	
Patient Sampling	Country: Italy Centres: multicentre Study design: non-comparative Recruitment: prospective Method of patient selection: consecutive

Romagnolo 2016 *(Continued)*

	Inappropriate exclusions: non-EOC excluded
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: not reported Sample size: 387 Age range: not reported Mean age: premenopausal 37.6 (SD 8.6) years, postmenopausal 63 (SD 9.5) years Percentage postmenopausal (n): 38% (148)
Index tests	Test: ROMA Prior test: ultrasound Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 13.1, postmenopausal 27.7 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: CA125 measured by a CMIA on the automated Architect i2000SR platform (Abbott Diagnostics, Chicago, IL, USA) and HE4 by the HE4 EIA assay (Fujirebio Diagnostics AB, Gothenburg, Sweden)
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 290, borderline 15, malignant 73 (EOC), 9 (non-EOC), metastatic and others 6 (not included in the analysis)
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		

Romagnolo 2016 *(Continued)*

B) Includes all stages and types of ovarian cancer	Yes	
C) Includes comorbidities such as infertility and endometriosis	Yes	
Could the selection of patients have introduced bias?		Low risk
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated		
B) Prior test in primary care: self-reported symptoms		
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound		
Are there concerns that the included patients and setting do not match the review question?		Unclear
DOMAIN 2: Index Test (ADNEX)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		

Romagnolo 2016 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		

Romagnolo 2016 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (LR2)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
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		

Romagnolo 2016 *(Continued)*

Could the patient flow have introduced bias?	High risk
DOMAIN 5: Comparative	
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?	
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	
Could the conduct of the comparative studies have introduced bias?	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	

Salim 2018

Study characteristics	
Patient Sampling	Country: Pakistan Centres: single Study design: non-comparative Recruitment: prospective Method of patient selection: unclear Inappropriate exclusions: none
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: postmenopausal women with ovarian mass (> 2 cm) on pelvic ultrasound examination, attending gynaecology clinics, planned for surgical intervention Sample size: 260 Age range: 40–65 years Mean age: 49.28 (SD 6.26) years Median age: 48 years Percentage postmenopausal (n): 100%
Index tests	Test: ROMA Prior test: not reported

Salim 2018 *(Continued)*

	Threshold for test positivity predefined: yes Threshold for test positivity: postmenopausal 27.7 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: serums were analysed for the quantification of CA125 and HE4 on automated immunoassay analyser, Abbot ARCHITECT i1000 by CMIA method.
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 138, borderline not reported, malignant 122, metastatic and others not reported
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Unclear		
C) Includes comorbidities such as infertility and endometriosis	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			

Salim 2018 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (RMI)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?

Salim 2018 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		

Salim 2018 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
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		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?	Yes		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	Yes		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	Yes		

Salim 2018 *(Continued)*

Could the conduct of the comparative studies have introduced bias?	Low risk
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	Low concern

Sayasneh 2013a

Study characteristics	
Patient Sampling	Country: UK Centres: multicentre (3) Study design: within-person comparison Recruitment: prospective cross-sectional study Method of patient selection: consecutive Inappropriate exclusions: none Comments (if applicable): N/A
Patient characteristics and setting	Clinical setting: mixed Study entry criteria: not reported Sample size: 255 (301 in Sayasneh 2013 secondary study) Age range: not reported Mean age: 46 years Median age: not reported Percentage postmenopausal (n): 35% (117) Comments: N/A
Index tests	Combination RMI I and LR2 Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: LR2-probability cut-off of 10% is considered malignant. RMI ≥ 200 Interval between application of index test and reference standard: < 120 days; 1 women excluded as surgery after 120 days Type of ultrasound (TAS, TVS or both): both Operator experience of sonographer (generalist, specialist or trainee): level 1 and level 2 (10 were excluded as level 3 scan) Type of technology or manufacturer of biomarker test: not reported

Sayasneh 2013a *(Continued)*

Target condition and reference standard(s)	Histology (%): 98; surgical mix but no histology in 5 cases (2 ovarian torsion and 3 tubo-ovarian abscess – abscess confirmed by microscopy culture) Follow-up: 2 of ovarian torsion after reporting were followed up Duration of follow-up: 6 months* Histology (n): benign 181, borderline 18, malignant 48, metastatic and others 8 Staging: early not reported, late not reported, unstaged not reported Comments: despite follow-up of 6 months reference standard classified as low concern as it combination of surgical visualisation and follow-up.
Flow and timing	
Comparative	LR2 vs RMI
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
A) Includes all ages regardless of menopausal status or justify restrictions	Unclear		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		High risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			

Sayasneh 2013a *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
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?	Yes	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Unclear
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		

Sayasneh 2013a *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ROMA)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (LR2)	
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
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes

Sayasneh 2013a *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
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?			
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?	Yes		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	Unclear		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	Unclear		

Sayasneh 2013a *(Continued)*

Could the conduct of the comparative studies have introduced bias?	Unclear risk
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	Low concern

Shen 2017

Study characteristics	
Patient Sampling	Country: China Centres: multicentre Study design: non-comparative Recruitment: prospective Method of patient selection: unclear Inappropriate exclusions: none
Patient characteristics and setting	Clinical setting: mixed Study entry criteria: women aged ≥ 18 years referred to a participating centre with a pelvic mass or an ovarian cyst and planning to undergo surgery Sample size: 684 Age range: 42–82 years Mean age: 58.8 (SD 8.6) years Percentage postmenopausal (n): 25% (174)
Index tests	Test: ROMA Prior test: pelvic USS, CT, MRI and medical history (diagnosis and treatment of pelvic mass and history of renal disease) Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 7.4, postmenopausal 25.3 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: CA125, HE4 measured using the Architect instrument and reagents (Abbott Diagnostics)
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 482, borderline 18, malignant 169, metastatic 7, others 8

Shen 2017 *(Continued)*

	Target condition: EOC		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			

Shen 2017 *(Continued)*

A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (RMI)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers

Shen 2017 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 3: Reference Standard		

Shen 2017 *(Continued)*

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		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
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		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?			
Could the conduct of the comparative studies have introduced bias?			
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?			

Stiekma 2014

Study characteristics	
Patient Sampling	Country: the Netherlands Centres: single Study design: within-person comparison

Stiekma 2014 *(Continued)*

	Recruitment: retrospective cross-sectional study Method of patient selection: convenience Inappropriate exclusions (all stages, all ages, included comorbidities such as infertility or endometriosis): BOT and non-EOC excluded
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: histologically confirmed EOC or benign ovarian disease referred to the institute Sample size: 181 Age range: not reported Mean age: benign 47 years, malignant 57 years Median age: not reported Percentage postmenopausal (n): 79% (143) Comments: none
Index tests	ROMA Prior test: unclear Threshold for test positivity predefined: no Threshold for test positivity: ROMA; premenopausal 0.129, postmenopausal 0.278 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: CA125 and HE4 (both Abbott) Comments: N/A
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 34, borderline excluded, malignant 147, metastatic and others not reported Staging: early 24, late 123
Flow and timing	
Comparative	
Notes	

Methodological quality

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

Stiekma 2014 *(Continued)*

Was a consecutive or random sample of patients enrolled?	No	
Was a case-control design avoided?	Yes	
Did the study avoid inappropriate exclusions?	No	
A) Includes all ages regardless of menopausal status or justify restrictions	Unclear	
B) Includes all stages and types of ovarian cancer	No	
C) Includes comorbidities such as infertility and endometriosis	Unclear	
Could the selection of patients have introduced bias?		High risk
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated		
B) Prior test in primary care: self-reported symptoms		
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound		
Are there concerns that the included patients and setting do not match the review question?		Unclear
DOMAIN 2: Index Test (ADNEX)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		

Stiekma 2014 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ACOG)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ROMA)	
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?	Yes
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Unclear

Stiekma 2014 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
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	
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear

Stiekma 2014 *(Continued)*

DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	No	
Could the patient flow have introduced bias?		High risk
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?		
Could the conduct of the comparative studies have introduced bias?		
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		

Szubert 2016a

Study characteristics	
Patient Sampling	Country: Poland Centres: single Study design: non-comparative Recruitment: retrospective Method of patient selection: unclear Inappropriate exclusions: (quote) "no specific exclusion criteria"
Patient characteristics and setting	Clinical setting: unclear, probably tertiary Study entry criteria: women needing surgery for an ovarian tumour Sample size: 204 Age range: 15–84 years Mean age: not reported

Szubert 2016a *(Continued)*

	Median age: 46 years Percentage postmenopausal (n): 54% (66)
Index tests	Test: ADNEX Prior test: not reported Threshold for test positivity predefined: yes Threshold for test positivity: 2000 IOTA criteria 10% Type of ultrasound (TAS, TVS or both): both Operator experience of sonographer (generalist, specialist or trainee): specialist Type of technology or manufacturer of biomarker test: tumours evaluated using Aloka Alpha 10 with 3.75–7.5 MHz endovaginal probe and Aloka 3500 with a 7.5 MHz endovaginal probe (Hitach Aloka, Tokyo, Japan). A transabdominal probe was used in case of large tumours.
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 134, borderline 12, malignant 58, metastatic and others not reported
Flow and timing	
Comparative	N/A
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			

Szubert 2016a *(Continued)*

C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound		
Are there concerns that the included patients and setting do not match the review question?		Unclear
DOMAIN 2: Index Test (ADNEX)		
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?	Yes	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk	
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		High
DOMAIN 2: Index Test (RMI)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		

Szubert 2016a *(Continued)*

DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ROMA)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (LR2)
Were the index test results interpreted without knowledge of the results of the reference standard?

Szubert 2016a *(Continued)*

If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			

Szubert 2016a *(Continued)*

For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	
Could the conduct of the comparative studies have introduced bias?	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	

Szubert 2016b

Study characteristics	
Patient Sampling	Country: Spain Centres: single Study design: non-comparative Recruitment: retrospective Method of patient selection: unclear Inappropriate exclusions: quote: "no specific exclusion criteria"
Patient characteristics and setting	Clinical setting: unclear, probably tertiary Study entry criteria: women needing surgery for an ovarian tumour Sample size: 128 Age range: 15–81 years Mean age: not reported Median age: 47 years Percentage postmenopausal (n): 42% (52)
Index tests	Test: ADNEX Prior test: not reported Threshold for test positivity predefined: yes Threshold for test positivity: 2000 IOTA criteria 10% Type of ultrasound (TAS, TVS or both): both Operator experience of sonographer (generalist, specialist or trainee): specialist Type of technology or manufacturer of biomarker test: TVS or transrectal ultrasound using a Voluson E8 equipped with an RIC5-9MHz endovaginal probe (GE Healthcare, Milwaukee, USA). A transabdominal probe was used in case of large tumours.
Target condition and reference standard(s)	Only surgical patients included

Szubert 2016b *(Continued)*

	Histology (n): benign 89, borderline 4, malignant 35, metastatic and others none		
Flow and timing			
Comparative	N/A		
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
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?	Yes		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes		

Szubert 2016b *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?	Unclear risk
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High
DOMAIN 2: Index Test (RMI)	
DOMAIN 2: Index Test (ACOG)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ROMA)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	

Szubert 2016b *(Continued)*

A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
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	
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	

Szubert 2016b *(Continued)*

Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	Unclear risk
DOMAIN 5: Comparative	
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?	
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	
Could the conduct of the comparative studies have introduced bias?	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	

Teh 2018

Study characteristics	
Patient Sampling	Country: Malaysia Centres: single Study design: non-comparative Recruitment: prospective Method of patient selection: unclear Inappropriate exclusions: low malignant potential tumours were included in the benign tumour group during analysis
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: women aged ≥ 18 years with pelvic mass(es) suspected of originating in the ovary who had been scheduled for surgery or radiological-guided biopsy Sample size: 129 Age range: not reported Mean age: not reported Median age: 37 (IQR 27.5–48.5) years Percentage postmenopausal (n): 21% (27)
Index tests	Test: ROMA

Teh 2018 *(Continued)*

	Prior test: not reported Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 11.4, postmenopausal 29.9 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: the serum samples were tested using the Elecsys HE4 assay (Roche Diagnostics, Mannheim, Germany) and Elecsys CA125 II assay (Roche Diagnostics, Mannheim, Germany) via electrochemiluminescence immunoassay technology.
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 97, borderline 10, malignant 27, metastatic and others 3
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			

Teh 2018 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?	Unclear
DOMAIN 2: Index Test (ADNEX)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (RMI)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ACOG)	

Teh 2018 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		

Teh 2018 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
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		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			

Teh 2018 *(Continued)*

For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	
Could the conduct of the comparative studies have introduced bias?	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	

Terlikowska 2016

Study characteristics	
Patient Sampling	Country: Poland Centres: multicentre Study design: non-comparative Recruitment: retrospective Method of patient selection: unclear Inappropriate exclusions: non-EOC excluded
Patient characteristics and setting	Clinical setting: mixed (secondary and tertiary) Study entry criteria: Caucasian women surgically treated on account of benign ovarian disease and epithelial cancer according to international treatment guidelines Sample size: 224 Age range: premenopausal 25–49 years, postmenopausal 53–74 years Mean age: not reported Median age: premenopausal 36 years, postmenopausal 63 years Percentage postmenopausal (n): 46% (104)
Index tests	Test: ROMA Prior test: not reported Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 11.4, postmenopausal 29.9 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: concentrations of HE4 and CA125 were assessed with the electrochemilumi-

Terlikowska 2016 *(Continued)*

	nescence (ECLIA) technique on Cobas e411 (Roche Diagnostics, Switzerland) analyser
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 128, borderline not reported, malignant 96, metastatic and others none reported
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			

Terlikowska 2016 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (RMI)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?

Terlikowska 2016 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		

Terlikowska 2016 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?			
Could the conduct of the comparative studies have introduced bias?			
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?			

Terzic 2013

Study characteristics	
Patient Sampling	Country: Serbia Centres: single Study design: non-comparative Recruitment: unclear Method of patient selection: consecutive Inappropriate exclusions (all stage, all ages, included comorbidities such as infertility or endometriosis): unclear Comments (if applicable): N/A
Patient characteristics and setting	Clinical setting: secondary Study entry criteria: women who had undergone surgery for adnexal mass Sample size: 540 Age range: 18–82 years Mean age: 53.44 (SD 16.82) Median age: not reported Percentage postmenopausal (n): 31.61% (184) Comments: 341 participants were symptomatic (benign 255, BOT 66, OC 66) but data could not be disaggregated as index test results were not given separately for test-positive and test-negative patients.
Index tests	Combination RMI I Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: > 250 Type of ultrasound (TAS, TVS or both): unclear Operator experience of sonographer (generalist, specialist or trainee): specialist Type of technology or manufacturer of biomarker test: not reported
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 435, borderline 20, malignant 85, metastatic and others not reported Staging: early not reported, late not reported, unstaged not reported
Flow and timing	
Comparative	N/A

Terzic 2013 *(Continued)*

Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Unclear		
C) Includes comorbidities such as infertility and endometriosis	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			

Terzic 2013 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
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?	Yes	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		High
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		

Terzic 2013 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (LR2)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
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?	

Terzic 2013 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?		
Could the conduct of the comparative studies have introduced bias?		
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		

Testa 2014

Study characteristics	
Patient Sampling	Country: Europe Centres: multicentre; 18 centres in 6 countries (Sweden, Belgium, Italy, Poland, Spain and Czech Republic) Study design: within-person comparison Recruitment: retrospective cross-sectional study Method of patient selection: convenience Inappropriate exclusions: none

Testa 2014 *(Continued)*

	Comment: N/A
Patient characteristics and setting	Clinical setting: mixed Study entry criteria: women presenting with adnexal mass and undergoing TVS by 1 of the principal investigators and surgery within 120 days after examination Sample size: 2403 Age range: 33–66 years Median age: benign 44 years, malignant 57 years Percentage postmenopausal (n): 44% (1049)
Index tests	Combination RMI I and LR2 Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: LR2-probability of malignancy ≥ 10%, RMI > 200 Type of ultrasound (TAS, TVS or both): both Operator experience of sonographer (generalist, specialist or trainee): specialist Type of technology or manufacturer of biomarker test: not reported CA125 results missing in 40% and multiple imputation was used to handle missing values.
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 1423, borderline 153, malignant 701, metastatic and others 126 Staging: early 316, late 470, unstaged 68 + 12 mets Pathologist was blinded to the outcome of index test
Flow and timing	Interval between application of index test and reference standard: ≤ 120 days, 66 women were excluded as surgery after 120 days. 13 women were excluded because of incomplete final histology.
Comparative	RMI vs LR2
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		

Testa 2014 *(Continued)*

Was a case-control design avoided?	Yes	
Did the study avoid inappropriate exclusions?	Yes	
A) Includes all ages regardless of menopausal status or justify restrictions	Yes	
B) Includes all stages and types of ovarian cancer	Yes	
C) Includes comorbidities such as infertility and endometriosis	Yes	
Could the selection of patients have introduced bias?		Unclear risk
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated		
B) Prior test in primary care: self-reported symptoms		
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound		
Are there concerns that the included patients and setting do not match the review question?		Unclear
DOMAIN 2: Index Test (ADNEX)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
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?	Yes	

Testa 2014 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		High
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		

Testa 2014 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (LR2)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		High
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?		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear

Testa 2014 *(Continued)*

DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	No	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	No	
Could the patient flow have introduced bias?		High risk
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?	Yes	
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	Unclear	
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	Unclear	
Could the conduct of the comparative studies have introduced bias?		Unclear risk
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		Low concern

Timmerman 2010

Study characteristics	
Patient Sampling	Country: Europe Centres: multicentre Study design: within-person comparison Recruitment: prospective cross-sectional study Method of patient selection: convenience Inappropriate exclusions: none Comments (if applicable): 15 patients who underwent surgery > 120 days after USS examination were excluded
Patient characteristics and setting	Clinical setting: mixed Study entry criteria: women with persistent adnexal mass undergoing surgery within 120 days Sample size: total 1938, 1522 women with CA125 included for RMI

Timmerman 2010 *(Continued)*

	Age range: 11–94 years Mean age: 46 years Percentage postmenopausal (n): 38% (742) Comments: 19 centres, 8 countries
Index tests	Combination RMI I and LR2 Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: LR2-probability of malignancy ≥ 10%, RMI > 200 Type of ultrasound (TAS, TVS or both): both Operator experience of sonographer (generalist, specialist or trainee): specialist Type of technology or manufacturer of biomarker test: not reported
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 542, borderline 111, malignant 373, metastatic and others 58 Staging: early 100 (BOT) + 100 (invasive), late 9 (BOT) + 232 (invasive), unstaged 2 (BOT) + 99 (invasive) Pathologist had no knowledge of the ultrasound results
Flow and timing	1501 women included for analysis for RMI; 1147 participants with CA125 results included
Comparative	RMI I vs LR2
Notes	Same cohort as Di Legge 2012 (see above). Data for RMI I extracted from Di Legge 2012.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		

Timmerman 2010 *(Continued)*

C) Includes comorbidities such as infertility and endometriosis	Yes
Could the selection of patients have introduced bias?	High risk
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated	
B) Prior test in primary care: self-reported symptoms	
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound	
Are there concerns that the included patients and setting do not match the review question?	High
DOMAIN 2: Index Test (ADNEX)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (RMI)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	

Timmerman 2010 *(Continued)*

A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ROMA)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers

Timmerman 2010 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (LR2)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Unclear
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?		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	No	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	No	
Could the patient flow have introduced bias?		High risk
DOMAIN 5: Comparative		

Timmerman 2010 *(Continued)*

For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?	Yes	
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	Unclear	
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	Unclear	
Could the conduct of the comparative studies have introduced bias?		Unclear risk
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		Low concern

van Calster 2014

Study characteristics	
Patient Sampling	Country: Europe Centres: multicentre (19) Study design: non-comparative Recruitment: prospective cross-sectional study Method of patient selection: consecutive Inappropriate exclusions (all stages, all ages, included comorbidities such as infertility or endometriosis): none Comments (if applicable): N/A
Patient characteristics and setting	Clinical setting: mixed (11/19 tertiary) Study entry criteria: women with an adnexal mass on USS and selected for surgery Sample size: 2403 (2124 analysed without metastatic and borderline) Age range: not reported Mean age: not reported Median age: not reported Percentage postmenopausal (n): not reported Comments: ADNEX includes age as a variable
Index tests	Combination ADNEX Prior test: unclear

van Calster 2014 *(Continued)*

	Threshold for test positivity predefined: no Threshold for test positivity: 3%, 5%, 10% and 15% disease positive probability of malignancy Interval between application of index test and reference standard: ≤ 120 days Type of ultrasound (TAS, TVS or both): both Operator experience of sonographer (generalist, specialist or trainee): not reported Type of technology or manufacturer of biomarker test: 5 manufacturers all using OC125 Ab
Target condition and reference standard(s)	Women selected for surgery OC; secondary metastatic OC Histology (n): benign 1423, borderline 153, malignant 701, metastasis or others 126 Staging: stage I 189, Stage II-IV 521
Flow and timing	
Comparative	N/A
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Low risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			

van Calster 2014 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?		Unclear
DOMAIN 2: Index Test (ADNEX)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Unclear
DOMAIN 2: Index Test (RMI)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ACOG)		

van Calster 2014 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ROMA)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (LR2)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?

van Calster 2014 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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?			
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			

van Calster 2014 *(Continued)*

For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	
Could the conduct of the comparative studies have introduced bias?	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	

van den Akker 2016

Study characteristics	
Patient Sampling	Only surgical patients included Histology (n): benign 128, borderline not reported, malignant 96, metastatic and others none reported
Patient characteristics and setting	Clinical setting: mixed (secondary and tertiary) Study entry criteria: women who were admitted for surgical treatment of an ovarian mass with unknown histology Sample size: 670 Age range: 13–93 years Mean age: not reported Median age: 54 years Percentage postmenopausal (n): 58% (390)
Index tests	Test: RMI I Prior test: not reported Threshold for test positivity predefined: yes Threshold for test positivity: 200 Type of ultrasound (TAS, TVS or both): both Operator experience of sonographer (generalist, specialist or trainee): specialist Type of technology or manufacturer of biomarker test: not reported; stated, "routine preoperative assessment included analysis of serum samples for cancer antigen 125 (CA125), and menopausal status was recorded".
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 531, borderline 46, malignant 93, metastatic and others not reported
Flow and timing	

van den Akker 2016 *(Continued)*

Comparative
Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			

van den Akker 2016 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
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?	Yes	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		High
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		

van den Akker 2016 *(Continued)*

DOMAIN 2: Index Test (ROMA)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (LR2)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 3: Reference Standard	
Is the reference standards likely to correctly classify the target condition?	Yes

van den Akker 2016 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?			
Could the conduct of the comparative studies have introduced bias?			
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?			

van Gorp 2011

Study characteristics	
Patient Sampling	Country: Belgium Centres: single Study design: within-person comparison Recruitment: prospective cross-sectional study Method of patient selection: consecutive

van Gorp 2011 *(Continued)*

	Inappropriate exclusions (all stages, all ages, included comorbidities such as infertility or endometriosis): none Comments (if applicable): N/A
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: women diagnosed with pelvic mass undergoing surgery Sample size: 389 Age range: not reported Mean age: benign 46.3 (SD 16) years, malignant 57.8 (SD 12.6) years Median age: not reported Percentage postmenopausal (n): 41.4% (161)
Index tests	Combination Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: CA125 35 U/mL, HE4 70 pmol/L and 150 pmol/L Interval between application of index tests: < 3 months' interval Interval between application of index test and reference standard: < 3 months' interval Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: EIA
Target condition and reference standard(s)	Only surgical patients included Follow-up: none Duration of follow-up: N/A Histology (n): benign 228, borderline not reported, malignant 135, metastatic and others 26 Staging: early 51, late 80, unstaged 0
Flow and timing	
Comparative	See van Gorp 2012 below
Notes	van Gorp 2012 (see below) is a secondary publication to this study. RMI results are presented only in this publication while ROMA results are presented in both publications. Since van Gorp 2011 has a bigger cohort, results for ROMA were considered from this publication and therefore treated as a separate study.

van Gorp 2011 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Unclear		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		High risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			

van Gorp 2011 *(Continued)*

DOMAIN 2: Index Test (RMI)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ACOG)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ROMA)	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes

van Gorp 2011 *(Continued)*

If a threshold was used, was it pre-specified?	Yes	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
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?		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk

van Gorp 2011 *(Continued)*

Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear
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	
Could the patient flow have introduced bias?		Low risk
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?		
Could the conduct of the comparative studies have introduced bias?		
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		

van Gorp 2012

Study characteristics	
Patient Sampling	Country: Belgium Centres: single Study design: within-person comparison Recruitment: prospective cross-sectional study Method of patient selection: convenience Inappropriate exclusions (all stages, all ages, included comorbidities such as infertility or endometriosis): none Comments (if applicable): N/A
Patient characteristics and setting	Clinical setting: tertiary

van Gorp 2012 *(Continued)*

	Study entry criteria: women with a pelvic mass, scheduled for surgery Sample size: 374 Age range: not reported Mean age: benign 46.2 years (95% CI 44.1 to 48.3), malignant 57.7 years (95% CI 55.7 to 59.8) Percentage postmenopausal (n): 52.4% (196) Comments: following participants were excluded: 6 with presumed benign disease, 6 had no cyst at time of surgery, 4 with conservative management due to poor prognosis.
Index tests	Combination ROMA, RMI I Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: ROMA; premenopausal 12.5%, postmenopausal 14.4%, RMI I cut-off 200 Type of ultrasound (TAS, TVS or both): both Operator experience of sonographer (generalist, specialist or trainee): mixed Type of technology or manufacturer of biomarker test: EIA Comments: ultrasound was performed by an experienced sonographer or supervised by an experienced sonographer; the sonographer blinded to CA125 but blinding to symptoms not given.
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 224, borderline 31, malignant 94, metastatic and others 25 Staging: early 49 (only for EOC + BOT), late 72 (only for EOC + BOC), unstaged 0
Flow and timing	There was < 3 months between the blood test and reference standard but interval between ultrasound and reference standard was unclear.
Comparative	ROMA vs RMI I
Notes	This is a secondary publication to van Gorp 2011. RMI results are presented only in this publication while ROMA results are presented in both publications. Since van Gorp 2011 has bigger cohort, results for ROMA were considered from this publication and therefore treated as a separate study.

Methodological quality

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

van Gorp 2012 *(Continued)*

Was a consecutive or random sample of patients enrolled?	No	
Was a case-control design avoided?	Yes	
Did the study avoid inappropriate exclusions?	Unclear	
A) Includes all ages regardless of menopausal status or justify restrictions	Unclear	
B) Includes all stages and types of ovarian cancer	Yes	
C) Includes comorbidities such as infertility and endometriosis	Yes	
Could the selection of patients have introduced bias?		High risk
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated		
B) Prior test in primary care: self-reported symptoms		
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound		
Are there concerns that the included patients and setting do not match the review question?		Unclear
DOMAIN 2: Index Test (ADNEX)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
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?	Yes	

van Gorp 2012 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Unclear	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		High
DOMAIN 2: Index Test (ACOG)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		

van Gorp 2012 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (LR2)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
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?	
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?	
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

van Gorp 2012 *(Continued)*

DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?	Yes	
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?	Unclear	
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	Unclear	
Could the conduct of the comparative studies have introduced bias?		Unclear risk
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		Low concern

Vural 2016

Study characteristics	
Patient Sampling	Country: Turkey Centres: single Study design: non-comparative Recruitment: retrospective Method of patient selection: unclear Inappropriate exclusions: none
Patient characteristics and setting	Clinical setting: tertiary Study entry criteria: postmenopausal women with adnexal masses who underwent surgery Sample size: 139 Age range: 42–87 years Mean age: 61.1 (SD 8.9) years

Vural 2016 *(Continued)*

	Percentage postmenopausal (n): 100%
Index tests	Test: RMI I Prior test: not reported Threshold for test positivity predefined: yes Threshold for test positivity: RMI I 200 Type of ultrasound (TAS, TVS or both): both Operator experience of sonographer (generalist, specialist or trainee): specialised gynaecologist Type of technology or manufacturer of biomarker test: grey scale ultrasonographic imaging of the cases was performed by an expert radiologist via ultrasound device with five MHz convex abdominal and 8 MHz vaginal probes.
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 87, borderline 8, malignant 44, metastatic and others 11 Target condition: OC/EOC (73% EOC)
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			

Vural 2016 *(Continued)*

C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound		
Are there concerns that the included patients and setting do not match the review question?		Unclear
DOMAIN 2: Index Test (ADNEX)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
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?	Yes	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		High

Vural 2016 *(Continued)*

DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ROMA)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (LR2)
Were the index test results interpreted without knowledge of the results of the reference standard?

Vural 2016 *(Continued)*

If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			

Vural 2016 *(Continued)*

For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?	
Could the conduct of the comparative studies have introduced bias?	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?	

Xu 2016

Study characteristics	
Patient Sampling	Country: China Centres: single Study design: non-comparative Recruitment: retrospective Method of patient selection: unclear Inappropriate exclusions: 29 women with non-EOC excluded from analysis
Patient characteristics and setting	Clinical setting: mixed Study entry criteria: women with a pelvic mass (defined as a simple, complex or solid ovarian cyst/pelvic mass) Sample size: 566 Age range: not reported Mean age: malignant 57 years, benign 42 years Percentage postmenopausal (n): 28% (166)
Index tests	Test: ROMA Prior test: not reported Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 11.4, postmenopausal 29.9 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: HE4 and CA125 were determined on the Roche Cobas E170 analyser with Elecsys HE4 kits (Roche, Mannheim, Germany) and Elecsys CA125 kits (Roche, Mannheim, Germany). This assay utilises an electrochemiluminescent immunoassay method.

Xu 2016 *(Continued)*

Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 311, borderline 45, malignant 210, metastatic and others none reported
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	No		
C) Includes comorbidities such as infertility and endometriosis	Unclear		
Could the selection of patients have introduced bias?		High risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (ADNEX)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			

Xu 2016 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (RMI)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers
Are there concerns that the index test, its conduct, or interpretation differ from the review question?
DOMAIN 2: Index Test (ACOG)
Were the index test results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it pre-specified?
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?
Could the conduct or interpretation of the index test have introduced bias?
A) Was ultrasound performed in all patients by non-specialised sonographers

Xu 2016 *(Continued)*

B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		

Xu 2016 *(Continued)*

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		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?			
Could the conduct of the comparative studies have introduced bias?			
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?			

Zhang 2015

Study characteristics	
Patient Sampling	Country: China Centres: multicentre

Zhang 2015 *(Continued)*

	Study design: non-comparative Recruitment: prospective Method of patient selection: unclear Inappropriate exclusions: non-EOC excluded
Patient characteristics and setting	Clinical setting: unclear Study entry criteria: women with and without pelvic mass on USS scheduled for surgery Sample size: 612 Age range: not reported Mean age: not reported Median age (25th centile, 75th centile): benign: premenopausal 41 (35, 46), postmenopausal 57 (54, 68); malignant (EOC): premenopausal 43 (38, 47), postmenopausal 59 (54, 65) Percentage postmenopausal (n): 37% (232)
Index tests	Test: ROMA Prior test: USS; adnexal lesions reported according to IOTA Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 11.4, postmenopausal 29.9 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: Roche Elecsys Cobas 601 platform and the matched reagents Roche Diagnostics (Basel, Switzerland)
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 348, borderline not reported, malignant 264, metastatic and others excluded
Flow and timing	
Comparative	N/A
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		

Zhang 2015 *(Continued)*

Was a case-control design avoided?	Unclear	
Did the study avoid inappropriate exclusions?	Yes	
A) Includes all ages regardless of menopausal status or justify restrictions	Yes	
B) Includes all stages and types of ovarian cancer	Yes	
C) Includes comorbidities such as infertility and endometriosis	Yes	
Could the selection of patients have introduced bias?		Unclear risk
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated		
B) Prior test in primary care: self-reported symptoms		
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound		
Are there concerns that the included patients and setting do not match the review question?		Unclear
DOMAIN 2: Index Test (ADNEX)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (RMI)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		

Zhang 2015 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ACOG)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ROMA)	
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
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes

Zhang 2015 *(Continued)*

If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
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	
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?		
Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear

Zhang 2015 *(Continued)*

DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk
DOMAIN 5: Comparative		
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?		
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?		
For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?		
Could the conduct of the comparative studies have introduced bias?		
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		

Zhang 2019

Study characteristics	
Patient Sampling	Country: China Centres: single Study design: non-comparative Recruitment: prospective Method of patient selection: unclear Inappropriate exclusions: borderline excluded from analysis
Patient characteristics and setting	Clinical setting: mixed Study entry criteria: women with ovarian tumour Sample size: 373 Age range: 12–77 years Mean age: 51 years

Zhang 2019 *(Continued)*

	Percentage postmenopausal (n): 50% (185)
Index tests	Test: ROMA Prior test: unclear Threshold for test positivity predefined: yes Threshold for test positivity: premenopausal 11.4, postmenopausal 29.9 Type of ultrasound (TAS, TVS or both): N/A Operator experience of sonographer (generalist, specialist or trainee): N/A Type of technology or manufacturer of biomarker test: HE4 and CA125 serum levels were analysed by Roche cobas 60 0 0 analyser using reagents that provided by Roche (Basel, Switzerland).
Target condition and reference standard(s)	Only surgical patients included Histology (n): benign 175, borderline 17, malignant 181, metastatic 4, others 4 stromal tumour, 1 germ cell tumour
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
A) Includes all ages regardless of menopausal status or justify restrictions	Yes		
B) Includes all stages and types of ovarian cancer	Yes		
C) Includes comorbidities such as infertility and endometriosis	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
A) All patients are symptomatic or symptomatic and asymptomatic can be disaggregated			
B) Prior test in primary care: self-reported symptoms			
C) Prior test secondary care: self-reported symptoms or self-reported symptoms plus one or more biochemical markers and ultrasound			

Zhang 2019 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?	Unclear
DOMAIN 2: Index Test (ADNEX)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (RMI)	
Were the index test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it pre-specified?	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	
Could the conduct or interpretation of the index test have introduced bias?	
A) Was ultrasound performed in all patients by non-specialised sonographers	
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
DOMAIN 2: Index Test (ACOG)	

Zhang 2019 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?		
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?		
Could the conduct or interpretation of the index test have introduced bias?		
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 2: Index Test (ROMA)		
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	
Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?	Yes	
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
A) Was ultrasound performed in all patients by non-specialised sonographers		
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers		
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 2: Index Test (LR2)		
Were the index test results interpreted without knowledge of the results of the reference standard?		
If a threshold was used, was it pre-specified?		

Zhang 2019 *(Continued)*

Were all components and thresholds of composite index test (including multivariable model) prespecified before their application?			
If a composite index test was used, were components of a composite index test/model defined and assessed in a similar way (e.g. in the same healthcare setting) for all participants?			
Could the conduct or interpretation of the index test have introduced bias?			
A) Was ultrasound performed in all patients by non-specialised sonographers			
B) i. Were symptoms interpreted without the knowledge of ultrasound and biomarkers; ii: was ultrasound interpreted without the knowledge of biomarkers			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
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		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Can borderline tumours be grouped with primary ovarian cancer for the purposes of analysis?			
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
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		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			
For studies comparing two or more index tests or testing strategies in different populations, were the selection criteria for participants receiving one or other index test or testing strategy the same?			
For within-study comparisons of index tests: was the interval between application of index test less than 3 months?			

Zhang 2019 *(Continued)*

For within-study comparison of individual index tests: were index tests interpreted blind to the results of other index test results?		
Could the conduct of the comparative studies have introduced bias?		
Is there concern that included patients have been selected in a different way to participants in non-comparative studies?		

CMIA: chemiluminescent microparticle immunoassay; CT: computed tomography; EOC: epithelial ovarian cancer; MRI: magnetic resonance imaging; N/A: not applicable; OC: ovarian cancer; ROMA: Risk of Ovarian Malignancy Algorithm; RMI: Risk of Malignancy Index; SD: standard deviation; TAS: transabdominal ultrasound; TVS: transvaginal ultrasound; USS: ultrasound scan.

DATA

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

Table Tests. Data tables by test

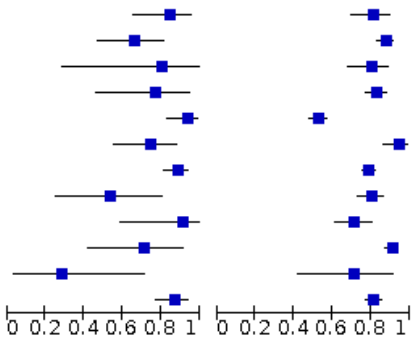
Test	No. of studies	No. of participants
1 ROMA 7.4 (± 2) premenopausal	12	3223
2 ROMA 25.3 (± 2) postmenopausal	15	2599
3 ROMA 12.5 premenopausal	3	302
4 ROMA 14.4 postmenopausal	3	299
5 ROMA 13.1 (± 2) premenopausal	27	4463
6 ROMA 27.7 (± 2) postmenopausal	13	2002
7 ROMA 7.4 premenopausal	10	3051
8 ROMA 25.3 postmenopausal	9	1386
9 ROMA 7.4/25.3 all	2	681
10 ROMA 12.5/14.4 all	3	601
11 ROMA 13.1 premenopausal	8	1353
12 ROMA 27.7 postmenopausal	9	1265
13 ROMA 13.1/27.7 all	5	1615
14 ROMA 11.4 premenopausal	11	2281
15 ROMA 29.9 postmenopausal	12	1797
18 ROMA mixed premenopausal	38	7616

Test	No. of studies	No. of participants
19 ROMA mixed postmenopausal	40	6099
20 ROMA mixed all	10	2897
21 RMI I 200 premenopausal	17	5233
22 RMI I 200 postmenopausal	17	4369
23 RMI I 200 all	5	4559
24 RMI I 250 premenopausal	2	461
25 RMI I 250 postmenopausal	2	220
26 RMI I 250 all	1	540
35 RMI mixed premenopausal	6	2990
36 RMI mixed postmenopausal	7	2099
37 RMI mixed all	6	5099
38 LR2 premenopausal	4	2843
39 LR2 postmenopausal	5	2157
40 LR2 all	3	4596
41 ADNEX 3% D+ probability all	1	2403
42 ADNEX 3% D+ probability premenopausal	1	1354
43 ADNEX 3% D+ probability postmenopausal	1	1049
44 ADNEX 5% D+ probability all	1	2403
45 ADNEX 5% D+ probability premenopausal	1	1354
46 ADNEX 5% D+ probability postmenopausal	1	1049
47 ADNEX 10% D+ probability all	1	2403
48 ADNEX 10% D+ probability premenopausal	4	1696
49 ADNEX 10% D+ probability postmenopausal	4	1365
50 ADNEX 15% D+ probability all	1	2403
51 ADNEX 15% D+ probability premenopausal	1	1354
52 ADNEX 15% D+ probability postmenopausal	1	1049
67 RMI I mixed premenopausal	19	5694
68 RMI I mixed postmenopausal	19	4589

Test 1. ROMA 7.4 (± 2) premenopausal

ROMA 7.4 (± 2) premenopausal

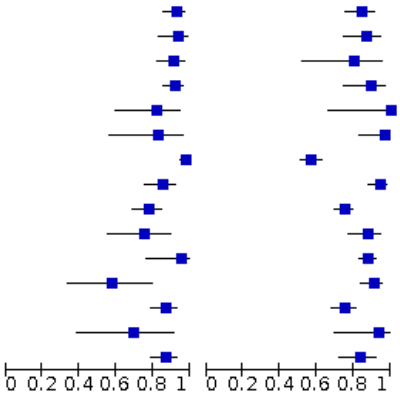
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Bandiera 2011	22	13	4	56	0.85 [0.65, 0.96]	0.81 [0.70, 0.90]		
Chan 2013	21	34	11	235	0.66 [0.47, 0.81]	0.87 [0.83, 0.91]		
Grenache 2015	4	13	1	52	0.80 [0.28, 0.99]	0.80 [0.68, 0.89]		
Huy 2018	10	37	3	180	0.77 [0.46, 0.95]	0.83 [0.77, 0.88]		
Karlsen 2012	46	251	3	279	0.94 [0.83, 0.99]	0.53 [0.48, 0.57]		
Kim 2011	23	4	8	67	0.74 [0.55, 0.88]	0.94 [0.86, 0.98]		
Li 2016	96	136	12	501	0.89 [0.81, 0.94]	0.79 [0.75, 0.82]		
Melo 2018	7	28	6	114	0.54 [0.25, 0.81]	0.80 [0.73, 0.86]		
Nikolova 2016	10	27	1	67	0.91 [0.59, 1.00]	0.71 [0.61, 0.80]		
Park 2019	10	29	4	282	0.71 [0.42, 0.92]	0.91 [0.87, 0.94]		
Richards 2015	2	4	5	10	0.29 [0.04, 0.71]	0.71 [0.42, 0.92]		
Shen 2017	60	79	9	347	0.87 [0.77, 0.94]	0.81 [0.77, 0.85]		



Test 2. ROMA 25.3 (± 2) postmenopausal

ROMA 25.3 (± 2) postmenopausal

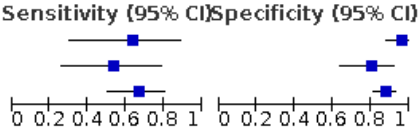
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Bandiera 2011	81	15	6	81	0.93 [0.86, 0.97]	0.84 [0.76, 0.91]		
Chan 2013	46	7	3	46	0.94 [0.83, 0.99]	0.87 [0.75, 0.95]		
Chen 2014	63	3	6	12	0.91 [0.82, 0.97]	0.80 [0.52, 0.96]		
Chudecka-Glaz 2015	114	4	10	33	0.92 [0.86, 0.96]	0.89 [0.75, 0.97]		
Farzaneh 2014	18	0	4	9	0.82 [0.60, 0.95]	1.00 [0.66, 1.00]		
Huy 2018	14	1	3	29	0.82 [0.57, 0.96]	0.97 [0.83, 1.00]		
Karlsen 2012	198	120	5	159	0.98 [0.94, 0.99]	0.57 [0.51, 0.63]		
Li 2016	70	5	12	85	0.85 [0.76, 0.92]	0.94 [0.88, 0.98]		
Liest 2019	91	76	26	230	0.78 [0.69, 0.85]	0.75 [0.70, 0.80]		
Melo 2018	21	8	7	56	0.75 [0.55, 0.89]	0.88 [0.77, 0.94]		
Novotny 2012	20	28	1	207	0.95 [0.76, 1.00]	0.88 [0.83, 0.92]		
Park 2019	11	9	8	90	0.58 [0.33, 0.80]	0.91 [0.83, 0.96]		
Partheen 2011a	81	41	12	124	0.87 [0.79, 0.93]	0.75 [0.68, 0.82]		
Richards 2015	9	1	4	15	0.69 [0.39, 0.91]	0.94 [0.70, 1.00]		
Shen 2017	91	9	14	47	0.87 [0.79, 0.93]	0.84 [0.72, 0.92]		



Test 3. ROMA 12.5 premenopausal

ROMA 12.5 premenopausal

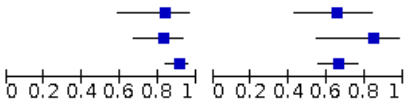
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Kadija 2012	7	2	4	54	0.64 [0.31, 0.89]	0.96 [0.88, 1.00]		
Montagnana 2011	8	7	7	29	0.53 [0.27, 0.79]	0.81 [0.64, 0.92]		
van Gorp 2011	28	17	14	125	0.67 [0.50, 0.80]	0.88 [0.82, 0.93]		



Test 4. ROMA 14.4 postmenopausal

ROMA 14.4 postmenopausal

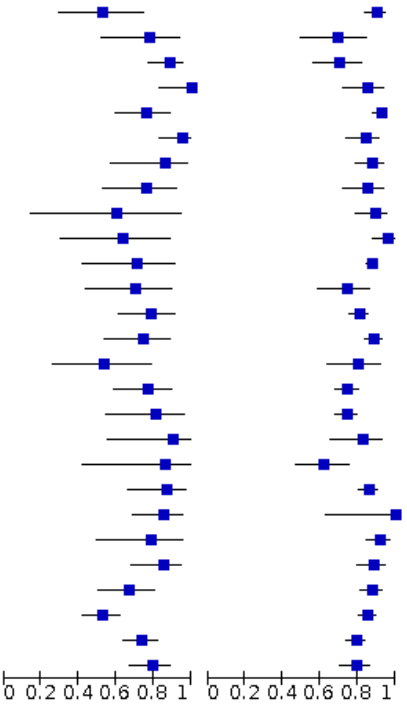
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Kadija 2012	15	8	3	15	0.83 [0.59, 0.96]	0.65 [0.43, 0.84]
Montagnana 2011	33	2	7	11	0.82 [0.67, 0.93]	0.85 [0.55, 0.98]
van Gorp 2011	108	29	11	57	0.91 [0.84, 0.95]	0.66 [0.55, 0.76]



Test 5. ROMA 13.1 (± 2) premenopausal

ROMA 13.1 (± 2) premenopausal

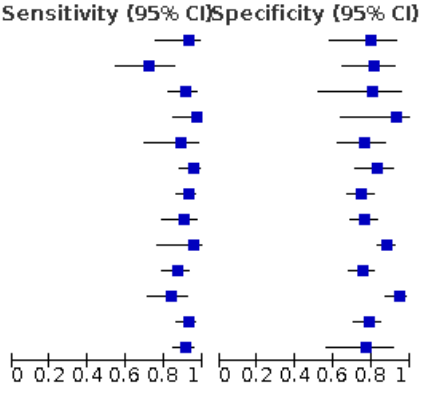
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Al Musalhi 2016	11	14	10	127	0.52 [0.30, 0.74]	0.90 [0.84, 0.94]
Anton 2012	14	9	4	20	0.78 [0.52, 0.94]	0.69 [0.49, 0.85]
Chen 2014	48	16	6	38	0.89 [0.77, 0.96]	0.70 [0.56, 0.82]
Chen 2015	20	7	0	41	1.00 [0.83, 1.00]	0.85 [0.72, 0.94]
Chudecka-Glaz 2015	29	16	9	198	0.76 [0.60, 0.89]	0.93 [0.88, 0.96]
Cradic 2018	37	12	2	63	0.95 [0.83, 0.99]	0.84 [0.74, 0.91]
Dikmen 2015	12	10	2	73	0.86 [0.57, 0.98]	0.88 [0.79, 0.94]
Farzaneh 2014	16	7	5	40	0.76 [0.53, 0.92]	0.85 [0.72, 0.94]
Grenache 2015	3	7	2	58	0.60 [0.15, 0.95]	0.89 [0.79, 0.96]
Kadija 2012	7	2	4	54	0.64 [0.31, 0.89]	0.96 [0.88, 1.00]
Kim 2019	10	71	4	496	0.71 [0.42, 0.92]	0.87 [0.84, 0.90]
Krascsenitis 2016	12	11	5	32	0.71 [0.44, 0.90]	0.74 [0.59, 0.86]
Lycke 2018	26	44	7	186	0.79 [0.61, 0.91]	0.81 [0.75, 0.86]
Molina 2011	20	25	7	201	0.74 [0.54, 0.89]	0.89 [0.84, 0.93]
Montagnana 2011	8	7	7	29	0.53 [0.27, 0.79]	0.81 [0.64, 0.92]
Moore 2009	26	51	8	151	0.76 [0.59, 0.89]	0.75 [0.68, 0.81]
Moore 2011	13	60	3	173	0.81 [0.54, 0.96]	0.74 [0.68, 0.80]
Ortiz-Munoz 2014	9	6	1	28	0.90 [0.55, 1.00]	0.82 [0.65, 0.93]
Prskalo 2015	6	19	1	31	0.86 [0.42, 1.00]	0.62 [0.47, 0.75]
Romagnolo 2016	20	30	3	186	0.87 [0.66, 0.97]	0.86 [0.81, 0.90]
Stiekma 2014	29	0	5	8	0.85 [0.69, 0.95]	1.00 [0.63, 1.00]
Teh 2018	11	7	3	81	0.79 [0.49, 0.95]	0.92 [0.84, 0.97]
Terlikowska 2016	28	10	5	77	0.85 [0.68, 0.95]	0.89 [0.80, 0.94]
van Gorp 2011	28	17	14	125	0.67 [0.50, 0.80]	0.88 [0.82, 0.93]
Xu 2016	56	38	51	226	0.52 [0.42, 0.62]	0.86 [0.81, 0.90]
Zhang 2015	70	59	25	226	0.74 [0.64, 0.82]	0.79 [0.74, 0.84]
Zhang 2019	50	24	13	91	0.79 [0.67, 0.89]	0.79 [0.71, 0.86]



Test 6. ROMA 27.7 (± 2) postmenopausal

ROMA 27.7 (± 2) postmenopausal

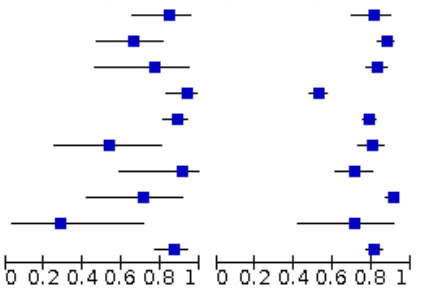
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Al Musalhi 2016	25	5	2	19	0.93 [0.76, 0.99]	0.79 [0.58, 0.93]
Anton 2012	26	7	10	30	0.72 [0.55, 0.86]	0.81 [0.65, 0.92]
Chen 2014	63	3	6	12	0.91 [0.82, 0.97]	0.80 [0.52, 0.96]
Dikmen 2015	32	1	1	12	0.97 [0.84, 1.00]	0.92 [0.64, 1.00]
Grenache 2015	23	12	3	38	0.88 [0.70, 0.98]	0.76 [0.62, 0.87]
Molina 2011	80	10	4	49	0.95 [0.88, 0.99]	0.83 [0.71, 0.92]
Moore 2009	108	38	9	112	0.92 [0.86, 0.96]	0.75 [0.67, 0.81]
Moore 2011	46	36	5	114	0.90 [0.79, 0.97]	0.76 [0.68, 0.83]
Novotny 2012	20	28	1	207	0.95 [0.76, 1.00]	0.88 [0.83, 0.92]
Partheen 2011a	81	41	12	124	0.87 [0.79, 0.93]	0.75 [0.68, 0.82]
Romagnolo 2016	50	5	10	83	0.83 [0.71, 0.92]	0.94 [0.87, 0.98]
Salim 2018	113	30	9	108	0.93 [0.86, 0.97]	0.78 [0.70, 0.85]
Stiekma 2014	103	6	10	20	0.91 [0.84, 0.96]	0.77 [0.56, 0.91]



Test 7. ROMA 7.4 premenopausal

ROMA 7.4 premenopausal

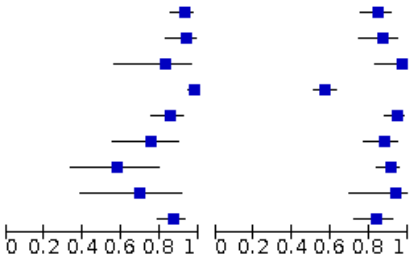
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Bandiera 2011	22	13	4	56	0.85 [0.65, 0.96]	0.81 [0.70, 0.90]		
Chan 2013	21	34	11	235	0.66 [0.47, 0.81]	0.87 [0.83, 0.91]		
Huy 2018	10	37	3	180	0.77 [0.46, 0.95]	0.83 [0.77, 0.88]		
Karlsen 2012	46	251	3	279	0.94 [0.83, 0.99]	0.53 [0.48, 0.57]		
Li 2016	96	136	12	501	0.89 [0.81, 0.94]	0.79 [0.75, 0.82]		
Melo 2018	7	28	6	114	0.54 [0.25, 0.81]	0.80 [0.73, 0.86]		
Nikolova 2016	10	27	1	67	0.91 [0.59, 1.00]	0.71 [0.61, 0.80]		
Park 2019	10	29	4	282	0.71 [0.42, 0.92]	0.91 [0.87, 0.94]		
Richards 2015	2	4	5	10	0.29 [0.04, 0.71]	0.71 [0.42, 0.92]		
Shen 2017	60	79	9	347	0.87 [0.77, 0.94]	0.81 [0.77, 0.85]		



Test 8. ROMA 25.3 postmenopausal

ROMA 25.3 postmenopausal

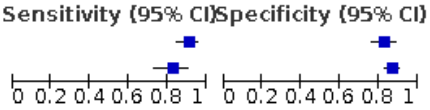
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Bandiera 2011	81	15	6	81	0.93 [0.86, 0.97]	0.84 [0.76, 0.91]		
Chan 2013	46	7	3	46	0.94 [0.83, 0.99]	0.87 [0.75, 0.95]		
Huy 2018	14	1	3	29	0.82 [0.57, 0.96]	0.97 [0.83, 1.00]		
Karlsen 2012	198	120	5	159	0.98 [0.94, 0.99]	0.57 [0.51, 0.63]		
Li 2016	70	5	12	85	0.85 [0.76, 0.92]	0.94 [0.88, 0.98]		
Melo 2018	21	8	7	56	0.75 [0.55, 0.89]	0.88 [0.77, 0.94]		
Park 2019	11	9	8	90	0.58 [0.33, 0.80]	0.91 [0.83, 0.96]		
Richards 2015	9	1	4	15	0.69 [0.39, 0.91]	0.94 [0.70, 1.00]		
Shen 2017	91	9	14	47	0.87 [0.79, 0.93]	0.84 [0.72, 0.92]		



Test 9. ROMA 7.4/25.3 all

ROMA 7.4/25.3 all

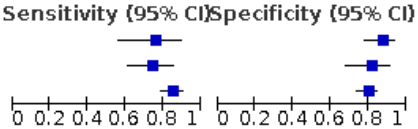
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Bandiera 2011	103	28	10	137	0.91 [0.84, 0.96]	0.83 [0.76, 0.88]		
Chan 2013	67	41	14	281	0.83 [0.73, 0.90]	0.87 [0.83, 0.91]		



Test 10. ROMA 12.5/14.4 all

ROMA 12.5/14.4 all

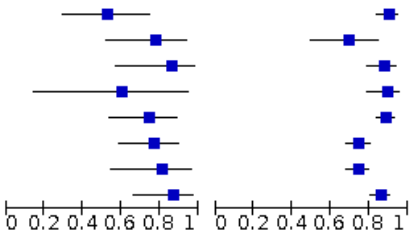
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Kadija 2012	22	10	7	69	0.76 [0.56, 0.90]	0.87 [0.78, 0.94]		
Montagnana 2011	41	9	14	40	0.75 [0.61, 0.85]	0.82 [0.68, 0.91]		
van Gorp 2011	137	46	24	182	0.85 [0.79, 0.90]	0.80 [0.74, 0.85]		



Test 11. ROMA 13.1 premenopausal

ROMA 13.1 premenopausal

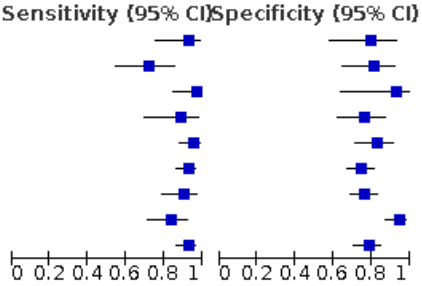
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Al Musalhi 2016	11	14	10	127	0.52 [0.30, 0.74]	0.90 [0.84, 0.94]
Anton 2012	14	9	4	20	0.78 [0.52, 0.94]	0.69 [0.49, 0.85]
Dikmen 2015	12	10	2	73	0.86 [0.57, 0.98]	0.88 [0.79, 0.94]
Grenache 2015	3	7	2	58	0.60 [0.15, 0.95]	0.89 [0.79, 0.96]
Molina 2011	20	25	7	201	0.74 [0.54, 0.89]	0.89 [0.84, 0.93]
Moore 2009	26	51	8	151	0.76 [0.59, 0.89]	0.75 [0.68, 0.81]
Moore 2011	13	60	3	173	0.81 [0.54, 0.96]	0.74 [0.68, 0.80]
Romagnolo 2016	20	30	3	186	0.87 [0.66, 0.97]	0.86 [0.81, 0.90]



Test 12. ROMA 27.7 postmenopausal

ROMA 27.7 postmenopausal

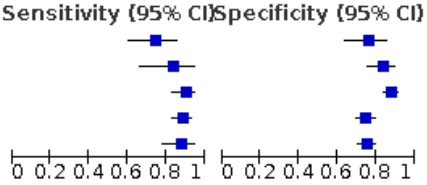
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Al Musalhi 2016	25	5	2	19	0.93 [0.76, 0.99]	0.79 [0.58, 0.93]
Anton 2012	26	7	10	30	0.72 [0.55, 0.86]	0.81 [0.65, 0.92]
Dikmen 2015	32	1	1	12	0.97 [0.84, 1.00]	0.92 [0.64, 1.00]
Grenache 2015	23	12	3	38	0.88 [0.70, 0.98]	0.76 [0.62, 0.87]
Molina 2011	80	10	4	49	0.95 [0.88, 0.99]	0.83 [0.71, 0.92]
Moore 2009	108	38	9	112	0.92 [0.86, 0.96]	0.75 [0.67, 0.81]
Moore 2011	46	36	5	114	0.90 [0.79, 0.97]	0.76 [0.68, 0.83]
Romagnolo 2016	50	5	10	83	0.83 [0.71, 0.92]	0.94 [0.87, 0.98]
Salim 2018	113	30	9	108	0.93 [0.86, 0.97]	0.78 [0.70, 0.85]



Test 13. ROMA 13.1/27.7 all

ROMA 13.1/27.7 all

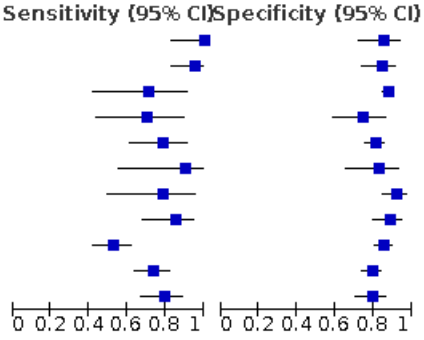
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Anton 2012	40	16	14	50	0.74 [0.60, 0.85]	0.76 [0.64, 0.85]
Grenache 2015	26	19	5	96	0.84 [0.66, 0.95]	0.83 [0.75, 0.90]
Molina 2011	100	35	11	250	0.90 [0.83, 0.95]	0.88 [0.83, 0.91]
Moore 2009	134	89	17	263	0.89 [0.83, 0.93]	0.75 [0.70, 0.79]
Moore 2011	59	96	8	287	0.88 [0.78, 0.95]	0.75 [0.70, 0.79]



Test 14. ROMA 11.4 premenopausal

ROMA 11.4 premenopausal

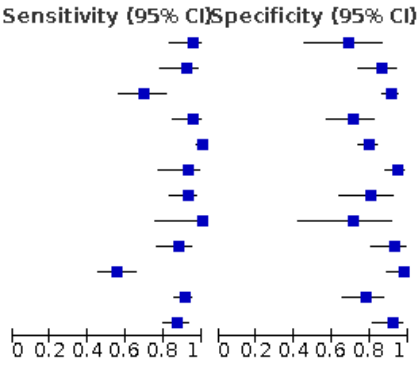
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Chen 2015	20	7	0	41	1.00 [0.83, 1.00]	0.85 [0.72, 0.94]
Cradic 2018	37	12	2	63	0.95 [0.83, 0.99]	0.84 [0.74, 0.91]
Kim 2019	10	71	4	496	0.71 [0.42, 0.92]	0.87 [0.84, 0.90]
Krascsenitis 2016	12	11	5	32	0.71 [0.44, 0.90]	0.74 [0.59, 0.86]
Lycke 2018	26	44	7	186	0.79 [0.61, 0.91]	0.81 [0.75, 0.86]
Ortiz-Munoz 2014	9	6	1	28	0.90 [0.55, 1.00]	0.82 [0.65, 0.93]
Teh 2018	11	7	3	81	0.79 [0.49, 0.95]	0.92 [0.84, 0.97]
Terlikowska 2016	28	10	5	77	0.85 [0.68, 0.95]	0.89 [0.80, 0.94]
Xu 2016	56	38	51	226	0.52 [0.42, 0.62]	0.86 [0.81, 0.90]
Zhang 2015	70	59	25	226	0.74 [0.64, 0.82]	0.79 [0.74, 0.84]
Zhang 2019	50	24	13	91	0.79 [0.67, 0.89]	0.79 [0.71, 0.86]



Test 15. ROMA 29.9 postmenopausal

ROMA 29.9 postmenopausal

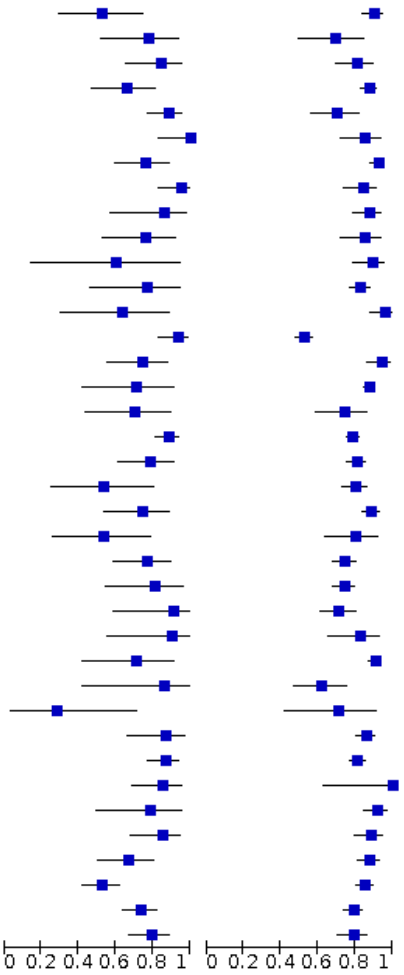
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Chen 2015	38	7	2	15	0.95 [0.83, 0.99]	0.68 [0.45, 0.86]
Cradic 2018	34	8	3	48	0.92 [0.78, 0.98]	0.86 [0.74, 0.94]
Kim 2019	39	17	17	178	0.70 [0.56, 0.81]	0.91 [0.86, 0.95]
Krascsenitis 2016	42	17	2	41	0.95 [0.85, 0.99]	0.71 [0.57, 0.82]
Lycke 2018	113	49	0	186	1.00 [0.97, 1.00]	0.79 [0.73, 0.84]
Ortiz-Munoz 2014	27	7	2	112	0.93 [0.77, 0.99]	0.94 [0.88, 0.98]
Prskalo 2015	61	7	5	29	0.92 [0.83, 0.97]	0.81 [0.64, 0.92]
Teh 2018	13	4	0	10	1.00 [0.75, 1.00]	0.71 [0.42, 0.92]
Terlikowska 2016	55	3	8	38	0.87 [0.77, 0.94]	0.93 [0.80, 0.98]
Xu 2016	57	1	46	46	0.55 [0.45, 0.65]	0.98 [0.89, 1.00]
Zhang 2015	154	14	15	49	0.91 [0.86, 0.95]	0.78 [0.66, 0.87]
Zhang 2019	103	5	15	55	0.87 [0.80, 0.93]	0.92 [0.82, 0.97]



Test 18. ROMA mixed premenopausal

ROMA mixed premenopausal

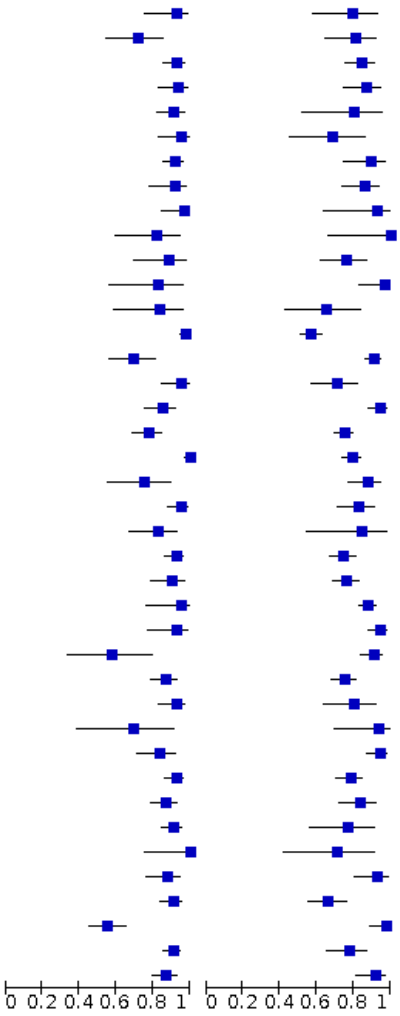
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Al Musalhi 2016	11	14	10	127	0.52 [0.30, 0.74]	0.90 [0.84, 0.94]
Anton 2012	14	9	4	20	0.78 [0.52, 0.94]	0.69 [0.49, 0.85]
Bandiera 2011	22	13	4	56	0.85 [0.65, 0.96]	0.81 [0.70, 0.90]
Chan 2013	21	34	11	235	0.66 [0.47, 0.81]	0.87 [0.83, 0.91]
Chen 2014	48	16	6	38	0.89 [0.77, 0.96]	0.70 [0.56, 0.82]
Chen 2015	20	7	0	41	1.00 [0.83, 1.00]	0.85 [0.72, 0.94]
Chudecka-Glaz 2015	29	16	9	198	0.76 [0.60, 0.89]	0.93 [0.88, 0.96]
Cradic 2018	37	12	2	63	0.95 [0.83, 0.99]	0.84 [0.74, 0.91]
Dikmen 2015	12	10	2	73	0.86 [0.57, 0.98]	0.88 [0.79, 0.94]
Farzaneh 2014	16	7	5	40	0.76 [0.53, 0.92]	0.85 [0.72, 0.94]
Grenache 2015	3	7	2	58	0.60 [0.15, 0.95]	0.89 [0.79, 0.96]
Huy 2018	10	37	3	180	0.77 [0.46, 0.95]	0.83 [0.77, 0.88]
Kadija 2012	7	2	4	54	0.64 [0.31, 0.89]	0.96 [0.88, 1.00]
Karlsen 2012	46	251	3	279	0.94 [0.83, 0.99]	0.53 [0.48, 0.57]
Kim 2011	23	4	8	67	0.74 [0.55, 0.88]	0.94 [0.86, 0.98]
Kim 2019	10	71	4	496	0.71 [0.42, 0.92]	0.87 [0.84, 0.90]
Krascsenitis 2016	12	11	5	32	0.71 [0.44, 0.90]	0.74 [0.59, 0.86]
Li 2016	96	136	12	501	0.89 [0.81, 0.94]	0.79 [0.75, 0.82]
Lycke 2018	26	44	7	186	0.79 [0.61, 0.91]	0.81 [0.75, 0.86]
Melo 2018	7	28	6	114	0.54 [0.25, 0.81]	0.80 [0.73, 0.86]
Molina 2011	20	25	7	201	0.74 [0.54, 0.89]	0.89 [0.84, 0.93]
Montagnana 2011	8	7	7	29	0.53 [0.27, 0.79]	0.81 [0.64, 0.92]
Moore 2009	26	51	8	151	0.76 [0.59, 0.89]	0.75 [0.68, 0.81]
Moore 2011	13	60	3	173	0.81 [0.54, 0.96]	0.74 [0.68, 0.80]
Nikolova 2016	10	27	1	67	0.91 [0.59, 1.00]	0.71 [0.61, 0.80]
Ortiz-Munoz 2014	9	6	1	28	0.90 [0.55, 1.00]	0.82 [0.65, 0.93]
Park 2019	10	29	4	282	0.71 [0.42, 0.92]	0.91 [0.87, 0.94]
Prskalo 2015	6	19	1	31	0.86 [0.42, 1.00]	0.62 [0.47, 0.75]
Richards 2015	2	4	5	10	0.29 [0.04, 0.71]	0.71 [0.42, 0.92]
Romagnolo 2016	20	30	3	186	0.87 [0.66, 0.97]	0.86 [0.81, 0.90]
Shen 2017	60	79	9	347	0.87 [0.77, 0.94]	0.81 [0.77, 0.85]
Stiekma 2014	29	0	5	8	0.85 [0.69, 0.95]	1.00 [0.63, 1.00]
Teh 2018	11	7	3	81	0.79 [0.49, 0.95]	0.92 [0.84, 0.97]
Terlikowska 2016	28	10	5	77	0.85 [0.68, 0.95]	0.89 [0.80, 0.94]
van Gorp 2011	28	17	14	125	0.67 [0.50, 0.80]	0.88 [0.82, 0.93]
Xu 2016	56	38	51	226	0.52 [0.42, 0.62]	0.86 [0.81, 0.90]
Zhang 2015	70	59	25	226	0.74 [0.64, 0.82]	0.79 [0.74, 0.84]
Zhang 2019	50	24	13	91	0.79 [0.67, 0.89]	0.79 [0.71, 0.86]



Test 19. ROMA mixed postmenopausal

ROMA mixed postmenopausal

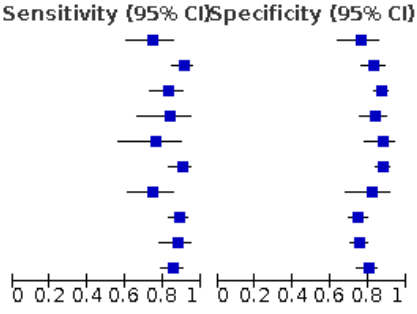
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Al Musalhi 2016	25	5	2	19	0.93 [0.76, 0.99]	0.79 [0.58, 0.93]
Anton 2012	26	7	10	30	0.72 [0.55, 0.86]	0.81 [0.65, 0.92]
Bandiera 2011	81	15	6	81	0.93 [0.86, 0.97]	0.84 [0.76, 0.91]
Chan 2013	46	7	3	46	0.94 [0.83, 0.99]	0.87 [0.75, 0.95]
Chen 2014	63	3	6	12	0.91 [0.82, 0.97]	0.80 [0.52, 0.96]
Chen 2015	38	7	2	15	0.95 [0.83, 0.99]	0.68 [0.45, 0.86]
Chudecka-Glaz 2015	114	4	10	33	0.92 [0.86, 0.96]	0.89 [0.75, 0.97]
Cradic 2018	34	8	3	48	0.92 [0.78, 0.98]	0.86 [0.74, 0.94]
Dikmen 2015	32	1	1	12	0.97 [0.84, 1.00]	0.92 [0.64, 1.00]
Farzaneh 2014	18	0	4	9	0.82 [0.60, 0.95]	1.00 [0.66, 1.00]
Grenache 2015	23	12	3	38	0.88 [0.70, 0.98]	0.76 [0.62, 0.87]
Huy 2018	14	1	3	29	0.82 [0.57, 0.96]	0.97 [0.83, 1.00]
Kadija 2012	15	8	3	15	0.83 [0.59, 0.96]	0.65 [0.43, 0.84]
Karlsen 2012	198	120	5	159	0.98 [0.94, 0.99]	0.57 [0.51, 0.63]
Kim 2019	39	17	17	178	0.70 [0.56, 0.81]	0.91 [0.86, 0.95]
Krascsenitis 2016	42	17	2	41	0.95 [0.85, 0.99]	0.71 [0.57, 0.82]
Li 2016	70	5	12	85	0.85 [0.76, 0.92]	0.94 [0.88, 0.98]
Liest 2019	91	76	26	230	0.78 [0.69, 0.85]	0.75 [0.70, 0.80]
Lycke 2018	113	49	0	186	1.00 [0.97, 1.00]	0.79 [0.73, 0.84]
Melo 2018	21	8	7	56	0.75 [0.55, 0.89]	0.88 [0.77, 0.94]
Molina 2011	80	10	4	49	0.95 [0.88, 0.99]	0.83 [0.71, 0.92]
Montagnana 2011	33	2	7	11	0.82 [0.67, 0.93]	0.85 [0.55, 0.98]
Moore 2009	108	38	9	112	0.92 [0.86, 0.96]	0.75 [0.67, 0.81]
Moore 2011	46	36	5	114	0.90 [0.79, 0.97]	0.76 [0.68, 0.83]
Novotny 2012	20	28	1	207	0.95 [0.76, 1.00]	0.88 [0.83, 0.92]
Ortiz-Munoz 2014	27	7	2	112	0.93 [0.77, 0.99]	0.94 [0.88, 0.98]
Park 2019	11	9	8	90	0.58 [0.33, 0.80]	0.91 [0.83, 0.96]
Partheen 2011a	81	41	12	124	0.87 [0.79, 0.93]	0.75 [0.68, 0.82]
Prskalo 2015	61	7	5	29	0.92 [0.83, 0.97]	0.81 [0.64, 0.92]
Richards 2015	9	1	4	15	0.69 [0.39, 0.91]	0.94 [0.70, 1.00]
Romagnolo 2016	50	5	10	83	0.83 [0.71, 0.92]	0.94 [0.87, 0.98]
Salim 2018	113	30	9	108	0.93 [0.86, 0.97]	0.78 [0.70, 0.85]
Shen 2017	91	9	14	47	0.87 [0.79, 0.93]	0.84 [0.72, 0.92]
Stiekma 2014	103	6	10	20	0.91 [0.84, 0.96]	0.77 [0.56, 0.91]
Teh 2018	13	4	0	10	1.00 [0.75, 1.00]	0.71 [0.42, 0.92]
Terlikowska 2016	55	3	8	38	0.87 [0.77, 0.94]	0.93 [0.80, 0.98]
van Gorp 2011	108	29	11	57	0.91 [0.84, 0.95]	0.66 [0.55, 0.76]
Xu 2016	57	1	46	46	0.55 [0.45, 0.65]	0.98 [0.89, 1.00]
Zhang 2015	154	14	15	49	0.91 [0.86, 0.95]	0.78 [0.66, 0.87]
Zhang 2019	103	5	15	55	0.87 [0.80, 0.93]	0.92 [0.82, 0.97]



Test 20. ROMA mixed all

ROMA mixed all

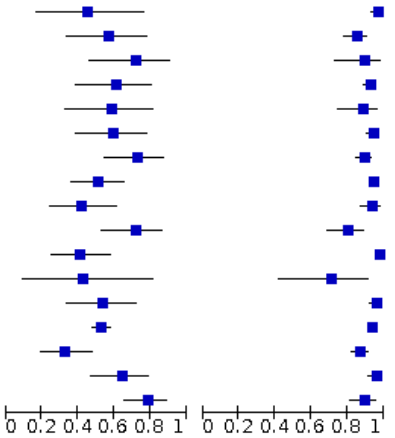
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Anton 2012	40	16	14	50	0.74 [0.60, 0.85]	0.76 [0.64, 0.85]
Bandiera 2011	103	28	10	137	0.91 [0.84, 0.96]	0.83 [0.76, 0.88]
Chan 2013	67	41	14	281	0.83 [0.73, 0.90]	0.87 [0.83, 0.91]
Grenache 2015	26	19	5	96	0.84 [0.66, 0.95]	0.83 [0.75, 0.90]
Kadija 2012	22	10	7	69	0.76 [0.56, 0.90]	0.87 [0.78, 0.94]
Molina 2011	100	35	11	250	0.90 [0.83, 0.95]	0.88 [0.83, 0.91]
Montagnana 2011	41	9	14	40	0.75 [0.61, 0.85]	0.82 [0.68, 0.91]
Moore 2009	134	89	17	263	0.89 [0.83, 0.93]	0.75 [0.70, 0.79]
Moore 2011	59	96	8	287	0.88 [0.78, 0.95]	0.75 [0.70, 0.79]
van Gorp 2011	137	46	24	182	0.85 [0.79, 0.90]	0.80 [0.74, 0.85]



Test 21. RMI I 200 premenopausal

RMI I 200 premenopausal

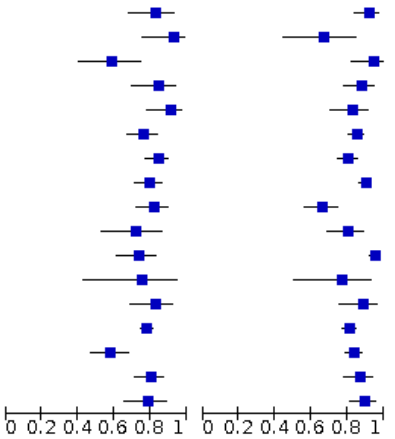
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Abdalla 2017	5	6	6	178	0.45 [0.17, 0.77]	0.97 [0.93, 0.99]		
Al Musalhi 2016	12	21	9	120	0.57 [0.34, 0.78]	0.85 [0.78, 0.91]		
Anton 2012	13	3	5	26	0.72 [0.47, 0.90]	0.90 [0.73, 0.98]		
Ertas 2016	14	20	9	248	0.61 [0.39, 0.80]	0.93 [0.89, 0.95]		
Krascsenitis 2016	10	5	7	38	0.59 [0.33, 0.82]	0.88 [0.75, 0.96]		
Liest 2019	16	18	11	287	0.59 [0.39, 0.78]	0.94 [0.91, 0.96]		
Lycke 2018	24	24	9	206	0.73 [0.54, 0.87]	0.90 [0.85, 0.93]		
Manegold-Brauer 2016	25	35	24	546	0.51 [0.36, 0.66]	0.94 [0.92, 0.96]		
Meys 2017	13	6	18	91	0.42 [0.25, 0.61]	0.94 [0.87, 0.98]		
Niemi 2017	23	13	9	53	0.72 [0.53, 0.86]	0.80 [0.69, 0.89]		
Radosa 2011	16	19	23	832	0.41 [0.26, 0.58]	0.98 [0.97, 0.99]		
Richards 2015	3	4	4	10	0.43 [0.10, 0.82]	0.71 [0.42, 0.92]		
Sayasneh 2013a	15	5	13	132	0.54 [0.34, 0.72]	0.96 [0.92, 0.99]		
Testa 2014	200	59	178	917	0.53 [0.48, 0.58]	0.94 [0.92, 0.95]		
van den Akker 2016	15	30	31	204	0.33 [0.20, 0.48]	0.87 [0.82, 0.91]		
van Gorp 2012	25	6	14	133	0.64 [0.47, 0.79]	0.96 [0.91, 0.98]		
Vural 2016	41	9	11	78	0.79 [0.65, 0.89]	0.90 [0.81, 0.95]		



Test 22. RMI I 200 postmenopausal

RMI I 200 postmenopausal

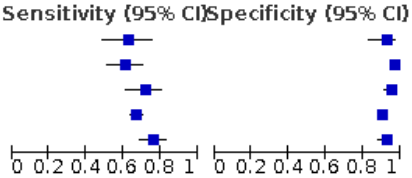
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Abdalla 2017	34	6	7	70	0.83 [0.68, 0.93]	0.92 [0.84, 0.97]		
Al Musalhi 2016	25	8	2	16	0.93 [0.76, 0.99]	0.67 [0.45, 0.84]		
Anton 2012	21	2	15	35	0.58 [0.41, 0.74]	0.95 [0.82, 0.99]		
Ertas 2016	37	9	7	64	0.84 [0.70, 0.93]	0.88 [0.78, 0.94]		
Krascsenitis 2016	40	10	4	48	0.91 [0.78, 0.97]	0.83 [0.71, 0.91]		
Liest 2019	89	46	28	260	0.76 [0.67, 0.83]	0.85 [0.80, 0.89]		
Lycke 2018	112	42	21	173	0.84 [0.77, 0.90]	0.80 [0.75, 0.86]		
Manegold-Brauer 2016	98	35	25	320	0.80 [0.71, 0.86]	0.90 [0.87, 0.93]		
Meys 2017	69	39	15	75	0.82 [0.72, 0.90]	0.66 [0.56, 0.74]		
Niemi 2017	23	13	9	53	0.72 [0.53, 0.86]	0.80 [0.69, 0.89]		
Radosa 2011	49	21	18	383	0.73 [0.61, 0.83]	0.95 [0.92, 0.97]		
Richards 2015	9	4	3	13	0.75 [0.43, 0.95]	0.76 [0.50, 0.93]		
Sayasneh 2013a	38	5	8	39	0.83 [0.69, 0.92]	0.89 [0.75, 0.96]		
Testa 2014	470	85	132	362	0.78 [0.75, 0.81]	0.81 [0.77, 0.85]		
van den Akker 2016	54	48	39	249	0.58 [0.47, 0.68]	0.84 [0.79, 0.88]		
van Gorp 2012	89	11	22	74	0.80 [0.72, 0.87]	0.87 [0.78, 0.93]		
Vural 2016	41	9	11	78	0.79 [0.65, 0.89]	0.90 [0.81, 0.95]		



Test 23. RMI I 200 all

RMI I 200 all

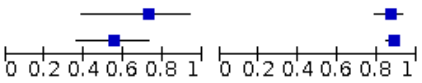
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Anton 2012	34	5	20	61	0.63 [0.49, 0.76]	0.92 [0.83, 0.97]		
Radosa 2011	65	40	41	1215	0.61 [0.51, 0.71]	0.97 [0.96, 0.98]		
Sayasneh 2013a	66	10	26	199	0.72 [0.61, 0.81]	0.95 [0.91, 0.98]		
Testa 2014	658	134	322	1289	0.67 [0.64, 0.70]	0.91 [0.89, 0.92]		
van Gorp 2012	114	17	36	207	0.76 [0.68, 0.83]	0.92 [0.88, 0.96]		



Test 24. RMI I 250 premenopausal

RMI I 250 premenopausal

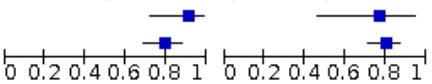
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Nikolova 2016	8	12	3	82	0.73 [0.39, 0.94]	0.87 [0.79, 0.93]		
Terzic 2013	17	38	14	287	0.55 [0.36, 0.73]	0.88 [0.84, 0.92]		



Test 25. RMI I 250 postmenopausal

RMI I 250 postmenopausal

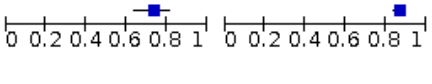
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Irshad 2013	21	3	2	10	0.91 [0.72, 0.99]	0.77 [0.46, 0.95]		
Terzic 2013	59	22	15	88	0.80 [0.69, 0.88]	0.80 [0.71, 0.87]		



Test 26. RMI I 250 all

RMI I 250 all

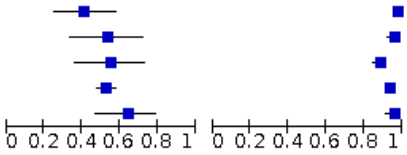
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Terzic 2013	77	57	28	378	0.73 [0.64, 0.81]	0.87 [0.83, 0.90]		



Test 35. RMI mixed premenopausal

RMI mixed premenopausal

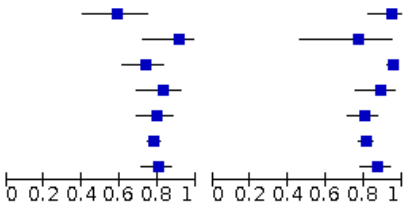
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Anton 2012	13	3	5	26	0.72 [0.47, 0.90]	0.90 [0.73, 0.98]		
Radosa 2011	16	19	23	832	0.41 [0.26, 0.58]	0.98 [0.97, 0.99]		
Sayasneh 2013a	15	5	13	132	0.54 [0.34, 0.72]	0.96 [0.92, 0.99]		
Terzic 2013	17	38	14	287	0.55 [0.36, 0.73]	0.88 [0.84, 0.92]		
Testa 2014	200	59	178	917	0.53 [0.48, 0.58]	0.94 [0.92, 0.95]		
van Gorp 2012	25	6	14	133	0.64 [0.47, 0.79]	0.96 [0.91, 0.98]		



Test 36. RMI mixed postmenopausal

RMI mixed postmenopausal

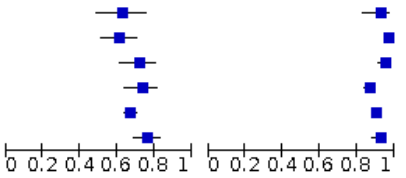
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Anton 2012	21	2	15	35	0.58 [0.41, 0.74]	0.95 [0.82, 0.99]		
Irshad 2013	21	3	2	10	0.91 [0.72, 0.99]	0.77 [0.46, 0.95]		
Radosa 2011	49	21	18	383	0.73 [0.61, 0.83]	0.95 [0.92, 0.97]		
Sayasneh 2013a	38	5	8	39	0.83 [0.69, 0.92]	0.89 [0.75, 0.96]		
Terzic 2013	59	22	15	88	0.80 [0.69, 0.88]	0.80 [0.71, 0.87]		
Testa 2014	470	85	132	362	0.78 [0.75, 0.81]	0.81 [0.77, 0.85]		
van Gorp 2012	89	11	22	74	0.80 [0.72, 0.87]	0.87 [0.78, 0.93]		



Test 37. RMI mixed all

RMI mixed all

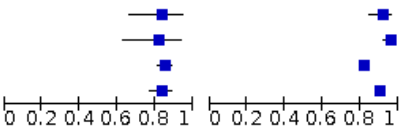
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Anton 2012	34	5	20	61	0.63 [0.49, 0.76]	0.92 [0.83, 0.97]		
Radosa 2011	65	40	41	1215	0.61 [0.51, 0.71]	0.97 [0.96, 0.98]		
Sayasneh 2013a	66	10	26	199	0.72 [0.61, 0.81]	0.95 [0.91, 0.98]		
Terzic 2013	77	57	28	378	0.73 [0.64, 0.81]	0.87 [0.83, 0.90]		
Testa 2014	658	134	322	1289	0.67 [0.64, 0.70]	0.91 [0.89, 0.92]		
van Gorp 2011	114	17	36	207	0.76 [0.68, 0.83]	0.92 [0.88, 0.96]		



Test 38. LR2 premenopausal

LR2 premenopausal

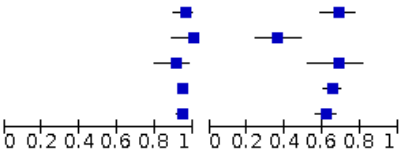
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Meys 2017	26	8	5	89	0.84 [0.66, 0.95]	0.92 [0.84, 0.96]		
Sayasneh 2013a	23	5	5	132	0.82 [0.63, 0.94]	0.96 [0.92, 0.99]		
Testa 2014	321	176	57	800	0.85 [0.81, 0.88]	0.82 [0.79, 0.84]		
Timmerman 2010	152	101	30	913	0.84 [0.77, 0.89]	0.90 [0.88, 0.92]		



Test 39. LR2 postmenopausal

LR2 postmenopausal

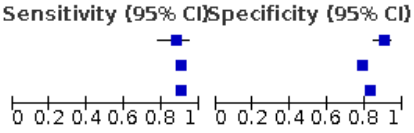
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Meys 2017	81	36	3	78	0.96 [0.90, 0.99]	0.68 [0.59, 0.77]		
Niemi 2017	32	42	0	24	1.00 [0.89, 1.00]	0.36 [0.25, 0.49]		
Sayasneh 2013a	42	14	4	30	0.91 [0.79, 0.98]	0.68 [0.52, 0.81]		
Testa 2014	566	156	36	291	0.94 [0.92, 0.96]	0.65 [0.60, 0.70]		
Timmerman 2010	339	138	21	224	0.94 [0.91, 0.96]	0.62 [0.57, 0.67]		



Test 40. LR2 all

LR2 all

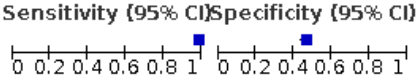
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Sayasneh 2013a	65	18	9	163	0.88 [0.78, 0.94]	0.90 [0.85, 0.94]		
Testa 2014	884	300	96	1123	0.90 [0.88, 0.92]	0.79 [0.77, 0.81]		
Timmerman 2010	490	239	52	1157	0.90 [0.88, 0.93]	0.83 [0.81, 0.85]		



Test 41. ADNEX 3% D+ probability all

ADNEX 3% D+ probability all

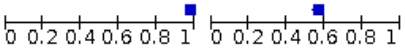
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
van Calster 2014	969	759	11	664	0.99 [0.98, 0.99]	0.47 [0.44, 0.49]		



Test 42. ADNEX 3% D+ probability premenopausal

ADNEX 3% D+ probability premenopausal

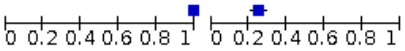
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
van Calster 2014	370	424	8	552	0.98 [0.96, 0.99]	0.57 [0.53, 0.60]		



Test 43. ADNEX 3% D+ probability postmenopausal

ADNEX 3% D+ probability postmenopausal

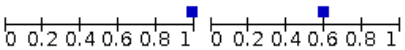
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
van Calster 2014	599	335	3	112	1.00 [0.99, 1.00]	0.25 [0.21, 0.29]		



Test 44. ADNEX 5% D+ probability all

ADNEX 5% D+ probability all

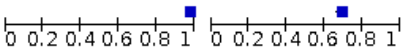
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
van Calster 2014	964	578	16	845	0.98 [0.97, 0.99]	0.59 [0.57, 0.62]		



Test 45. ADNEX 5% D+ probability premenopausal

ADNEX 5% D+ probability premenopausal

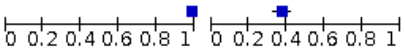
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
van Calster 2014	369	298	9	678	0.98 [0.96, 0.99]	0.69 [0.66, 0.72]		



Test 46. ADNEX 5% D+ probability postmenopausal

ADNEX 5% D+ probability postmenopausal

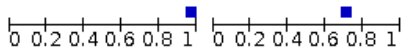
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
van Calster 2014	595	280	7	167	0.99 [0.98, 1.00]	0.37 [0.33, 0.42]		



Test 47. ADNEX 10% D+ probability all

ADNEX 10% D+ probability all

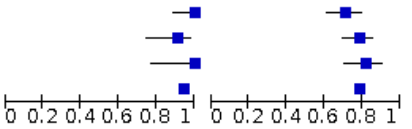
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
van Calster 2014	946	408	34	1015	0.97 [0.95, 0.98]	0.71 [0.69, 0.74]		



Test 48. ADNEX 10% D+ probability premenopausal

ADNEX 10% D+ probability premenopausal

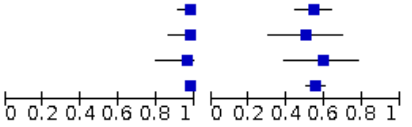
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Meys 2017	31	28	0	69	1.00 [0.89, 1.00]	0.71 [0.61, 0.80]		
Szubert 2016a	29	23	3	83	0.91 [0.75, 0.98]	0.78 [0.69, 0.86]		
Szubert 2016b	14	11	0	51	1.00 [0.77, 1.00]	0.82 [0.70, 0.91]		
van Calster 2014	358	209	20	767	0.95 [0.92, 0.97]	0.79 [0.76, 0.81]		



Test 49. ADNEX 10% D+ probability postmenopausal

ADNEX 10% D+ probability postmenopausal

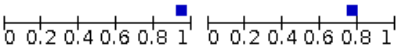
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Meys 2017	82	52	2	62	0.98 [0.92, 1.00]	0.54 [0.45, 0.64]		
Szubert 2016a	37	14	1	14	0.97 [0.86, 1.00]	0.50 [0.31, 0.69]		
Szubert 2016b	24	11	1	16	0.96 [0.80, 1.00]	0.59 [0.39, 0.78]		
van Calster 2014	588	199	14	248	0.98 [0.96, 0.99]	0.55 [0.51, 0.60]		



Test 50. ADNEX 15% D+ probability all

ADNEX 15% D+ probability all

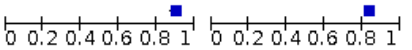
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
van Calster 2014	923	325	57	1098	0.94 [0.93, 0.96]	0.77 [0.75, 0.79]		



Test 51. ADNEX 15% D+ probability premenopausal

ADNEX 15% D+ probability premenopausal

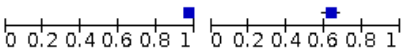
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
van Calster 2014	342	162	36	814	0.90 [0.87, 0.93]	0.83 [0.81, 0.86]		



Test 52. ADNEX 15% D+ probability postmenopausal

ADNEX 15% D+ probability postmenopausal

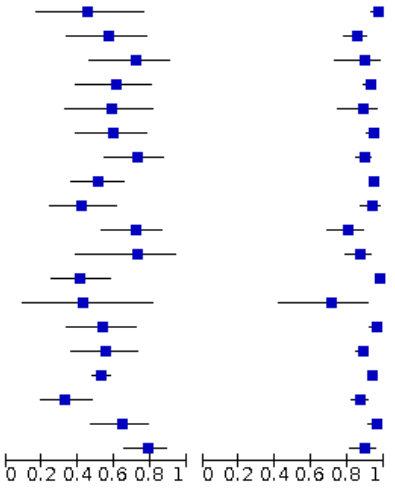
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
van Calster 2014	581	163	21	284	0.97 [0.95, 0.98]	0.64 [0.59, 0.68]		



Test 67. RMI I mixed premenopausal

RMI I mixed premenopausal

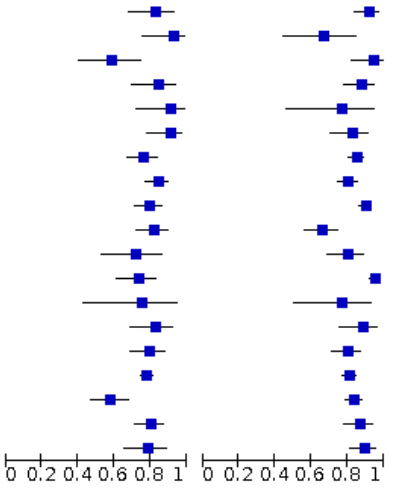
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abdalla 2017	5	6	6	178	0.45 [0.17, 0.77]	0.97 [0.93, 0.99]
Al Musalhi 2016	12	21	9	120	0.57 [0.34, 0.78]	0.85 [0.78, 0.91]
Anton 2012	13	3	5	26	0.72 [0.47, 0.90]	0.90 [0.73, 0.98]
Ertas 2016	14	20	9	248	0.61 [0.39, 0.80]	0.93 [0.89, 0.95]
Krascsenitis 2016	10	5	7	38	0.59 [0.33, 0.82]	0.88 [0.75, 0.96]
Liest 2019	16	18	11	287	0.59 [0.39, 0.78]	0.94 [0.91, 0.96]
Lycke 2018	24	24	9	206	0.73 [0.54, 0.87]	0.90 [0.85, 0.93]
Manegold-Brauer 2016	25	35	24	546	0.51 [0.36, 0.66]	0.94 [0.92, 0.96]
Meys 2017	13	6	18	91	0.42 [0.25, 0.61]	0.94 [0.87, 0.98]
Niemi 2017	23	13	9	53	0.72 [0.53, 0.86]	0.80 [0.69, 0.89]
Nikolova 2016	8	12	3	82	0.73 [0.39, 0.94]	0.87 [0.79, 0.93]
Radosa 2011	16	19	23	832	0.41 [0.26, 0.58]	0.98 [0.97, 0.99]
Richards 2015	3	4	4	10	0.43 [0.10, 0.82]	0.71 [0.42, 0.92]
Sayasneh 2013a	15	5	13	132	0.54 [0.34, 0.72]	0.96 [0.92, 0.99]
Terzic 2013	17	38	14	287	0.55 [0.36, 0.73]	0.88 [0.84, 0.92]
Testa 2014	200	59	178	917	0.53 [0.48, 0.58]	0.94 [0.92, 0.95]
van den Akker 2016	15	30	31	204	0.33 [0.20, 0.48]	0.87 [0.82, 0.91]
van Gorp 2012	25	6	14	133	0.64 [0.47, 0.79]	0.96 [0.91, 0.98]
Vural 2016	41	9	11	78	0.79 [0.65, 0.89]	0.90 [0.81, 0.95]



Test 68. RMI I mixed postmenopausal

RMI I mixed postmenopausal

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abdalla 2017	34	6	7	70	0.83 [0.68, 0.93]	0.92 [0.84, 0.97]
Al Musalhi 2016	25	8	2	16	0.93 [0.76, 0.99]	0.67 [0.45, 0.84]
Anton 2012	21	2	15	35	0.58 [0.41, 0.74]	0.95 [0.82, 0.99]
Ertas 2016	37	9	7	64	0.84 [0.70, 0.93]	0.88 [0.78, 0.94]
Irshad 2013	21	3	2	10	0.91 [0.72, 0.99]	0.77 [0.46, 0.95]
Krascsenitis 2016	40	10	4	48	0.91 [0.78, 0.97]	0.83 [0.71, 0.91]
Liest 2019	89	46	28	260	0.76 [0.67, 0.83]	0.85 [0.80, 0.89]
Lycke 2018	112	42	21	173	0.84 [0.77, 0.90]	0.80 [0.75, 0.86]
Manegold-Brauer 2016	98	35	25	320	0.80 [0.71, 0.86]	0.90 [0.87, 0.93]
Meys 2017	69	39	15	75	0.82 [0.72, 0.90]	0.66 [0.56, 0.74]
Niemi 2017	23	13	9	53	0.72 [0.53, 0.86]	0.80 [0.69, 0.89]
Radosa 2011	49	21	18	383	0.73 [0.61, 0.83]	0.95 [0.92, 0.97]
Richards 2015	9	4	3	13	0.75 [0.43, 0.95]	0.76 [0.50, 0.93]
Sayasneh 2013a	38	5	8	39	0.83 [0.69, 0.92]	0.89 [0.75, 0.96]
Terzic 2013	59	22	15	88	0.80 [0.69, 0.88]	0.80 [0.71, 0.87]
Testa 2014	470	85	132	362	0.78 [0.75, 0.81]	0.81 [0.77, 0.85]
van den Akker 2016	54	48	39	249	0.58 [0.47, 0.68]	0.84 [0.79, 0.88]
van Gorp 2012	89	11	22	74	0.80 [0.72, 0.87]	0.87 [0.78, 0.93]
Vural 2016	41	9	11	78	0.79 [0.65, 0.89]	0.90 [0.81, 0.95]



ADDITIONAL TABLES

Table 1. Details of included test combinations

Index test combination	Details	Test positivity thresholds included
RMI I *U* × *M* × CA125 Jacobs 1990	**Ultrasound (U):** (1 point for each of multilocular cysts, solid areas, metastases, ascites and bilateral lesions) where a total ultrasound point score of 0 = 0, a point score of 1 = 1, and a point score of ≥ 2 = 3 **Menopausal status (M):** premenopausal = 1 and postmenopausal = 3	200, 250

Table 1. Details of included test combinations *(Continued)*

Serum CA125: CA125 U/mL applied directly to the calculation

ROMA Bandiera 2011 Moore 2009 van Gorp 2011	**Premenopausal** PI = −12.0 + 2.38 × LN(HE4) + 0.0626 × LN(CA125) **Postmenopausal** PI = −8.09 + 1.04 × LN(HE4) + 0.732 × LN(CA125) **Predicted probability (ROMA score)** = exp(PI)/[1 + exp(PI)] × 100	Premenopausal 7.4 and postmenopausal 25.3 Premenopausal 12.5 and postmenopausal 14.4 Premenopausal 13.1 and postmenopausal 27.7 ± 2% from common (above) thresholds Premenopausal: 7.4 (5.4 to 9.4%), 12.5 (10.5 to 14.5%), 14.4 (12.4 to 16.4%) Postmenopausal: 25.3 (23.3 to 27.3%), 27.7 (25.7 to 29.7%)
LR2 Timmerman 2010	**(3)** age of the woman (in years) **(6)** presence of ascites (yes, 1; no, 0) **(7)** presence of blood flow within a solid papillary projection (yes, 1; no, 0) **(9)** maximum diameter of the solid component of the adnexal mass (expressed in millimetres, but with no increase 950 mm) **(10)** irregular internal cyst walls (yes, 1; no, 0) **(11)** presence of acoustic shadows (yes, 1; no, 0) The probability of malignancy is calculated using the formula y = 1/(1 + exp(jz)), where z = j5.3718 + 0.0354 **(3)** + 1.6159 **(6)** + 1.1768 **(7)** + 0.0697 **(9)** + 0.9586 **(10)** j 2.9486 **(11)**. The probability y is dichotomised at 0.1 to give a predictive diagnosis of cancer.	10% probability of ovarian cancer
ADNEX van Calster 2014	Age (years) Serum CA125 level (log transformed) Type of centre (oncology centres vs other hospitals) Maximum diameter of the lesion (log transformed) Proportion of solid tissue (with quadratic term) Number of papillary projections > 10 cyst locules Acoustic shadows Ascites	3%, 5%, 10% and 15% probability of ovarian cancer

ADNEX: Assessment of Different NEoplasias in the adneXa model; LR2: Logistic Regression Model 2; RMI: Risk of Malignancy Index; ROMA: Risk of Ovarian Malignancy Algorithm.

Table 2. Summary bivariate estimates of RMI I, ROMA, LR2 and ADNEX at all thresholds in pre- and postmenopausal women

Pooled sensitivity and specificity of RMI, ROMA, ADNEX, and LR2 at thresholds reported in included studies					
Score, threshold and menopause status	**Studies**	**Participants**	**OC cases**	**Pooled sensitivity % (95% CI)**	**Pooled specificity % (95% CI)**
ROMA					
7.4 (premenopausal)	10	3051	342	80.7 (69.6 to 88.5)	80.5 (73.8 to 85.9)
25.3 (post-menopausal)	9	1386	603	86.8 (77.9 to 92.5)	87.6 (80.2 to 92.6)
11.4 (premenopausal)	11	2281	445	80.9 (71.0 to 88.0)	84.1 (81.2 to 86.7)
29.9 (post-menopausal)	12	1797	851	91.6 (84.2 to 95.7)	86.3 (80.1 to 90.7)
12.5 (premenopausal)	3	302	68	63.5 (51.0 to 74.4)	89.3 (80.8 to 94.3)
14.4 (post-menopausal)	3	299	177	88.0 (80.6 to 92.8)	68.3 (57.4 to 77.4)
13.1 (premenopausal)	8	1353	158	75.2 (67.0 to 81.9)	84.0 (78.4 to 88.3)
27.7 (post-menopausal)	9	1265	556	90.5 (86.2 to 93.6)	81.1 (75.7 to 85.5)
7.4 ± 2 (pre-menopausal)	12	3223	378	80.6 (71.5 to 87.3)	81.7 (75.7 to 86.5)
25.3 ± 2 (post-menopausal)	15	2599	1049	87.2 (81.7 to 91.3)	86.0 (80.3 to 90.3)
13.1 ± 2 (pre-menopausal)	27	4463	825	77.8 (72.5 to 82.4)	84.3 (81.3 to 86.8)
27.7 ± 2 (post-menopausal)	13	2002	852	90.4 (87.4 to 92.7)	81.3 (76.9 to 85.0)
RMI I					
200 (premenopausal)	17	5233	851	57.1 (50.6 to 63.4)	92.5 (90.0 to 94.4)
200 (postmenopausal)	17	4369	1664	78.7 (74.3 to 82.5)	85.5 (81.3 to 88.9)
Difference in sensitivity and specificity premenopausal vs postmenopausal				21.6 (13.9 to 29.2); P < 0.0001	−6.9 (−11.3 to −2.6); P = 0.002
250 (premenopausal)	2	461	42	59.5 (44.3 to 73.1)	88.1 (84.6 to 90.8)
250 (postmenopausal)	2	220	97	82.5 (73.6 to 88.8)	79.7 (71.6 to 85.9)

Table 2. Summary bivariate estimates of RMI I, ROMA, LR2 and ADNEX at all thresholds in pre- and postmenopausal women *(Continued)*

Difference in sensitivity and specificity premenopausal vs postmenopausal				23.0 (6.3 to 39.6); P = 0.007	−8.4 (−16.2 to −0.6); P = 0.034
LR2					
10 (premenopausal)	4	2843	619	83.2 (78.6 to 87.0)	90.4 (84.6 to 94.1)
10 (postmenopausal)	5	2157	1124	94.5 (92.8 to 95.7)	60.5 (49.3 to 70.7)
Difference in sensitivity and specificity premenopausal vs postmenopausal				11.2 (6.6 to 15.9); P < 0.0001	−29.9 (−41.7 to −18.0); P < 0.0001
ADNEX D+					
3 (premenopausal)	1	1354	378	97.9 (95.9 to 99.1)	56.6 (53.4 to 59.7)
3 (postmenopausal)	1	1049	602	99.5 (98.6 to 99.9)	25.1 (21.1 to 29.3)
5 (premenopausal)	1	1354	378	97.6 (95.5 to 98.9)	69.5 (66.5 to 72.3)
5 (postmenopausal)	1	1049	602	98.8 (97.6 to 99.5)	37.4 (32.9 to 42.0)
10 (premenopausal)	4	1696	455	94.9 (92.5 to 96.6)	78.2 (75.8 to 80.4)
10 (postmenopausal)	4	1365	749	97.6 (96.2 to 98.5)	55.2 (51.2 to 59.1)
Difference in sensitivity and specificity premenopausal vs postmenopausal				2.7 (0.4 to 4.9); P = 0.023	−23.0 (−27.5 to −18.4); P < 0.0001
15 (premenopausal)	1	1354	378	90.5 (87.1 to 93.2)	83.4 (80.9 to 85.7)
15 (postmenopausal)	1	1049	602	96.5 (94.7 to 97.8)	63.5 (58.9 to 68.0)

ADNEX: Assessment of Different NEoplasias in the adneXa model; CI: confidence interval; LR2: Logistic Regression model 2; OC: ovarian cancer; RMI: Risk of Malignancy Index; ROMA: Risk of Ovarian Malignancy Algorithm.

Table 3. Study characteristics: RMI I

Author year Country	Setting	Participants characteristics	Index test threshold
Abdalla 2017 Poland	**Study criteria:** women scheduled to undergo surgery for adnexal tumours **Clinical setting:** mixed **Prior tests:** USS assessment of adnexal mass and measurement of tumour markers CA125 and HE4 within 5 days before surgical intervention **Exclusions:** presence of fibroids > 5 cm were excluded **Centre:** single	**n:** 312 **Postmenopausal n (%):** 117 (37) **Ovarian cancer n (%):** 45 (15) **Borderline n (%):** 7 (2) **Age:** range 18–85 years **Separated by menopausal status:** yes	**Threshold:** 200 **Prespecified:** yes

Table 3. Study characteristics: RMI I *(Continued)*

Al Musalhi 2016 Oman	**Study criteria:** women with an ovarian mass **Clinical setting:** mixed **Prior tests:** unclear but assume USS **Exclusions:** none reported **Centre:** single	**n:** 213 **Postmenopausal n (%):** 51 (24) **Ovarian cancer n (%):** 48 (23) **Borderline n (%):** 7 (3) **Age:** not reported **Separated by menopausal status:** yes	**Threshold:** 200 **Prespecified:** yes
Anton 2012 Brazil	**Study criteria:** women referred with pelvic mass diagnosed by USS, CT or MRI with signs of carcinomatosis undergoing surgery or image-guided biopsy **Clinical setting:** secondary care **Prior tests:** unclear **Exclusions:** none reported **Centre:** single	**n:** 120 **Postmenopausal n (%):** 73 (60) **Ovarian cancer n(%):** 30 (25) **Borderline n (%):** 17 (14) **Mean age:** malignant 54.7 years, borderline 56.4 years, benign 50.7 years **Separated by menopausal status:** yes	**Threshold:** 200 **Prespecified:** yes
Ertas 2016 Turkey	**Study criteria:** women with adnexal masses that underwent surgery **Clinical setting:** tertiary **Prior tests:** unclear **Exclusions:** none reported **Centre:** single	**n:** 408 **Postmenopausal n (%):** 117 (71.4) **Ovarian cancer n (%):** 55 (13) **Borderline n (%):** 12 (3) **Mean age:** benign 40.8 (SD 13.8) years, malignant 54.4 (SD 13.6) years **Separated by menopausal status:** yes	**Threshold:** 200 **Prespecified:** yes
Irshad 2013 Pakistan	**Study criteria:** unclear (ovarian masses) **Clinical setting:** secondary **Prior test:** unclear **Exclusions:** unclear **Centre:** single	**n:** 36 **Postmenopausal n (%):** 36 (100) **Ovarian cancer n (%)**: 24 (37) **Borderline n (%):** not reported **Mean age:** 58 years **Separated by menopausal status:** yes	**Thresholds:** 250 **Prespecified:** yes
Krascsenitis 2016 Hungary	**Study criteria:** women diagnosed with an ovarian tumour of unknown significance admitted for surgery **Clinical setting:** tertiary **Prior tests:** not reported **Exclusions:** none reported **Centre:** single	**n:** 162 **Postmenopausal n (%):** 102 (63) **Ovarian cancer n (%):** 34 (21) **Borderline n (%):** 11 (7) **Mean age:** 55 years	**Threshold:** 200 **Prespecified:** yes

Table 3. Study characteristics: RMI I *(Continued)*

		Separated by menopausal status: yes	
Liest 2019 Sweden	**Study criteria:** women with a pelvic mass of probable ovarian origin and scheduled for surgery **Clinical setting:** tertiary **Prior tests:** preoperative USS **Exclusions:** none reported **Centre:** multicentre	**n:** 784 **Postmenopausal n (%):** 117 (81) **Ovarian cancer n (%):** 144 (18) (include borderline) **Borderline n (%):** not reported **Age:** not reported **Separated by menopausal status:** yes	**Threshold:** 200 **Prespecified:** yes
Lycke 2018 Sweden	**Study criteria:** women planned for a surgical procedure for a symptomatic/suspected malignant ovarian cyst or pelvic tumour **Clinical setting:** mixed **Prior tests:** unclear but assume history and examination, and USS from participant selection **Exclusions:** none reported **Centre:** multicentre	**n:** 638 **Postmenopausal n (%):** 348 (55) **Ovarian cancer n (%):** 162 (25) **Borderline n (%):** 31 (5) **Mean age:** benign 50.76 years, BOT 55.58 years, EOC 62.67 **Separated by menopausal status:** yes	**Threshold:** 200 **Prespecified:** yes
Manegold-Brauer 2016 Switzerland	**Study criteria:** women who had USS examination for an adnexal mass with histology and CA125 results available **Clinical setting:** secondary **Prior tests:** not reported **Exclusions:** none reported **Centre:** single	**n:** 1108 **Postmenopausal n (%):** 478 (43) **Ovarian cancer n (%):** 118 (11) **Borderline n (%):** 33 (3) **Median age:** 48 years **Separated by menopausal status:** yes	**Threshold:** 200 **Prespecified:** yes
Meys 2017 Netherlands	**Study criteria**: women with adnexal pathology **Clinical setting:** tertiary **Prior tests:** not reported **Exclusions:** none reported **Centre:** single	**n:** 326 **Postmenopausal n (%):** 198 (61) **Ovarian cancer n (%):** 115 (35) **Borderline n (%):** 27 (8) **Median age:** benign 53.2 (IQR 16.1–87.2) years, malignant 67.7 (IQR 32.3–87) years **Separated by menopausal status:** yes	**Threshold:** 200 **Prespecified:** yes
Niemi 2017 Finland	**Study criteria:** women aged > 50 years presenting with an abnormal adnexal mass(es) **Clinical setting:** tertiary	**n:** 98 **Postmenopausal n (%):** 98 (100) **Ovarian cancer n (%):** 23 (23)	**Threshold:** 200 **Prespecified:** yes

Table 3. Study characteristics: RMI I *(Continued)*

	Prior tests: not reported **Exclusions:** overtly benign or malignant-appearing tumours such as unilocular simple ovarian cysts and tumours associated with marked ascites (depth of the greatest pool > 10 cm) **Centre:** single	**Borderline n (%):** 7 (7) **Median age:** 61 (range 50–84) years **Separated by menopausal status:** only postmenopausal included	
Nikolova 2016 Macedonia	**Study criteria:** premenopausal women with USS confirming an ovarian cyst/mass and undergoing surgery **Clinical setting:** tertiary **Prior test:** unclear **Exclusions:** postmenopausal women **Centre:** single	**n:** 105 (analysed) **Postmenopausal n (%):** 0 **Ovarian cancer n (%):** 11 (10%) **Borderline n (%):** not reported **Mean age:** ovarian cancer 42.46 (SD 8.21) years, benign 36.90 (SD 10.12) years **Separated by menopausal status:** only premenopausal women included	**Threshold:** 250 **Prespecified:** yes
Radosa 2011 Germany	**Study criteria:** women with adnexal mass who subsequently underwent surgery were selected **Clinical setting:** tertiary **Prior test:** unclear **Exclusions:** none **Centre:** single	**n:** 442 **Postmenopausal n (%):** 141 (32) **Ovarian cancer n (%)**: 79 **Borderline n (%):** 19 **Mean age:** 43.3 years **Separated by menopausal status:** yes	**Thresholds:** 200 **Prespecified:** yes
Richards 2015 Australia	**Study criteria:** women who were undergoing surgery for a complex pelvic mass, presumed to be arising from the ovary **Clinical setting:** mixed **Prior tests:** unclear **Exclusions:** none reported **Centre:** single	**n:** 50 **Postmenopausal n (%):** 29 (58) **Ovarian cancer n (%):** 16 (32) **Borderline n (%):** 4 (8) **Median age:** 60 years **Separated by menopausal status:** yes	**Threshold:** 200 **Prespecified:** yes
Sayasneh 2013a UK	**Study criteria:** women presenting with adnexal mass and undergoing surgery within 120 days after examination **Clinical setting:** mixed **Prior test:** unclear **Exclusions:** none **Centre:** multicentre	**n:** 255 **Postmenopausal n (%):** 117 (46) **Ovarian cancer n (%)**: 48 (19) **Borderline n (%):** 18 (7) **Mean age:** 46 years **Separated by menopausal status:** yes	**Thresholds:** 200 **Prespecified:** yes

Table 3. Study characteristics: RMI I *(Continued)*

Terzic 2013 Serbia	**Study criteria:** women treated for adnexal tumours **Clinical setting:** secondary **Prior test:** unclear **Exclusions:** none **Centre:** single	**n:** 689 **Postmenopausal n (%):** 138 (20) **Ovarian cancer n (%)**: 112 (16) **Borderline n (%):** 33 (5) **Mean age:** benign 42.8 years, borderline: 53.6 years, malignant 57.25 years **Separated by menopausal status:** yes	**Thresholds:** 250 **Prespecified:** yes
Testa 2014 European countries	**Study criteria:** women presenting with adnexal mass and undergoing TVS by 1 of the principal investigators and surgery within 120 days after examination **Clinical setting:** mixed **Prior test:** unclear **Exclusions:** none **Centre:** single	**n:** 2403 **Postmenopausal n (%):** 1049 (44) **Ovarian cancer n (%)**: 701 (29) **Borderline n (%):** 153 (6) **Age:** not reported **Separated by menopausal status:** yes	**Thresholds:** 200 **Prespecified:** yes
van den Akker 2016 Netherlands	**Study criteria:** women admitted for surgical treatment of an ovarian mass with unknown histology **Clinical setting:** mixed **Prior tests:** not reported **Exclusions:** women with clear evidence of malignancy found before or during the surgical procedure (e.g. pleural effusions and evidence of distal organ involvement) **Centre:** multicentre	**n:** 670 **Postmenopausal n (%):** 390 (58) **Ovarian cancer n (%):** 93 (14) **Borderline n (%):** 46 (6) **Median age:** 54 years **Separated by menopausal status:** yes	**Threshold:** 200 **Prespecified:** yes
van Gorp 2012 Belgium	**Study criteria:** women with a pelvic mass, scheduled for surgery **Clinical setting:** secondary **Prior test:** unclear **Exclusions:** none **Centre:** single	**n:** 374 **Postmenopausal n (%):** 196 (52) **Ovarian cancer n (%)**: 94 (25) **Borderline n (%):** 31 (8) **Mean age:** benign 46.2 years, malignant 57.7 years **Separated by menopausal status:** yes	**Thresholds:** 200 **Prespecified:** yes
Vural 2016 Turkey	**Study criteria**: postmenopausal women with adnexal masses who underwent surgery **Clinical setting:** tertiary **Prior tests:** not reported **Exclusions:** premenopausal women	**n:** 139 **Postmenopausal n (%):** 139 (100) **Ovarian cancer n (%):** 44 (32) **Borderline n (%):** 8 (6)	**Threshold:** 200 **Prespecified:** yes

Table 3. Study characteristics: RMI I *(Continued)*

	Centre: single	**Mean age:** 61.1 (SD 8.9) years (range 42–87 years) **Separated by menopausal status:** yes

*Thresholds extracted for RMI I: 200 and 250.
BOT: borderline ovarian tumour; CT: computed tomography; EOC: epithelial ovarian cancer; HE4: Human Epididymis protein; IQR: interquartile range; MRI: magnetic resonance imaging; n: number of participants; RMI I: Risk of Malignancy Index I; SD: standard deviation; TVS: transvaginal ultrasound; USS: ultrasound scan.

Table 4. Study characteristics: ROMA

Author year Country	Setting	Participant characteristics	Index test threshold*
Al Musalhi 2016 Oman	**Study criteria:** women with an ovarian mass **Clinical setting:** mixed **Prior tests:** unclear but assumed USS **Exclusions:** none reported **Centre:** single	**n:** 213 **Postmenopausal n (%):** 51 (24) **Ovarian cancer n (%):** 48 (23) **Borderline n (%):** 7 (3) **Age:** not reported **Separated by menopausal status:** yes	**Threshold:** premenopausal 13.1, postmenopausal 27.7 **Prespecified:** yes
Anton 2012 Brazil	**Study criteria:** women with signs of carcinomatosis with a pelvic mass diagnosed by US, CT or MRI undergoing surgery or image-guided biopsy **Clinical setting:** secondary care **Prior tests:** not reported **Exclusions**: none reported **Centre:** single	**n:** 120 **Postmenopausal n (%):** 73 (60.8%) **Ovarian cancer n (%):** 30 (25%) **Borderline n (%):** 17 (14%) **Mean age:** malignant 54.7 years, borderline 56.4 years, benign 50.73 years **Separated by menopausal status:** yes	**Thresholds:** premenopausal 13.1, postmenopausal 27.7 **Prespecified:** yes
Bandiera 2011 USA	**Study criteria:** not reported **Clinical setting:** tertiary care **Prior tests:** not reported **Exclusions:** non-EOC **Centre:** single	**n:** 278 **Postmenopausal n (%):** 183 (65.8) **Ovarian cancer n (%):** 113 (41) **Borderline n (%):** not reported **Mean age:** premenopausal: malignant 44.7 years, benign 41.5 years; postmenopausal: malignant 66.3 years, benign 64.0 years **Separated by menopausal status:** yes	**Thresholds:** premenopausal 7.4, postmenopausal 25.3 **Prespecified:** yes
Chan 2013 Asia-Pacific region	**Study criteria:** women aged > 18 years diagnosed with adnexal mass diagnosed by any imaging method (US, CT or MRI)	**n:** 414 **Postmenopausal n (%):** 26 (108)	**Thresholds:** premenopausal 7.4,

Table 4. Study characteristics: ROMA *(Continued)*

	Clinical setting: unclear **Prior test:** unclear **Exclusions:** none **Centre:** multicentre	**Ovarian cancer n (%):** 74 (18) **Borderline n (%):** 16 (4) **Age mean:** not reported **Separated by menopausal status:** yes	postmenopausal 25.3 **Prespecified:** yes
Chen 2015 China	**Study criteria:** women with pelvic masses scheduled for surgery **Clinical setting:** unclear **Prior test:** unclear **Exclusions:** none **Centre:** single	**n:** 130 **Postmenopausal n (%):** 62 (48) **Ovarian cancer n (%):** 60 (46) **Borderline n (%):** not reported **Median age:** benign 34 years, malignant 53 years **Separated by menopausal status:** yes	**Thresholds:** premenopausal 11.4, postmenopausal 29.9 **Prespecified:** yes
Chen 2014 China	**Study criteria:** women with EOC and benign lesions **Clinical setting:** tertiary **Prior test:** unclear **Exclusions:** women with non-EOC **Centre:** single	**n:** 192 **Postmenopausal n (%):** 84 (44) **Ovarian cancer n (%):** 123 (64) **Borderline n (%):** not reported **Age mean:** not reported **Separated by menopausal status:** yes	**Thresholds:** premenopausal 12.2, postmenopausal 25.8 **Prespecified:** yes
Chudecka-Glaz 2015 Poland (ROMA and ROMA-P)	**Study criteria:** consecutive women who attended the hospital presenting with suspected ovarian cancer (ovarian tumour, ovarian cyst, or ascites) **Clinical setting:** tertiary **Prior test:** not reported **Exclusions:** none reported **Centre:** single	**n:** 413 **Postmenopausal (%):** 251 (61) **Ovarian cancer n (%):** 162 (39%) **Borderline n (%):** not reported **Age median:** benign 35 years, malignant 59.7 years **Separated by menopausal status:** yes	**a) ROMA** **Thresholds:** premenopausal 14.1, postmenopausal 25 **Prespecified:** yes **b) ROMA-P** **Thresholds:** determined by age group in both pre- and postmenopausal; age group included: < 20 years, 21–30 years, 31–40 years, 41–50 years, 51–60 years, 61–70 years, 71–80 years, and > 80 years **Prespecified:** no
Cradic 2018 USA	**Study criteria:** women with EOC or benign ovarian lesions **Clinical setting:** tertiary **Prior test:** not reported	**n:** 207 **Postmenopausal n (%):** 93 (45) **Ovarian cancer n (%):** 76 (37) (EOC) **Borderline n (%):** not reported	**Thresholds:** premenopausal 11.4, postmenopausal 29.9 **Prespecified:** yes

Table 4. Study characteristics: ROMA *(Continued)*

	Centre: single	**Age mean:** not reported **Separated by menopausal status:** yes	
Dikmen 2015 Turkey	**Study criteria:** women were 'preoperative' **Clinical setting:** unclear **Prior test:** unclear **Exclusions:** none reported **Centre:** unclear	**n:** 143 **Postmenopausal n (%):** 46 (32%) **Ovarian cancer n (%):** 47 (33%) **Borderline n (%):** not reported **Age mean:** benign 42 (SD 10) years, malignant 56 (SD 14) years **Separated by menopausal status:** yes	**Thresholds:** pre-menopausal 13.1, postmenopausal 27.7 **Prespecified:** yes
Farzaneh 2014 Iran	**Study criteria:** women with adnexal mass undergoing surgery and having attained menarche 12 months before presenting with adnexal mass **Clinical setting:** secondary **Prior test:** unclear **Exclusions:** non-EOC **Centre:** single	**n:** 99 **Postmenopausal n (%):** 31 (31) **Ovarian cancer n (%):** 43 (43) (EOC) **Borderline n (%):** not reported **Mean age:** benign 39 years, malignant 51 years **Separated by menopausal status:** yes	**Thresholds:** pre-menopausal 11.5, postmenopausal 25.5 **Prespecified:** yes
Grenache 2015 USA	**Study criteria:** women with abnormal adnexal mass detected on physical examination and imaging Included USS, CT or MRI) followed by surgery **Clinical setting:** unclear **Prior test:** unclear **Exclusions:** unclear **Centre:** multicentre	**n:** 146 **Postmenopausal n (%):** 76 (52) **Ovarian cancer n (%):** 19 (13) **Borderline n (%):** 7 (5) **Mean age:** 52 years **Separated by menopausal status:** yes	**Thresholds:** pre-menopausal 8.6 and 13.1, post-menopausal 27.7 **Prespecified:** yes
Huy 2018 Vietnam	**Study criteria:** women with sufficient personal information, clinical symptoms, data on serum CA125 and serum HE4 levels, and postoperative pathologic findings **Clinical setting:** mixed **Prior test:** not reported **Exclusions:** unclear borderline cases **Centre:** single	**n:** 277 **Postmenopausal n (%):** 47 (17) **Ovarian cancer n (%):** 30 (11) (EOC only) **Borderline n (%):** not reported **Age:** not reported **Separated by menopausal status:** yes	**Thresholds:** pre-menopausal 7.4, postmenopausal 25.3 **Prespecified:** yes
Karlsen 2012 Denmark	**Study criteria:** women admitted to surgery for pelvic mass or pelvic pain potentially caused by malignant disease or endometriosis **Clinical setting:** secondary **Prior test:** unclear	**n:** 1218 **Postmenopausal n (%):** 621 (51) **Ovarian cancer n (%):** 261 (21) **Borderline n (%):** 79 (6) **Age mean:** not reported	**Thresholds:** pre-menopausal 7.4, postmenopausal 25.3 **Prespecified:** yes

Table 4. Study characteristics: ROMA *(Continued)*

	Exclusions: none **Centre:** single	**Separated by menopausal status:** yes	
Kadija 2012 Serbia	**Study criteria:** women diagnosed with adnexal mass scheduled to undergo surgery **Clinical setting:** secondary **Prior test:** unclear **Exclusions:** none **Centre:** single	**n:** 108 **Postmenopausal n (%):** 41 (38) **Ovarian cancer n (%):** 24 (22) **Borderline n (%):** 5 (5) **Age:** not reported **Separated by menopausal status:** yes	**Thresholds:** premenopausal 12.5, postmenopausal 14.4 **Prespecified:** no
Kim 2011 South Korea	**Study criteria:** women diagnosed with adnexal mass on the first visit to the gynaecological oncology clinic and underwent surgery **Clinical setting:** tertiary **Prior test:** unclear **Exclusions:** only EOC included **Centre:** single	**n:** 159 **Postmenopausal n (%):** 108 (68) **Ovarian cancer n (%):** 68 (43) **Borderline n (%):** 10 (6) **Mean age:** benign 35.7, malignant 51.7 **Separated by menopausal status:** **yes	**Threshold:** premenopausal 7.6 **Prespecified:** yes
Kim 2019 Korea	**Study criteria:** women with suspected gynaecological disease **Clinical setting:** tertiary **Prior test:** unclear **Exclusions:** unclear; presume BOT excluded as retrospective **Centre:** single	**n:** 832 **Postmenopausal n (%):** 251 (30) **Ovarian cancer n (%):** 70 (8) **Borderline n (%):** not reported **Median age:** benign 45.0 (IQR 36.0–51.0) years, malignant 64.0 (IQR 50.9–77.0) years **Separated by menopausal status:** yes	**Thresholds:** premenopausal 11.4, postmenopausal 29.9 **Prespecified:** yes
Krascsenitis 2016 Hungary	**Study criteria:** women diagnosed with an ovarian tumour of unknown significance admitted for surgery **Clinical setting:** tertiary **Prior tests:** not reported **Exclusions:** none reported **Centre:** single	**n:** 162 **Postmenopausal n (%):** 102 (63) **Ovarian cancer n (%):** 34 (21) **Borderline n (%):** 11 (7) **Mean age:** 55 years **Separated by menopausal status:** yes	**Thresholds:** premenopausal 11.4, postmenopausal 29.9 **Prespecified:** yes
Li 2016 **China**	**Study criteria:** women diagnosed with gynaecological diseases by US, CT scan, PET-CT scan or MRI **Clinical setting:** unclear **Prior test:** not reported **Exclusions:** none	**n:** 916 **Postmenopausal n (%):** 172 (19) **Ovarian cancer n (%):** 190 **Borderline n (%):** not reported **Median age:** 50 years (range 18–82 years)	**Thresholds:** premenopausal 7.4, postmenopausal 25.3 **Prespecified:** yes

Table 4. Study characteristics: ROMA *(Continued)*

	Centre: single	**Separated by menopausal status:** yes	
Liest 2019 Sweden	**Study criteria:** women with a pelvic mass of probable ovarian origin and scheduled for surgery **Clinical setting:** tertiary **Prior tests:** preoperative US **Exclusions:** none reported **Centre:** multicentre	**n:** 784 **Postmenopausal n (%):** 117 (81) **Ovarian cancer n (%):** 144 (18) (EOC + borderline) **Borderline n (%):** not reported **Mean age:** not reported **Separated by menopausal status:** yes	**Thresholds:** pre-menopausal 11, postmenopausal 25 **Prespecified:** yes
Lycke 2018 Sweden	**Study criteria:** women planned for a surgical procedure for a symptomatic/suspected malignant ovarian cyst or pelvic tumour **Clinical setting:** mixed **Prior tests:** unclear but assume history and examination, and US from patient selection **Exclusions:** none **Centre:** multicentre	**n:** 638 **Postmenopausal n (%):** 348 (55) **Ovarian cancer n (%):** 162 (25) (EOC only) **Borderline n (%):** 31 (5) **Mean age:** benign 50.76 years, BOT 55.58 years, EOC 62.67 years **Separated by menopausal status:** yes	**Thresholds:** pre-menopausal 11.4, postmenopausal 29.9 **Prespecified:** yes
Melo 2018 Portugal	**Study criteria:** women with adnexal neoplasia submitted to surgical treatment, with a histological diagnosis and in which ROMA had been determined **Clinical setting:** tertiary **Prior test:** unclear **Exclusions:** none reported but age group unclear **Centre:** single	**n:** 247 **Postmenopausal n (%):** 92 (37) **Ovarian cancer n (%):** 34 (14) **Borderline n (%):** 7 (3) **Age:** not reported **Separated by menopausal status:** yes	**Thresholds:** pre-menopausal 7.4, postmenopausal 25.3 **Prespecified:** yes
Molina 2011 Spain	**Study criteria:** not reported **Clinical setting:** unclear **Prior test:** unclear **Exclusions:** none **Centre:** single	**n:** 396 **Postmenopausal n (%):** 143 (36) **Ovarian cancer n (%):** 111 (28) **Borderline n (%):** not reported **Age:** not reported **Separated by menopausal status:** yes	**Thresholds:** pre-menopausal 13.1, postmenopausal 27.7 **Prespecified:** yes
Montagnana 2011 Italy	**Study criteria:** women with pelvic mass scheduled to have radical surgery **Clinical setting:** secondary **Prior test:** unclear **Exclusions:** only EOC included	**n:** 104 **Postmenopausal n (%):** 53 (51) **Ovarian cancer n (%):** 55 (53) **Borderline n (%):** excluded	**Thresholds:** pre-menopausal 12.5, postmenopausal 14.4 **Prespecified:** yes

Table 4. Study characteristics: ROMA *(Continued)*

	Centre: single	**Mean age:** malignant 56.9 years, benign 42 years **Separated by menopausal status:** yes	
Moore 2009 USA	**Study criteria:** women with ovarian cyst scheduled to undergo surgery **Clinical setting:** unclear **Prior test:** unclear **Exclusions:** none **Centre:** multicentre	**n:** 513 **Postmenopausal n (%):** 150 (29) **Ovarian cancer n (%):** 143 (28) **Borderline n (%):** 22 (4) **Mean age:** 54 years **Separated by menopausal status:** yes	**Thresholds:** premenopausal 13.1, postmenopausal 27.7 **Prespecified:** yes
Moore 2011 USA	**Study criteria:** women with ovarian cyst scheduled to undergo surgery **Clinical setting:** mixed **Prior test:** unclear **Exclusions:** none **Centre:** multicentre	**n:** 472 **Postmenopausal n (%):** 217 (46) **Ovarian cancer n (%):** 68 (14) **Borderline n (%):** 19 (4) **Mean age:** 50.3 years **Separated by menopausal status:** yes	**Thresholds:** premenopausal 13.1, postmenopausal 27.7 **Prespecified:** yes
Nikolova 2016 Macedonia	**Study criteria:** premenopausal women to have an USS confirming an ovarian cyst/mass and to undergo surgery **Clinical setting:** tertiary **Prior test:** unclear **Exclusions:** postmenopausal women **Centre:** single	**n:** 105 (analysed) **Postmenopausal n (%):** 0 **Ovarian cancer n (%):** 11 (10%) (EOC only) **Borderline n (%):** not reported **Mean age:** malignant 42.46 (SD 8.21) years, benign 36.90 (SD 10.12) years **Separated by menopausal status:** only premenopausal women included	**Thresholds:** premenopausal 7.4 **Prespecified:** yes
Novotny 2012 Czech Republic	**Study criteria:** women with pelvic abnormalities **Clinical setting:** secondary **Prior test:** unclear **Exclusions:** premenopausal women **Centre:** single	**n:** 256 **Postmenopausal n (%):** 256 (100) **Ovarian cancer n (%):** 21 (8) **Borderline n (%):** not reported **Mean age:** benign 65.28 years, malignant 64.37 years **Separated by menopausal status:** yes	**Thresholds:** postmenopausal 26.3 **Prespecified:** no
Ortiz-Munoz 2014 Spain	**Study criteria:** women with gynaecological symptoms, diagnosed with primary ovarian cancer **Clinical setting:** tertiary **Prior test:** symptoms	**n:** 148 **Postmenopausal n (%):** 104 (70) **Ovarian cancer n (%):** 29 (20) **Borderline n (%):** not reported	**Thresholds:** premenopausal 11.4, postmenopausal 29.9 **Prespecified:** yes

Table 4. Study characteristics: ROMA *(Continued)*

	Exclusions: none **Centre:** single	**Age:** not reported **Separated by menopausal status:** **yes	
Park 2019 Korea	**Study criteria:** women for whom gynaecologists had requested HE4, CA125 and ROMA tests to evaluate a pelvic mass **Clinical setting:** secondary **Prior test:** USS, CT or MRI **Exclusions:** 2 cases of non-EOC excluded from analysis **Centre:** single	**n:** 433 (biopsy 309; follow-up 134) **Postmenopausal n (%):** biopsy: 81 (26), follow-up: 37 (28) **Ovarian cancer n (%):** 18 (4) **Borderline n (%):** 15 (3) **Median age:** benign 43.0 (SD 21.0) years, malignant 52.3 (SD 6.1) years, BOT 47.8 (SD 12.9) years **Separated by menopausal status:** yes	**Thresholds:** premenopausal 7.4, postmenopausal 25.3 **Prespecified:** yes
Partheen 2011a Sweden	**Study criteria:** women with complex cystic mass and suspicious of malignancy undergoing surgery **Clinical setting:** tertiary **Prior test:** unclear **Exclusions:** solid and unilocular mass **Centre:** single	**n:** 374 **Postmenopausal n (%):** 276 (74) **Ovarian cancer n (%):** 108 (29) **Borderline n (%):** 45 (12) **Age:** not reported **Separated by menopausal status:** **yes	**Thresholds:** premenopausal 17.3, postmenopausal 26.0 **Prespecified:** yes
Prskalo 2015 Croatia	**Study criteria:** women with suspected adnexal mass on a TVS scheduled for elective surgery **Clinical setting:** mixed **Prior test:** unclear **Exclusions:** none **Centre:** single	**n:** 159 **Postmenopausal n (%):** 102 (64) **Ovarian cancer n (%):** 43 (27) **Borderline n (%):** 11 (7) **Mean age:** premenopausal 36.9 (SD 8.9) years; postmenopausal 60.2 (SD 9.6) years **Separated by menopausal status:** yes	**Thresholds:** premenopausal 11.7, postmenopausal 29.9 **Prespecified:** yes
Richards 2015 Australia	**Study criteria:** women who were undergoing surgery for a complex pelvic mass, presumed to be arising from the ovary **Clinical setting:** mixed **Prior tests:** unclear **Exclusions:** none reported **Centre:** single	**n:** 50 **Postmenopausal n (%):** 29 (58) **Ovarian cancer n (%):** 16 (32) (EOC only) **Borderline n (%):** 4 (8) **Median age:** 60 years **Separated by menopausal status:** yes	**Thresholds:** premenopausal 7.4, postmenopausal 25.3 **Prespecified:** yes
Romagnolo 2016 Italy	**Study criteria:** women referred to gynaecological oncologist with a suspicious pelvic mass requiring surgery **Clinical setting:** tertiary	**n:** 387 **Postmenopausal n (%):** 148 (38)	**Thresholds:** premenopausal 13.1, postmenopausal 27.7

Table 4. Study characteristics: ROMA *(Continued)*

	Prior test: pelvic masses confirmed by USS prior to inclusion **Exclusions:** non-EOC **Centre:** multicentre	**Ovarian cancer n (%):** 73 (19) (EOC only) **Borderline n (%):** 15 (3.9) **Mean age:** premenopausal 37.6 (SD 8.6) years, postmenopausal 63 (SD 9.5) years **Separated by menopausal status:** yes	**Prespecified:** yes
Salim 2018 Pakistan	**Study criteria:** postmenopausal women with ovarian mass (> 2 cm) on pelvic ultrasound examination, attending gynaecology clinics, planned for surgical intervention **Clinical setting:** secondary **Prior test:** not reported **Exclusions:** only postmenopausal women included **Centre:** single	**n:** 260 **Postmenopausal n (%):** 260 (100) **Ovarian cancer n (%):** 122 (47) **Borderline n (%):** NR **Mean age:** 49.28 (SD 6.26) years **Separated by menopausal status:** only postmenopausal women included	**Thresholds:** postmenopausal 27.7 **Prespecified:** yes
Shen 2017 China	**Study criteria:** women referred to a participating centre with a pelvic mass or an ovarian cyst and planning to undergo surgery **Clinical setting:** mixed **Prior test:** pelvic USS, CT, MRI and the medical history (the diagnosis and treatment of pelvic mass and history of renal disease) **Exclusions:** none **Centre:** multicentre	**n:** 684 **Postmenopausal n (%):** 174 (25) **Ovarian cancer n (%):** 169 (25) (EOC + BOT) **Borderline n (%):** 18 (3) **Mean age:** 58.8 (SD 8.6) years **Separated by menopausal status:** yes	**Thresholds:** premenopausal 7.4, postmenopausal 25.3 **Prespecified:** yes
Stiekma 2014 Netherlands	**Study criteria:** histologically confirmed EOC or benign ovarian disease referred to the institute **Clinical setting:** tertiary **Prior test:** unclear **Exclusions:** BOT **Centre:** single	**n:** 181 **Postmenopausal n (%):** 143 (79) **Ovarian cancer n (%):** 147 (81) **Borderline n (%):** excluded **Mean age:** benign 47 years, malignant 57 years **Separated by menopausal status:** yes	**Thresholds:** premenopausal 12.9, postmenopausal 27.8 **Prespecified:** yes
Teh 2018 Malaysia	**Study criteria:** women with pelvic mass(es) suspected of originating in the ovary who had been scheduled for surgery or radiological-guided biopsy **Clinical setting:** tertiary **Prior test:** not reported **Exclusions:** unclear; low malignant potential tumours were included in the benign tumour group for analysis	**n:** 129 **Postmenopausal n (%):** 27 (21) **Ovarian cancer n (%):** 27 (21) **Borderline n (%):** 10 (8) **Median age:** 37 (IQR 27.5–48.5) years **Separated by menopausal status:** yes	**Thresholds:** premenopausal 11.4, postmenopausal 29.9 **Prespecified:** yes

Table 4. Study characteristics: ROMA *(Continued)*

	Centre: single		
Terlikowska 2016 Poland	**Study criteria:** Caucasian women surgically treated on account of benign ovarian disease and epithelial cancer according to international treatment guidelines **Clinical setting:** mixed **Prior test:** not reported **Exclusions:** non-EOC **Centre:** multicentre	**n:** 224 **Postmenopausal n (%):** 104 (46) **Ovarian cancer n (%):** 96 (43) (EOC only) **Borderline n (%):** not reported **Median age:** premenopausal 36, postmenopausal 63 **Separated by menopausal status:** yes	**Thresholds:** premenopausal 11.4, postmenopausal 29.9 **Prespecified:** yes
van Gorp 2011 (van Gorp 2012 secondary publication; smaller cohort) Belgium	**Study criteria:** women diagnosed with pelvic mass undergoing surgery **Clinical setting:** unclear **Prior test:** unclear **Exclusions:** none **Centre:** single	**n:** 389 **Postmenopausal n (%):** 161 (41) **Ovarian cancer n (%):** 161 (41) **Borderline n (%):** not reported **Mean age:** benign 46.3 years, malignant 57.8 years **Separated by menopausal status:** yes	**Thresholds:** premenopausal 12.5, postmenopausal 14.4 **Prespecified:** yes
Xu 2016 China	**Study criteria:** women with a pelvic mass (defined as a simple, complex or solid ovarian cyst/pelvic mass) and healthy women from the Physical Examination Center **Clinical setting:** mixed **Prior test:** not reported **Exclusions:** non-EOC **Centre:** single	**n:** 566 **Postmenopausal n (%):** 159 (28) **Ovarian cancer n (%):** 210 (37) (EOC only) **Borderline n (%):** 45 (8) **Mean age:** benign 42 years, malignant 57 years **Separated by menopausal status:** yes	**Thresholds:** premenopausal 11.4, postmenopausal 29.9 **Prespecified:** yes
Zhang 2015 China	**Study criteria:** all women scheduled for surgery, with and without pelvic mass on USS **Clinical setting:** unclear **Prior test:** USS; adnexal lesions reported according to IOTA **Exclusions:** non-EOC excluded **Centre:** multicentre	**n:** 612 **Postmenopausal n (%):** 232 (37) **Ovarian cancer n (%):** 264 (43) (EOC only) **Borderline n (%):** not reported **Median age (25th centile, 75th centile):** benign: premenopausal 41 (35, 46), postmenopausal 57 (54, 68); malignant premenopausal 43 (38, 47), postmenopausal 59 (54, 65) **Separated by menopausal status:** yes	**Thresholds:** premenopausal 11.4, postmenopausal 29.9 **Prespecified:** yes
Zhang 2019 China	**Study criteria:** women with ovarian tumour **Clinical setting:** tertiary **Prior test:** unclear	**n:** 373 **Postmenopausal n (%):** 185 (50) **Ovarian cancer n (%):** 181 (48)	**Thresholds:** premenopausal 11.4, postmenopausal 29.9

Table 4. Study characteristics: ROMA *(Continued)*

	Exclusions: borderline excluded from analysis **Centre:** single	**Borderline n (%):** 17 (5) **Mean age:** 51 years **Separated by menopausal status:** yes	**Prespecified:** yes

*ROMA thresholds most commonly reported and included: **premenopausal** 7.4 (± 2); 12.5; 13.1 (± 2); **postmenopausal** 25.3 (± 2); 14.4; 27.7 (± 2)
**Threshold for premenopausal women OR postmenopausal women reported in the study not included in analysis.
BOT: borderline ovarian tumour; CT: computed tomography; EOC: epithelial ovarian cancer; HE4: Human Epididymis protein; IQR: interquartile range; IOTA: International Ovarian Tumour Analysis; MRI: magnetic resonance imaging; n: number of participants; PET-CT: positron emission tomography–computed tomography; ROMA: Risk of Ovarian Malignancy Algorithm; ROMA-P: a modified ROMA; TVS: transvaginal ultrasound; USS: ultrasound scan.

Table 5. Study characteristics: LR2

Author year country	Setting*	Participant characteristics	Index test threshold
Meys 2017 **Netherlands**	**Study criteria**: women with adnexal pathology **Clinical setting:** tertiary **Prior tests:** not reported **Exclusions:** none reported **Centre:** single	**n:** 326 **Postmenopausal n (%):** 198 (61) **Ovarian cancer n (%):** 115 (35) **Borderline n (%):** 27 (8) **Median age:** malignant 67.7 (IQR 32.3–87) years, borderline 53.2 (016.1–87.2) years **Separated by menopausal status:** yes	**Threshold:** 10% post-test probability of malignancy **Prespecified:** yes
Niemi 2017 **Finland**	**Study criteria**: women aged > 50 years presenting with an abnormal adnexal mass(es) **Clinical setting:** tertiary **Prior tests:** not reported **Exclusions:** overtly benign or malignant-appearing tumours such as unilocular simple ovarian cysts and tumours associated with marked ascites (depth of the greatest pool > 10 cm) **Centre:** single	**n:** 98 **Postmenopausal n (%):** 98 (100) **Ovarian cancer n (%):** 23 (23) **Borderline n (%):** 7 (7) **Median age:** 61 (range 50–84) years **Separated by menopausal status:** only postmenopausal included	**Threshold:** 10%, 25% and 43% of post-test probability of malignancy **Prespecified:** yes
Sayasneh 2013a **Secondary study:** Sayasneh 2013 (see under Sayasneh 2013a) **UK**	**Study criteria:** women presenting with adnexal mass and undergoing surgery within 120 days after examination **Clinical setting:** mixed secondary and tertiary care **Prior tests:** not reported **Exclusions:** none reported **Centre:** multicentre	**n:** 255 **Postmenopausal n (%):** 117 (45.9) **Malignant n (%):** 48 (18.8) **Borderline n (%):** 18 (7.1) **Mean age:** 46 years **Separated by menopausal status:** yes	**Threshold:** 10% post-test probability of malignancy **Prespecified threshold:** yes

Table 5. Study characteristics: LR2 *(Continued)*

Testa 2014 **Europe**	**Study criteria:** women presenting with adnexal mass on TVS and undergoing surgery within 120 days. **Clinical setting:** mixed secondary and tertiary care **Prior tests:** not reported **Exclusions:** none reported **Centre:** multicentre	**n:** 2403 **Postmenopausal n (%):** 1049 (43.7) **Malignant n (%):** 701(18.8) **Borderline n (%):** 153 (6.4) **Median age:** malignant 57 (range 33–66) years; benign 44 (range not reported) years **Separated by menopausal status:** yes	**Threshold:** 10% post-test probability of malignancy **Prespecified threshold:** yes
Timmerman 2010 Secondary study: Di Legge 2012 **Europe**	**Study criteria:** women with persistent adnexal mass undergoing surgery within 120 days **Clinical setting:** mixed secondary and tertiary **Prior tests:** not reported **Exclusions:** none reported **Centre:** multicentre	**n:** 1938 **Postmenopausal n (%):** 742 (38.0) **Malignant n (%):** 373 (19.2) **Borderline n (%):** 111 (5.7) **Mean age:** 46 years **Separated by menopausal status:** yes	**Threshold:** 10% post-test probability of malignancy **Prespecified threshold:** yes

*Setting: secondary care: dedicated gynaecologist in a general hospital; tertiary care: gynaecological oncology centre.
IQR: interquartile range; n: number of participants; TVS: transvaginal ultrasound.

Table 6. Study characteristics: ADNEX

Author year country	Setting*	Participants characteristics	Index test threshold
Meys 2017 **Netherlands**	**Study criteria**: women with adnexal pathology **Clinical setting:** tertiary **Prior tests:** not reported **Exclusions:** none reported **Centre:** single	**n:** 326 **Postmenopausal n (%):** 198 (61) **Ovarian cancer n (%):** 115 (35) **Borderline n (%):** 27 (8) **Median age:** benign 53.2 (IQR 16.1–87.2) years, malignant 67.7 (IQR 32.3–87) years **Separated by menopausal status:** yes	**Threshold:** 10% post-test probability of malignancy **Prespecified:** yes
Szubert 2016a **Poland**	**Study criteria:** women with a 'need for surgery due to an ovarian tumour' **Clinical setting:** unclear, probably tertiary **Prior test:** not reported **Exclusions:** none reported **Centre:** single	**n:** 204 **Postmenopausal n (%):** 66 (54) **Ovarian cancer n (%):** 58 (28) **Borderline n (%):** 12 (6) **Median age:** 46 **Separated by menopausal status:** yes	**Thresholds:** 2000 IOTA criteria 10% **Prespecified: y**es

Table 6. Study characteristics: ADNEX *(Continued)*

Szubert 2016b **Spain**	**Study criteria:** women with a 'need for surgery due to an ovarian tumour' **Clinical setting:** unclear, probably tertiary **Prior test:** not reported **Exclusions:** none reported **Centre:** single	**n:** 128 **Postmenopausal n (%):** 52 (42) **Ovarian cancer n (%):** 35 (27) **Borderline n (%):** 4 (3) **Median age:** 47 years **Separated by menopausal status:** yes	**Thresholds:** 2000 IOTA criteria 10% **Prespecified:** yes
van Calster 2014 **Europe**	**Study criteria:** women presenting with adnexal mass on US and selected for surgery **Clinical setting:** mixed secondary and tertiary care **Prior tests:** not reported **Exclusions:** none reported **Centre:** multicentre	**n:** 2403 **Postmenopausal n (%):** 1049 (43.7)** **Malignant n (%):** 827 (34.4) **Borderline n (%):** 153 (6.4) **Age:** not reported **Separated by menopausal status:** yes**	**Threshold:** 3, 5, 10 and 15% post-test probability of malignancy **Prespecified threshold:** yes

*Setting: secondary care: dedicated gynaecologist in a general hospital; tertiary care: gynaecological oncology centre.
**Contact with authors
IOTA: International Ovarian Tumour Analysis; IQR: interquartile range; n: number of participants.

Table 7. HSROC analysis: comparison of sensitivity at a fixed specificity of 80% and 90%: all studies, all thresholds, pre- and postmenopausal women separately

HSROC analysis: comparison of ROMA, LR2 and ADNEX compared to RMI I. Mixed test positivity threshold analysis at fixed specificities of 80% and 90%									
Test	**Studies**	**Partici-pants (OC cases)**	**Diagnostic odds ratio (95% CI)**	**Relative di-agnostic odds ratio (95% CI)**	**P value**	**Sensitivity at fixed specificity of 80%**		**Sensitivity at fixed specificity of 90%**	
						Sensitivity (95% CI)	**Difference from RMI I (95% CI)**	**Sensitivity (95% CI)**	**Difference from RMI I (95% CI)**
Premenopausal									
RMI I 200/250	19	5694 (893)	15.5 (9.0 to 26.5)	—	—	79.4 (69.5 to 86.7)	—	65.1 (57.2 to 72.2)	—
ROMA mixed	38	7616 (1198)	18.5 (14.3 to 23.9)	1.19 (0.69 to 2.07)	0.5202	82.0 (77.9 to 85.5)	2.6 (–5.5 to 10.7)	68.8 (61.8 to 75.0)	3.7 (–7.3 to 14.7)
LR2	4	2843 (619)	33.9 (21.5 to 53.3)	2.19 (1.18 to 4.06)	0.014	89.0 (83.8 to 92.7)	9.6 (2.2, 17.0)	79.7 (71.3 to 86.1)	14.6 (5.6 to 23.6)
ADNEX 10%	4	1696 (455)	72.6 (29.4 to 179.2)	4.70 (1.45 to 15.20)	0.0108	94.4 (88.3 to 7.4)	14.9 (5.4 to 24.5)	89.0 (77.6 to 95.0)	23.9 (12.0 to 35.8)
Postmenopausal									
RMI I 200/250	19	4589 (1761)	22.8 (17.3 to 30.1)	—	—	85.1 (80.9 to 88.5)	—	71.8 (65.4 to 77.4)	—
ROMA mixed	40	6099 (2746)	40.0 (31.5 to 50.8)	1.75 (1.23 to 2.50)	0.0024	90.9 (88.8 to 92.7)	5.8 (2.1 to 9.6)	81.7 (76.8 to 85.7)	9.9 (4.0 to 15.8)
LR2 10%	5	2157 (1124)	39.5 (22.6 to 69.0)	1.73 (0.97 to 3.09)	0.0622	90.8 (85.9 to 94.1)	5.7 (0.7 to 10.7)	81.5 (70.0 to 89.2)	9.7 (2.0 to 17.4)
ADNEX 10%	4	1365 (749)	56.7 (21.9 to 146.8)	2.48 (0.90 to 6.85)	0.0776	93.4 (85.9 to 97.1)	8.3 (1.5 to 15.1)	86.3 (70.2 to 94.4)	14.6 (3.4 to 25.7)

Notes to table: ADNEX 10% & LR2 10%: threshold to achieve a post-test probability of ovarian cancer of 10%. ADNEX and LR2 studies reported a range of thresholds but all included a threshold of 10%. For RMI I and ROMA studies, each included study contributed a different test positivity threshold.
ADNEX: Assessment of Different NEoplasias in the adneXa model; CI: confidence interval; HSROC: hierarchical summary receiver operating characteristic; LR2: Logistic Regression Model 2; OC: ovarian cancer; RMI I: Risk of Malignancy Index I; ROMA: Risk of Ovarian Malignancy Algorithm.

Table 8. Bivariate comparisons of ROMA, LR2 and ADNEX compared to RMI I in premenopausal women

Bivariate model-pairwise comparisons: premenopausal women					
Absolute sensitivity difference (95% CI); P value for comparison **Absolute specificity difference (95% CI); P value for comparison**			**RMI I** **(200)**	**ROMA** **(13.1 ± 2)**	**LR2** **(10)**
		Studies (participants)	17 (5233)	27 (4463)	4 (2843)
		Sensitivity % (95% CI)	57.2 (50.3 to 63.8)	77.4 (72.7 to 81.5)	83.3 (74.7 to 89.5)
		Specificity % (95% CI)	92.5 (90.3 to 94.2)	84.3 (81.2 to 87.0)	90.4 (84.6 to 94.1)
	Studies (participants)				
ROMA (13.1 ± 2)	27 (4463)	77.4 (95% CI 72.7 to 81.5) 84.3 (95% CI 81.2 to 87.0)	20.2 (12.2 to 28.3); P < 0.0001 −8.2 (−11.7 to −4.7); P < 0.0001	—	—
LR2 (10)	4 (2843)	83.3 (95% CI 74.7 to 89.5) 90.4 (95% CI 84.6 to 94.1)	26.2 (16.2 to 36.2); P < 0.0001 −2.1 (−7.2 to 2.9); P = 0.404	6.0 (−2.6 to 14.5); P = 0.170 6.1 (0.6 to 11.5); P = 0.029	—
ADNEX (10)	4 (1696)	95.5 (95% CI 91.0 to 97.8) 77.8 (95% CI 67.4 to 85.5)	38.3 (30.9 to 45.8); P < 0.0001 −14.8 (−24.0 to −5.5); P = 0.002	18.1 (12.7 to 23.5); P = 0.0001 −6.5 (−16.0 to 3.0); P = 0.178	12.1 (4.2 to 20.1); P = 0.003 −12.6 (−22.8 to −2.4); P = 0.015

ADNEX: Assessment of Different NEoplasias in the adneXa model; CI: confidence interval; LR2: Logistic Regression Model 2; OC: ovarian cancer; RMI I: Risk of Malignancy Index I; ROMA: Risk of Ovarian Malignancy Algorithm.

Table 9. Bivariate comparisons of ROMA, LR2 and ADNEX compared to RMI I in postmenopausal women

Bivariate model-pairwise comparisons: postmenopausal women				
Absolute sensitivity difference (95% CI); Pvalue for comparison		**RMI I** **(200)**	**ROMA** **(27.7 ± 2)**	**LR2** **(10)**

Table 9. Bivariate comparisons of ROMA, LR2 and ADNEX compared to RMI I in postmenopausal women *(Continued)*

		Studies (participants)	17 (4369)	13 (2002)	5 (2157)
		Sensitivity % (95% CI) **Specificity % (95% CI)**	78.4 (74.6 to 81.7) 85.4 (82.0 to 88.2)	90.3 (87.5 to 92.6) 81.5 (76.5 to 85.5)	94.8 (92.3 to 96.6) 60.6 (50.5 to 69.9)
	Studies (participants)				
ROMA (27.7 ± 2)	13 (2002)	90.3 (87.5 to 92.6) 81.5 (76.5 to 85.5)	11.9 (7.6 to 16.3); P < 0.0001 −3.9 (−9.4 to 1.5); P = 0.157	—	—
LR2 (10)	5 (2157)	94.8 (92.3 to 96.6) 60.6 (50.5 to 69.9)	16.4 (12.3 to 20.5); P < 0.0001 −24.8 (−35.1 to −14.5); P < 0.0001	4.5 (1.2 to 7.8); P = 0.008 −20.9 (−31.7 to −10.1); P < 0.0001	—
ADNEX (10)	4 (1365)	97.6 (95.6 to 98.7) 55.0 (42.8 to 66.6)	19.2 (15.4 to 23.1); P < 0.0001 −30.4 (−42.9 to −17.9); P < 0.0001	7.3 (4.3 to 10.2); P < 0.0001 −26.5 (−39.4 to −13.6); P < 0.0001	2.8 (0.2 to 5.3); P = 0.034 −5.6 (−21.2 to 10.0); P = 0.480

ADNEX: Assessment of Different NEoplasias in the adneXa model; CI: confidence interval; LR2: Logistic Regression Model 2; OC: ovarian cancer; RMI I: Risk of Malignancy Index I; ROMA: Risk of Ovarian Malignancy Algorithm.

Table 10. Sensitivity analysis: borderline ovarian tumours

Sensitivity analysis:sensitivity at fixed specificities of 80% and 90% for RMI I and ROMA (all thresholds) for studies grouping borderline ovarian tumours with malignant for the estimation of test accuracy (BOT=1) compared to studies that excluded borderline tumours or where their management for the estimation of test accuracy was unclear (BOT=2/3)

Test	Studies	Participants	OC Cases	DOR (95% CI)	Relative DOR (95% CI)	P value	Sensitivity at fixed specificity of 80%		Sensitivity at fixed specificity of 90%	
							Sensitivity (95% CI)	Difference from BOT=1 (95% CI)	Sensitivity (95% CI)	Difference from BOT=1 (95% CI)
Premenopausal										
RMI I 200/250										
BOT=1	16	4861	801	11.7 (5.3 to 25.9)	—	—	74.9 (59.6 to 85.8)	—	62.2 (53.1 to 70.5)	—
BOT=2/3	3	833	92	11.5 (4.2 to 31.6)	0.98 (0.37 to 2.60)	0.9699	74.6 (55.0 to 87.6)	−0.3 (−16.1 to 15.5)	61.8 (43.3 to 77.4)	−0.4 (−20.1 to 19.4)
ROMA mixed thresholds										
BOT=1	15	2737	363	13.9 (9.0 to 21.7)	—	—	77.6 (69.1 to 84.3)	—	59.2 (47.0 to 70.3)	—
BOT=2/3	23	4879	835	22.3 (15.9 to 31.3)	1.60 (0.94 to 2.74)	0.0837	84.9 (79.7 to 89.0)	7.4 (−1.2 to 15.9)	70.2 (60.3 to 78.6)	11.1 (−1.3 to 23.5)
Postmenopausal										
ROMA mixed thresholds										
BOT=1	15	2289	882	27.4 (18.6 to 40.4)	—	—	87.7 (82.3 to 91.7)	—	72.4 (59.6 to 82.4)	—
BOT=2/3	25	3810	1864	56.3 (40.5 to 78.1)	2.06 (1.24 to 3.40)	0.0062	94.1 (91.3 to 96.0)	6.4 (1.2, 11.5)	85.4 (79.6 to 89.8)	13.0 (1.9 to 24.0)

BOT=1: borderline tumours grouped with malignant ovarian tumours for estimation of test accuracy; BOT=2/3: borderline tumours excluded, grouped with benign or management unclear for estimation of test accuracy; CI: confidence interval; DOR: diagnostic odds ratio; OC: ovarian cancer; RMI: Risk of Malignancy Index; ROMA: Risk of Ovarian Malignancy Algorithm.

Table 11. Excluded studies: no 2 × 2 table

No 2 × 2 table
1. Ovarian cancer: screening, treatment, and follow up. NIH Consensus Statement. 1994;12(3):1-30.
2. Ovarian cancer. Journal of Practical Nursing. 2009;59(3):12-4.
3. CA125 remains best predictor of ovarian cancer. Contemporary OB/GYN. 2011;56(5):16.
4. Clinical use of tumor markers in China. Tumor Biology. 2014;35:S11.
5. A.N AL-N, Ahmed M, Petersen CB. Epithelial ovarian cancer. Obstetrics and Gynecology Clinics of North America. 2012;39(2):269-83.
6. Abbas AM. A new scoring model for characterization of adnexal masses based on two-dimensional gray-scale and colour Doppler sonographic features. Facts Views & Vision in Obgyn. 2014;6(2):68-74.
7. Abdalla N, Bachanek M, Kowalska J, Cendrowski K, Sawicki W. Role of HE4 and simple ultrasound rules proposed by iota group in preoperative evaluation of adnexal masses: A prospective study. International Journal of Gynecological Cancer. 2013;1):835.
8. Abdalla YN, Bachanek M, Kowalska J, Cendrowski K, Sawicki W. The role of risk of malignancy algorithm in the presurgical assessment of adnexal tumors: A prospective study. International Journal of Gynecological Cancer. 2013;1):441.
9. Abdulrahman GO, Jr, McKnight L, Lutchman Singh K. The risk of malignancy index (RMI) in women with adnexal masses in Wales. Taiwanese Journal of Obstetrics & Gynecology. 2014;53(3):376-81.
10. Aggarwal P, Kehoe S. Serum tumour markers in gynaecological cancers. Maturitas. 2010;67(1):46-53.
11. Aguado Romeo MJ, Llanos Mendez A. Serum biomarkers panel for detecting early stage ovarian cancer (Structured abstract). Health Technology Assessment Database [Internet]. 2010 [cited HTA N]; (1). Available from: http://onlinelibrary.wiley.com/o/cochrane/clhta/articles/HTA-32011001516/frame.html.
12. Aiken C, Deakin N, Mehasseb M, Baldwin P. Surgical management of ovarian cysts in premenopausal women. Gynecological Surgery. 2012;1):S24.
13. Akdeniz N, Kuyumcuoglu U, Kale A, Erdemoglu M, Caca F. Risk of malignancy index for adnexal masses. European Journal of Gynaecological Oncology. 2009;30(2):178-80.
14. Alborzi S, Keramati P, Younesi M, Samsami A, Dadras N. The impact of laparoscopic cystectomy on ovarian reserve in patients with unilateral and bilateral endometriomas. Fertility & Sterility. 2014;101(2):427-34.
15. Alcazar JL. Three-dimensional static ultrasound and 3D power doppler in gynecologic pelvic tumors. Donald School Journal of Ultrasound in Obstetrics and Gynecology. 2013;7(2):187-99.
16. Alcazar JL, Errasti T, Laparte C, Jurado M, Lopez-Garcia G. Assessment of a new logistic model in the preoperative evaluation of adnexal masses. Journal of Ultrasound in Medicine. 2001;20(8):841-8.
17. Alcazar JL, Guerriero S, Laparte C, Ajossa S, Jurado M. Contribution of power Doppler blood flow mapping to gray-scale ultrasound for predicting malignancy of adnexal masses in symptomatic and asymptomatic women. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2011;155(1):99-105.
18. Alcazar JL, Utrilla-Layna J, Minguez JA, Jurado M. Clinical and ultrasound features of type I and type II epithelial ovarian cancer. International Journal of Gynecological Cancer. 2013;23(4):680-4.
19. Amayo AA, Kuria JG. Clinical application of tumour markers: a review. East African Medical Journal. 2009;86(12 Suppl):S76-83.
20. Amor F, Vaccaro H, Alcazar JL, Leon M, Craig JM, Martinez J. Gynecologic imaging reporting and data system: a new proposal for classifying adnexal masses on the basis of sonographic findings. Journal of Ultrasound in Medicine. 2009;28(3):285-91.

Table 11. Excluded studies: no 2 × 2 table *(Continued)*

21. Anthoulakis C, Nikoloudis N. Pelvic MRI as the "gold standard" in the subsequent evaluation of ultrasound-indeterminate adnexal lesions: a systematic review. Gynecologic Oncology. 2014;132(3):661-8.
22. Arab M, Honarvar Z, Hosseini-Zijoud S-M. Ovarian Malignancy Probability Score (OMPS) for appropriate referral of adnexal masses. Asian Pacific Journal of Cancer Prevention: Apjcp. 2014;15(20):8647-50.
23. Arab M, Yaseri M, Farzaneh M, Moridi A, Tehranian A, Sheibani K. The construction and validation of a new ovarian malignancy probability score (omps) for prediction of ovarian malignancy. Iranian Journal of Cancer Prevention. 2010;3(3):132-8.
24. Aslam N, Banerjee S, Carr JV, Savvas M, Hooper R, Jurkovic D. Prospective evaluation of logistic regression models for the diagnosis of ovarian cancer. Obstetrics & Gynecology. 2000;96(1):75-80.
25. Autelitano DJ, Raineri L, Knight K, Bannister K, Rice GE. Performance of a multianalyte test as an aid for the diagnosis of ovarian cancer in symptomatic women. Journal of Translational Medicine. 2012;10:45.
26. Azzam AZ, Hashad DI, Kamel NAF. Evaluation of HE4 as an extrabiomarker to CA125 to improve detection of ovarian carcinoma: is it time for a step forward? Archives of Gynecology & Obstetrics. 2013;288(1):167-72.
27. Bandiera E, Ragnoli M, Zanotti L, Bignotti E, Romani C, Tassi R, et al. Clinical role of human epididymis protein 4 (HE4) in epithelial ovarian cancer. European Journal of Cancer, Supplement. 2009;7 (2-3):143.
28. Bankhead CR, Kehoe ST, Austoker J. Symptoms associated with diagnosis of ovarian cancer: a systematic review. BJOG: An International Journal of Obstetrics & Gynaecology. 2005;112(7):857-65.
29. Bast RC, Jr. Biomarkers for ovarian cancer: new technologies and targets to address persistently unmet needs. Cancer Biomarkers: Section A of Disease Markers. 2010;8(4-5):161-6.
30. Bast RC, Jr, Skates S, Lokshin A, Moore RG. Differential diagnosis of a pelvic mass: improved algorithms and novel biomarkers. International Journal of Gynecological Cancer. 2012;22 Suppl 1:S5-8.
31. Behtash N, Rahmani M, Ghotbizadeh F, Karimi Zarchi M, Mousavi A. Ultrasonography and computed tomography for management of adnexal masses in Iranian patients with suspected ovarian cancer: results of a prospective study. Asian Pacific Journal of Cancer Prevention: Apjcp. 2009;10(2):201-4.
32. Berlanda N, Ferrari MM, Mezzopane R, Boero V, Grijuela B, Ferrazzi E, et al. Impact of a multiparameter, ultrasound-based triage on surgical management of adnexal masses. Ultrasound in Obstetrics & Gynecology. 2002;20(2):181-5.
33. Bharwani N, Reznek RH, Rockall AG. Ovarian Cancer Management: The role of imaging and diagnostic challenges. European Journal of Radiology. 2011;78(1):41-51.
34. Bhoola S, Hoskins WJ. Diagnosis and management of epithelial ovarian cancer. Obstetrics and gynecology. 2006;107(6):1399-410.
35. Braicu E, Mecke H, Hellmeyer L, Mueller B, Nohe G, Rodekamp E, et al. HE4 and its value for the gynecologic oncologist-first results from the berlin-roma study. Tumor Biology. 2014;35:S6.
36. Braicu EI, Pietzner K, Henrich W, Sehouli J. Possibilities and limits of diagnostic tests in ovarian cancer. Onkologe. 2014;20(7):650-+.
37. Braicu EI, Van Gorp T, Nassir M, Richter R, Chekerov R, Gasimli K, et al. Preoperative HE4 and ROMA values do not improve the CA125 diagnostic value for borderline tumors of the ovary (BOT) - a study of the TOC Consortium. Journal of ovarian research. 2014;7:49.
38. Brown DL, Dudiak KM, Laing FC. Adnexal masses: US characterization and reporting. Radiology. 2010;254(2):342-54.
39. Bruchim I, Aviram R, Halevy RS, Beyth Y, Tepper R. Contribution of Sonographic Measurement of Ovarian Volume to Diagnosing Ovarian Tumors in Postmenopausal Women. Journal of Clinical Ultrasound. 2004;32(3):107-14.

Table 11. Excluded studies: no 2 × 2 table *(Continued)*

40. Burges A, Schmalfeldt B. Ovarian cancer: diagnosis and treatment. Deutsches Arzteblatt International. 2011;108(38):635-41.
41. Buy JN, Ghossain MA, Hugol D, Hassen K, Sciot C, Truc JB, et al. Characterization of adnexal masses: combination of color Doppler and conventional sonography compared with spectral Doppler analysis alone and conventional sonography alone. AJR American Journal of Roentgenology. 1996;166(2):385-93.
42. Candido Dos Reis FJ, De Andrade JM, Bighetti S. CA 125 and vascular endothelial growth factor in the differential diagnosis of epithelial ovarian tumors. Gynecologic and Obstetric Investigation. 2002;54(3):132-6.
43. Cardillo G, Di Biase I, Biscotti U. The bayes theorem for the evidence based ovarian cancer diagnosis. International Journal of Gynecological Cancer. 2014;2.
44. Cesario S. Advances in the early detection of ovarian cancer. Nursing for Women's Health. 2010;14(3):222-34.
45. Chauhan SC, Singh AP, Ruiz F, Johansson SL, Jain M, Smith LM, et al. Aberrant expression of MUC4 in ovarian carcinoma: Diagnostic significance alone and in combination with MUC1 and MUC16 (CA125). Modern Pathology. 2006;19(10):1386-94.
46. Chia YN, Marsden DE, Robertson G, Hacker NF. Triage of ovarian masses. The Australian & New Zealand journal of obstetrics & gynaecology. 2008;48(3):322-8.
47. Choudhury S, Mohiuddin ASM, Ahmed AU, Ahsan S. Preoperative discrimination of benign and malignant ovarian tumors using color Doppler sonography and its correlation with histopathology. Bangladesh Medical Research Council Bulletin. 2005;31(1):21-6.
48. Chu MM, Fishman D. Risk assessment for epithelial ovarian cancer: proposing a new approach to a deadly problem. Scandinavian Journal of Clinical and Laboratory Investigation Supplement. 2014;244:63-7; discussion 6-7.
49. Chudecka-Glaz AM. ROMA, an algorithm for ovarian cancer. Clinica Chimica Acta. 2015;440:143-51.
50. Coccia ME, Rizzello F, Romanelli C, Capezzuoli T. Adnexal masses: what is the role of ultrasonographic imaging? Archives of Gynecology & Obstetrics. 2014;290(5):843-54.
51. Cohen J, Yu X, Deshmukh H, Zhang R, Shin JY, Osann K, et al. The association of serial ultrasounds and CA-125 prior to diagnosis of ovarian cancer: Do they improve early detection? Gynecologic Oncology. 2009;112(2):S65-S.
52. Colak A, Dikmen ZG, Basaran D, Tuncer S, Akbiyik F. Human epididymis protein 4 (HE4) as a new tumour marker for diagnosis of ovarian carcinoma. FEBS Journal. 2012;279:117.
53. Coleman RL. Two-step Ovarian Cancer Screening: A 'Scissor-step' Forward? Clinical Oncology Alert. 2013;29(10):78-9.
54. Collins GS, Altman DG. Identifying women with undetected ovarian cancer: independent and external validation of QCancer() (Ovarian) prediction model. European Journal of Cancer Care. 2013;22(4):423-9.
55. Cramer DW, Bast RC, Jr, Berg CD, Diamandis EP, Godwin AK, Hartge P, et al. Ovarian cancer biomarker performance in prostate, lung, colorectal, and ovarian cancer screening trial specimens. Cancer Prevention Research. 2011;4(3):365-74.
56. Cramer DW, Vitonis AF. Thoughts on the prevention and early detection of postmenopausal ovarian cancer. Sexuality, Reproduction and Menopause. 2011;9(1):1-7.
57. Cui H, Li D, Chang X, Ye X, Cheng H. The values of serum human epididymis secretory protein 4 and CA125 assay in the diagnosis of ovarian malignance. International Journal of Gynecology and Obstetrics. 2009;107:S240.
58. Daponte A, Stergioti E, Messinis IE. Risk scoring for adnexal masses and endoscopic management. International Journal of Gynecology and Obstetrics.96(1):42-3.
59. de Gauna BR, Sanchez P, Pineda L, Utrilla-Layna J, Juez L, Alcazar JL. Interobserver agreement in describing adnexal masses using the International Ovarian Tumor Analysis simple rules in a real-time setting and using three-dimensional ultrasound volumes and digital clips. Ultrasound in Obstetrics & Gynecology. 2014;44(1):95-9.

Table 11. Excluded studies: no 2 × 2 table *(Continued)*

60. del Carmen MG, Birrer M, Schorge JO. Clear cell carcinoma of the ovary: a review of the literature. Gynecologic Oncology. 2012;126(3):481-90.

61. del Carmen MG, Birrer M, Schorge JO. Carcinosarcoma of the ovary: a review of the literature. Gynecologic Oncology. 2012;125(1):271-7.

62. Diaz E, Burky RE, Hummel CS, Farias-Eisner R. Early detection of ovarian cancer. Expert Review of Obstetrics and Gynecology. 2013;8(2):169-79.

63. Diaz-Padilla I, Razak ARA, Minig L, Bernardini MQ, Maria Del Campo J. Prognostic and predictive value of CA-125 in the primary treatment of epithelial ovarian cancer: potentials and pitfalls. Clinical & Translational Oncology: Official Publication of the Federation of Spanish Oncology Societes & of the National Cancer Institute of Mexico. 2012;14(1):15-20.

64. Dilks P, Narayanan P, Reznek R, Sahdev A, Rockall A. Can quantitative dynamic contrastenhanced MRI independently characterize an ovarian mass? European Radiology. 2010;20(9):2176-83.

65. Djordjevic B, Stojanovic S, Conic I, Jankovic-Velickovic L, Vukomanovic P, Zivadinovic R, et al. Current approach to epithelial ovarian cancer based on the concept of cancer stem cells. Journal of BUOn. 2012;17(4):627-36.

66. Dochez V, Randet M, Renaudeau C, Dimet J, Le Thuaut A, Caillon H, et al. Diagnostic performances of HE4, CA125, RMI and ROMA for the detection of ovarian cancer in presumed benign ovarian tumours. BJOG: An International Journal of Obstetrics and Gynaecology. 2017;124 (Supplement 1):77.

67. Domali E, Ficherova D, Kyriakopoulos K, Papamanolis V, Rodolakis A, Chaidopoulos D, et al. Endometrial and ovarian malignancies: Diagnostic involvement of ultrasound and CT scans. International Journal of Gynecological Cancer. 2011;3):S400.

68. Donach M, Yu Y, Artioli G, Banna G, Feng W, Bast Jr RC, et al. Combined use of biomarkers for detection of ovarian cancer in high-risk women. Tumor Biology. 2010;31(3):209-15.

69. Dorigo O, Berek JS. Personalizing CA125 levels for ovarian cancer screening. Cancer Prevention Research. 2011;4(9):1356-9.

70. Dotlic J, Terzic M, Likic I, Brndusic N, Milenkovic S, Maricic S, et al. New models for preoperative triage of women with benignant and malignant adnexal tumours. BJOG: An International Journal of Obstetrics and Gynaecology. 2013;120:236.

71. Drake J. Diagnosis and management of the adnexal mass. American Family Physician. 1998;57(10):2471-6, 9-80.

72. Durno CA, Aronson M, Tabori U, Malkin D, Chan H, Gallinger S. Oncologic surveillance for subjects with biallelic mismatch repair gene mutations-10 year follow-up in a kindred. Hereditary Cancer in Clinical Practice. 2011;9:9.

73. Dutta S, Wang F-q, Fleischer AC, Fishman DA. New frontiers for ovarian cancer risk evaluation: proteomics and contrast-enhanced ultrasound. AJR American Journal of Roentgenology. 2010;194(2):349-54.

74. Dutta S, Wang F-q, Phalen A, Fishman DA. Biomarkers for ovarian cancer detection and therapy. Cancer Biology & Therapy. 2010;9(9):668-77.

75. Escudero JM, Auge JM, Filella X, Torne A, Pahisa J, Molina R. Comparison of serum human epididymis protein 4 with cancer antigen 125 as a tumor marker in patients with malignant and nonmalignant diseases. Clinical Chemistry. 2011;57(11):1534-44.

76. Farghaly SA. Current diagnosis and management of ovarian cysts. Clinical & Experimental Obstetrics & Gynecology. 2014;41(6):609-12.

77. Farnaz, Hassan L. Serum CA-125 versus ultrasound examination in differentiating between benign and malignant ovarian masses. JPMI - Journal of Postgraduate Medical Institute. 2012;26(3):291-5.

78. Faschingbauer F, Benz M, Haberle L, Goecke TW, Beckmann MW, Renner S, et al. Subjective assessment of ovarian masses using pattern recognition: the impact of experience on diagnostic performance and interobserver variability. Archives of Gynecology & Obstetrics. 2012;285(6):1663-9.

Table 11. Excluded studies: no 2 × 2 table *(Continued)*

79. Fischerova D, Burgetova A. Imaging techniques for the evaluation of ovarian cancer. Best Practice & Research in Clinical Obstetrics & Gynaecology. 2014;28(5):697-720.

80. Fischerova D, Zikan M, Dundr P, Cibula D. Diagnosis, treatment, and follow-up of borderline ovarian tumors. Oncologist. 2012;17(12):1515-33.

81. Fischerova D, Zikan M, Pinkavova I, Slama S, Fruhauf F, Freitag P, et al. [The rational preoperative diagnosis of ovarian tumors - imaging techniques and tumor biomarkers (review)]. Ceska Gynekologie. 2012;77(4):272-87.

82. Fleischer AC, Lyshchik A, Andreotti RF, Hwang M, Jones HW, 3rd, Fishman DA. Advances in sonographic detection of ovarian cancer: depiction of tumor neovascularity with microbubbles. AJR American Journal of Roentgenology. 2010;194(2):343-8.

83. Fleischer AC, Lyshchik AP, Jones IHW, Fishman DA. Early detection of ovarian cancer with contrast enhanced transvaginal sonography. MedicaMundi. 2009;53(1):49-54+7+9+60+1.

84. Fleischer AC, Rodgers WH, Kepple DM, Williams LL, Jones HW, 3rd. Color Doppler sonography of ovarian masses: a multiparameter analysis. Journal of ultrasound in medicine: official journal of the American Institute of Ultrasound in Medicine. 1993;12(1):41-8.

85. Florinas S, Kim J, Nam K, Janat-Amsbury MM, Kim SW. Ultrasound-assisted siRNA delivery via arginine-grafted bioreducible polymer and microbubbles targeting VEGF for ovarian cancer treatment. Journal of Controlled Release. 2014;183:1-8.

86. Franchi D, Sandri M, Boveri S, Peiretti M, Radice D, Preti E, et al. Ultrasound imaging compared to a multivariate predictive algorithm combining HE4 and CA 125 (roma) in patients with adnexal masses. International Journal of Gynecological Cancer. 2011;3):S364.

87. Fuste P. HE-4 in the diagnosis of abdominal masses. Comparison with ultrasonography and RMI. Tumor Biology. 2014;35:S9.

88. Gasiorowska E, Michalak M, Warchol W, Lemanska A, Jasinski P, Spaczynski M, et al. Clinical application of HE4 and CA125 in ovarian cancer type I and type II detection and differential diagnosis. Ginekologia Polska. 2015;86(2):88-93.

89. Geetha P, Nair MK. Granulosa cell tumours of the ovary. Australian & New Zealand Journal of Obstetrics & Gynaecology. 2010;50(3):216-20.

90. Geomini P, Coppus S, Kluivers KB, Bremer GL, Kruitwagen R, Mol BWJ. Is three-dimensional ultrasonography of additional value in the assessment of adnexal masses? Gynecologic Oncology. 2007;106(1):153-9.

91. Gilbert L, Basso O, Sampalis J, Karp I, Martins C, Feng J, et al. Assessment of symptomatic women for early diagnosis of ovarian cancer: results from the prospective DOvE pilot project. Lancet Oncology. 2012;13(3):285-91.

92. Givens V, Mitchell GE, Harraway-Smith C, Reddy A, Maness DL. Diagnosis and management of adnexal masses. American Family Physician. 2009;80(8):815-20.

93. Goff B. Symptoms Associated With Ovarian Cancer. Clinical Obstetrics and Gynecology. 2012;55(1):36-42.

94. Gotlieb WH, Soriano D, Achiron R, Zalel Y, Davidson B, Kopolovic J, et al. CA 125 measurement and ultrasonography in borderline tumors of the ovary. American journal of obstetrics and gynecology. 2000;183(3):541-6.

95. Goyal S. Assessment of adnexal masses by color Doppler study and serum CA125 assay. BJOG: An International Journal of Obstetrics and Gynaecology. 2014;121:49.

96. Grasso M, Blanco S, Grasso AAC, Nespoli L. Inguinal polypropylene plug: A cause of unusual testicular tumor pelvic metastasis. Archivio Italiano di Urologia e Andrologia. 2013;85(1):47-9.

97. Grewal K, Hamilton W, Sharp D. Ovarian cancer: A scoring system for primary care. BJOG: An International Journal of Obstetrics and Gynaecology. 2012;119:158.

Table 11. Excluded studies: no 2 × 2 table *(Continued)*

98. Griffin N, Grant LA, Sala E. Adnexal masses: characterization and imaging strategies. Seminars in Ultrasound, CT & MR. 2010;31(5):330-46.

99. Gubbels JA, Claussen N, Kapur AK, Connor JP, Patankar MS. The detection, treatment, and biology of epithelial ovarian cancer. Journal of ovarian research. 2010;3:8.

100. Guerriero S, Ajossa S, Garau N, Piras B, Paoletti AM, Melis GB. Ultrasonography and color Doppler-based triage for adnexal masses to provide the most appropriate surgical approach. American journal of obstetrics and gynecology. 2005;192(2):401-6.

100. Guerriero S, Ajossa S, Lai MP, Alcazar JL, Paoletti AM, Marisa O, et al. The diagnosis of functional ovarian cysts using transvaginal ultrasound combined with clinical parameters, CA125 determinations, and color Doppler. European Journal of Obstetrics & Gynecology and Reproductive Biology. 2003;110(1):83-8.

101. 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.

102. Guerriero S, Saba L, Ajossa S, Peddes C, Sedda F, Piras A, et al. Assessing the reproducibility of the IOTA simple ultrasound rules for classifying adnexal masses as benign or malignant using stored 3D volumes. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2013;171(1):157-60.

103. Haiyan H, Min D. Ultrasound operators' confidence influences diagnosis of ovarian tumors - a study in China. Asian Pacific Journal of Cancer Prevention: Apjcp. 2011;12(5):1275-7.

104. Hamilton W, Roobottom C. Early diagnosis of cancer by imaging: The primary care perspective. Radiography. 2012;18(1):5-8.

105. Hamper UM, Sheth S, Abbas FM, Rosenshein NB, Aronson D, Kurman RJ. Transvaginal color Doppler sonography of adnexal masses: differences in blood flow impedance in benign and malignant lesions. AJR American Journal of Roentgenology. 1993;160(6):1225-8.

106. Haoula Z, Deshpande R, Jayaprakasan K, Raine-Fenning N. Doppler imaging in the diagnosis of ovarian disease. Expert Opinion on Medical Diagnostics. 2012;6(1):59-73.

107. Hartman CA, Juliato CRT, Sarian LO, Toledo MC, Jales RM, Morais SS, et al. Ultrasound criteria and CA 125 as predictive variables of ovarian cancer in women with adnexal tumors. Ultrasound in Obstetrics & Gynecology. 2012;40(3):360-6.

108. Hata K, Akiba S, Hata T, Miyazaki K. A multivariate logistic regression analysis in predicting malignancy for patients with ovarian tumors. Gynecologic Oncology. 1998;68(3):256-62.

109. Hata T, Hata K, Noguchi J, Kanenishi K, Shiota A. Ultrasound for evaluation of adnexal malignancy: from 2D to 3D ultrasound. Journal of Obstetrics & Gynaecology Research. 2011;37(10):1255-68.

110. Hellstrom I, Heagerty PJ, Swisher EM, Liu P, Jaffar J, Agnew K, et al. Detection of the HE4 protein in urine as a biomarker for ovarian neoplasms. Cancer Letters. 2010;296(1):43-8.

111. Hellstrom I, Hellstrom KE. Two new biomarkers, mesothelin and HE4, for diagnosis of ovarian carcinoma. Expert Opinion on Medical Diagnostics. 2011;5(3):227-40.

112. Hertlein L, Stieber P, Kirschenhofer A, Krocker K, Nagel D, Lenhard M, et al. Human epididymis protein 4 (HE4) in benign and malignant diseases. Clinical Chemistry & Laboratory Medicine. 2012;50(12):2181-8.

113. Hodeib M, Bristow RE, Smith A, Zhang Z, Chan DW, Fung ET, et al. Impact of a multivariate index assay on referral patterns for surgical management of an adnexal mass. Gynecologic Oncology. 2013;131 (1):258.

114. Hogdall E, Aarnstrup Rasmussen M, Christensen IJ, Lundvall L, Engelholm SA, Nedergaard L, et al. Diagnostic value of he4, CA125 and the roma index in ovarian cancer patients from a tertiary center. International Journal of Gynecological Cancer. 2011;3):S607.

Table 11. Excluded studies: no 2 × 2 table *(Continued)*

115. Holdenrieder S, Molina R, Gion M, Gressner A, Troalen F, Maria Auge J, et al. Alternative antibody for the detection of CA125 antigen: A European multicenter study for the evaluation of the analytical and clinical performance of the Access OV Monitor assay on the UniCel DxI 800 Immunoassay System. Clinical Chemistry and Laboratory Medicine. 2008;46(5):588-99.
116. Horala A, Swiatly A, Lorek J, Kokot ZJ, Matysiak J and Nowak-Markwitz E. Assessment of diagnostic utility of multivariate diagnostic models in differential diagnosis of ovarian tumors. Ginekologia Polska 2018;89(10): 568-572
117. Hossain F, Khan N, Hussain R. Preoperative detection of ovarian cancer by colour Doppler ultrasonography and CA125. BJOG: An International Journal of Obstetrics and Gynaecology. 2012;119:160.
118. Ibrahim M, Bahaa A, Ibrahim A, El Hakem AA, Abo-El Noor A, El Tohamy U. Evaluation of serum mesothelin in malignant and benign ovarian masses. Archives of Gynecology & Obstetrics. 2014;290(1):107-13.
119. Jacobs I, Menon U. The sine qua non of discovering novel biomarkers for early detection of ovarian cancer: carefully selected preclinical samples. Cancer Prevention Research. 2011;4(3):299-302.
120. Jayde V, White K, Blomfield P. Symptoms and diagnostic delay in ovarian cancer: a summary of the literature. Contemporary Nurse. 2009;34(1):55-65.
121. Jelovac D, Armstrong DK. Recent progress in the diagnosis and treatment of ovarian cancer. CA: a Cancer Journal for Clinicians. 2011;61(3):183-203.
122. Kaijser J, Bourne T, Valentin L, Sayasneh A, Van Holsbeke C, Vergote I, et al. Improving strategies for diagnosing ovarian cancer: a summary of the International Ovarian Tumor Analysis (IOTA) studies. Ultrasound in Obstetrics & Gynecology. 2013;41(1):9-20.
123. Kaijser J, Van Gorp T, Van Holsbeke C, Sayasneh A, Vergote I, Bourne T, et al. Diagnostic test performance of CA125, HE4, ROMA and IOTA's LR2 in adnexal tumours of different size. BJOG: An International Journal of Obstetrics and Gynaecology. 2013;120:360-1.
124. Kaijser J, Vandecaveye V, Deroose CM, Rockall A, Thomassin-Naggara I, Bourne T, et al. Imaging techniques for the pre-surgical diagnosis of adnexal tumours. Best Practice & Research in Clinical Obstetrics & Gynaecology. 2014;28(5):683-95.
125. Kalmantis K, Papageorgiou T, Rodolakis A, Lymberopoulos E, Daskalakis G, Voulgaris Z, et al. The role of three-dimensional (3D) sonography and 3D power Doppler in the preoperative assessment of borderline ovarian tumors. European Journal of Gynaecological Oncology. 2007;28(5):381-5.
126. Kazantseva M, Porhanova N, Syrota E, Potemin S. Sensitivity of the serum HE4 and CA-125 combination in the ovarian cancer diagnosis. International Journal of Gynecological Cancer. 2011;3):S415.
127. Kesterson J, Chan JK, Tian C, Monk BJ, Bell JG. Symptoms of women with high-risk early-stage ovarian cancer: A Gynecologic Oncology Group study. Gynecologic Oncology. 2009;112(2):S36-S7.
128. Khan A, Sultana K. Presenting signs and symptoms of ovarian cancer at a tertiary care hospital. JPMA The Journal of the Pakistan Medical Association. 2010;60(4):260-2.
129. Kim J, Lee HN, Park H, Lee DW, Kim MJ, Shin JE. Diagnostic accuracy of risk of ovarian malignancy algorithm (ROMA) experienced in the clinical practice at a single hospital. International Journal of Gynecological Cancer. 2018;28 (Supplement 3):58
130. Kim KH, Alvarez RD. Using a multivariate index assay to assess malignancy in a pelvic mass. Obstetrics & Gynecology. 2012;119(2 Pt 1):365-7.
131. Kim PS, Djazayeri S, Zeineldin R. Novel nanotechnology approaches to diagnosis and therapy of ovarian cancer. Gynecologic Oncology. 2011;120(3):393-403.
132. Kirschenhofer A, Lenhard M, Hertlein L, Furst S, Mayr D, Nagel D, et al. HE4 and CA125in benignandmalignantdiseases. Clinical Chemistry and Laboratory Medicine. 2011;49:S253.

Table 11. Excluded studies: no 2 × 2 table *(Continued)*

133. Knafel A, Nocun A, Banas T, Wiechec M, Jach R, Pietrus M, et al. Iota simple ultrasound-based rules: Why do we have inconclusive results? International Journal of Gynecological Cancer. 2013;1):155-6.

134. Knudsen UB, Tabor A, Mosgaard B, Andersen ES, Kjer JJ, Hahn-Pedersen S, et al. Management of ovarian cysts. Acta Obstetricia et Gynecologica Scandinavica. 2004;83(11):1012-21.

135. Koleva-Topova V. Diagnostic usefulness of multi-marker panels in preoperative prediction of ovarian cancer diagnosis. Clinica Chimica Acta. 2019;493 (Supplement 1):S147.

136. Konforte D, Diamandis EP. Is early detection of cancer with circulating biomarkers feasible? Clinical Chemistry. 2013;59(1):35-7.

137. Kotowicz B, Fuksiewicz M, Kowalska M, Sobiczewski P, Skrzypczak M, Spiewankiewicz B, et al. He4 and ca125 in ovarian cancer patients. International Journal of Gynecological Cancer. 2013;1):923.

138. Kurjak A, Schulman H, Sosic A, Zalud I, Shalan H. Transvaginal ultrasound, color flow, and Doppler waveform of the postmenopausal adnexal mass. Obstetrics & Gynecology. 1992;80(6):917-21.

139. Kurman RJ, Visvanathan K, Roden R, Wu TC, Shih IM. Early detection and treatment of ovarian cancer: shifting from early stage to minimal volume of disease based on a new model of carcinogenesis. American journal of obstetrics and gynecology. 2008;198(4):351-6.

140. Lachance JA, Choudhri AF, Sarti M, Modesitt SC, Jazaeri AA, Stukenborg GJ. A nomogram for estimating the probability of ovarian cancer. Gynecologic Oncology. 2011;121(1):2-7.

141. Lalwani N, Dubinsky TJ. Clinically suspected adnexal mass. Ultrasound Quarterly. 2013;29(1):87-8.

142. Langhe R, Norris L, Petzold M, Abu Saadeh F, Kamran W, Ibrahim N, et al. Roma index improves specificity of ovarian cancer diagnosis compared to CA125 alone. International Journal of Gynecological Cancer. 2013;1):588.

143. Langmar Z, Nemeth M, Vlesko G, Kiraly M, Hornyak L, Bosze P. HE4--a novel promising serum marker in the diagnosis of ovarian carcinoma. European Journal of Gynaecological Oncology. 2011;32(6):605-10.

144. Lawicki S, Bedkowska GE, Gacuta-Szumarska E, Szmitkowski M. The plasma concentration of VEGF, HE4 and CA125 as a new biomarkers panel in different stages and sub-types of epithelial ovarian tumors. Journal of ovarian research. 2013;6(1):45.

145. Leandersson P, Kalapotharakos G, Henic E, Borgfeldt H, Petzold M, Hoyer-Hansen G, et al. A Biomarker Panel Increases the Diagnostic Performance for Epithelial Ovarian Cancer Type I and II in Young Women. Anticancer Research. 2016;36(3):957-65

146. Lee TS, Kim JW, Park NH, Song YS, Kang SB, Lee HP. Assessing clinical performance of gynecology residents: sonographic evaluation of adnexal masses based on morphological scoring systems. Ultrasound in Obstetrics & Gynecology. 2005;26(7):776-9.

147. Leeners B, Schild RL, Funk A, Hauptmann S, Kemp B, Schroder W, et al. Colour Doppler sonography improves the pre-operative diagnosis of ovarian tumours made using conventional transvaginal sonography. European Journal of Obstetrics Gynecology and Reproductive Biology. 1996;64(1):79-85.

148. Lerner JP, Timor-Tritsch IE, Federman A, Abramovich G. Transvaginal ultrasonographic characterization of ovarian masses with an improved, weighted scoring system. American Journal of Obstetrics & Gynecology. 1994;170(1 Pt 1):81-5.

149. Levine D, Feldstein VA, Babcook CJ, Filly RA. Sonography of ovarian masses: poor sensitivity of resistive index for identifying malignant lesions. AJR American Journal of Roentgenology. 1994;162(6):1355-9.

150. Li AJ. New biomarkers for the diagnosis of ovarian carcinoma: OVA1 and ROMA. [Italian] Nuovi biomarcatori per la diagnosi di carcinoma ovarico: OVA1 e ROMA. Giornale Italiano di Ostetricia e Ginecologia. 2012;34(3):409-14.

151. Li J, Dowdy S, Tipton T, Podratz K, Lu W-G, Xie X, et al. HE4 as a biomarker for ovarian and endometrial cancer management. Expert Review of Molecular Diagnostics. 2009;9(6):555-66.

Table 11. Excluded studies: no 2 × 2 table *(Continued)*

152. Likic I, Terzic MM, Bila J, Pilic I, Berisavac M, Dotlic J, et al. Adnexal masses: Reliability of serum tumor marker levels. International Journal of Gynecological Cancer. 2011;3):S363.

153. Lin JY, Qin JB, Sangvatanakul V. Human epididymis protein 4 for differential diagnosis between benign gynecologic disease and ovarian cancer: a systematic review and meta-analysis (Provisional abstract). European Journal of Obstetrics and Gynecology and Reproductive Biology [Internet]. 2013 [cited DARE Y/U]; (1):[81-5 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/cldare/articles/DARE-12013017358/frame.html.

154. Liu JH, Zanotti KM. Management of the adnexal mass. Obstetrics & Gynecology. 2011;117(6):1413-28.

155. Lutz AM, Willmann JK, Drescher CW, Ray P, Cochran FV, Urban N, et al. Early diagnosis of ovarian carcinoma: is a solution in sight? Radiology. 2011;259(2):329-45.

156. Macuks R, Baidekalna I, Donina S. An ovarian cancer malignancy risk index composed of HE4, CA125, ultrasonographic score, and menopausal status: use in differentiation of ovarian cancers and benign lesions. Tumour Biology. 2012;33(5):1811-7.

157. Malek M. Imaging strategy for diagnosis of adnexal masses including US,CT Scan and MRI. International Journal of Fertility and Sterility Conference: 13th Congress on Reproductive Biomedicine and 7th Royan Nursing and Midwifery Seminar Tehran Iran, Islamic Republic of Conference Start. 2012;6(pp 27).

158. Manganaro L, Yazdanian D, Vinci V, Imperiale L, Anastasi E, Porpora MG. Avoiding inappropriate surgery on undetermined mass: May the combination of Magnetic Resonance Imaging (MRI) and ovarian markers be helpful? Journal of Minimally Invasive Gynecology. 2014;1):S69.

159. Manolov V, Avramova V, Radenkov R. He 4 - A new marker for ovarian cancer. Clinical Chemistry and Laboratory Medicine. 2011;49:S254.

160. Mark DH. Multianalyte testing for the evaluation of adnexal masses (Provisional abstract). Database of Abstracts of Reviews of Effects [Internet]. 2013 [cited DARE Y/U]; (1):[1 p.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/cldare/articles/DARE-12013062634/frame.html.

161. Marret H, Ecochard R, Giraudeau B, Golfier F, Raudrant D, Lansac J. Color Doppler energy prediction of malignancy in adnexal masses using logistic regression models. Ultrasound in Obstetrics & Gynecology. 2002;20(6):597-604.

162. Marret H, Sauget S, Giraudeau B, Brewer M, Ranger-Moore J, Body G, et al. Contrast-enhanced sonography helps in discrimination of benign from malignant adnexal masses. Journal of Ultrasound in Medicine. 2004;23(12):1629-39; quiz 41-42.

163. McGuckin MA, Ramm LE, Joy GJ, Free KE, Ward BG. Preoperative discrimination between ovarian-carcinoma, non-ovarian gynecological malignancy and benign adnexal masses using serum levels of ca125 and the polymorphic epithelial mucin antigens CASA, OSA AND MSA. International Journal of Gynecological Cancer. 1992;2(3):119-28.

164. McGuire WP. HE4--another marker for gynecologic cancers: do we really need one? Oncology (Williston Park). 2013;27(6):563, 6.

165. McIntosh MW, Drescher C, Karlan B, Scholler N, Urban N, Hellstrom KE, et al. Combining CA 125 and SMR serum markers for diagnosis and early detection of ovarian carcinoma. Gynecol Oncol. 2004;95(1):9-15.

166. Medeiros LR, Rosa DD, da Rosa MI, Bozzetti MC. Accuracy of ultrasonography with color Doppler in ovarian tumor: a systematic quantitative review. International Journal of Gynecological Cancer. 2009;19(7):1214-20.

167. Medeiros LR, Rosa DD, da Rosa MI, Bozzetti MC. Accuracy of CA 125 in the diagnosis of ovarian tumors: a quantitative systematic review. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2009;142(2):99-105.

168. Miller RW, Ueland FR. Risk of malignancy in sonographically confirmed ovarian tumors. Clinical Obstetrics & Gynecology. 2012;55(1):52-64.

169. Miller RW, van Nagell JR. Preoperative evaluation of adnexal masses. Women's health. 2011;7(1):37-9.

Table 11. Excluded studies: no 2 × 2 table *(Continued)*

170. Mohaghegh P, Rockall AG. Imaging strategy for early ovarian cancer: characterization of adnexal masses with conventional and advanced imaging techniques. Radiographics. 2012;32(6):1751-73.
171. Molina R, Escudero JM, Auge JM, Filella X. Evaluation of serum human epididymis protein 4 (HE4) in patients with malignant and non malignant diseases: Comparison with CA 125. Clinical Chemistry and Laboratory Medicine. 2011;49:S259.
172. Molina R, Escudero JM, Auge JM, Filella X, Ferro L, Torne A, et al. HE-4, CA 125 and ROMA in the differential diagnosis of patients with gynecological diseases. Tumor Biology. 2011;32:S17-S8.
173. Molina R, Escudero JM, Auge JM, Filella X, Foj L, Torne A, et al. He4, CA 125 and roma in the differential diagnosis of patients with gynecological diseases. Clinical Chemistry and Laboratory Medicine. 2011;49:S258.
174. Molina R, Escudero JM, Auge JM, Filella X, Torne A, Pahisa J, et al. CA 19.9 and useful complementary tumor marker in ovarian cancer: Comparison with HE4, CA 125 and ROMA in Gynaecological diseases. Tumor Biology. 2012;33:S39.
175. Molina R, Escudero JM, Fuste P. HE-4 levels in gynaecological patients undergoing surgical treatment for suspected malignancies. Systems to increase efficiency. Tumor Biology. 2014;35:S9.
176. Molina R, Stieber P, Heinz-Jurgen R, Escudero JM, Auge JM, Filella X. A european multicenter study for the evaluation of the analytical and clinical performance of the Elecsys HE-4 assay on Cobas e411 from roche diagnostics. Tumor Biology. 2012;33:S79.
177. Molina R, Fernandez-Bonifacio R, Mellado B, Fernandez-Galan E, Carreras N, Rodriguez I et al. New algorithm ROMA (risk of ovarian malignancy algorithm)using CA 125 from Siemens and HE4 from Roche. Clinica Chimica Acta. 2019;493 (Supplement 1): S155
178. Momtaz M. 3D ultrasonography of ovarian tumors. International Journal of Fertility and Sterility Conference: 15th Congress on Reproductive Biomedicine and 9th Royan Nursing and Midwifery Seminar Tehran Iran, Islamic Republic of Conference Start. 2014;8(pp 23-24).
179. Montagnana M, Lippi G, Ruzzenente O, Bresciani V, Danese E, Scevarolli S, et al. The utility of serum human epididymis protein 4 (HE4) in patients with a pelvic mass. Journal of Clinical Laboratory Analysis. 2009;23(5):331-5.
180. Moore RG, Bast Jr RC. How do you distinguish a malignant pelvic mass from a benign pelvic mass? Imaging, biomarkers, or none of the above. Journal of Clinical Oncology. 2007;25(27):4159-61.
181. Moore RG, Maclaughlan S. Current clinical use of biomarkers for epithelial ovarian cancer. Current Opinion in Oncology. 2010;22(5):492-7.
182. Moore RG, MacLaughlan S, Bast RC, Jr. Current state of biomarker development for clinical application in epithelial ovarian cancer. Gynecologic Oncology. 2010;116(2):240-5.
183. Morotti M, Menada MV, Gillott DJ, Venturini PL, Ferrero S. The preoperative diagnosis of borderline ovarian tumors: a review of current literature. Archives of Gynecology & Obstetrics. 2012;285(4):1103-12.
184. Moszynski R, Zywica P, Wojtowicz A, Szubert S, Sajdak S, Stachowiak A, et al. Menopausal status strongly influences the utility of predictive models in differential diagnosis of ovarian tumors: an external validation of selected diagnostic tools. Ginekologia Polska. 2014;85(12):892-9.
185. Mozaffari-Kermani R, Montaser-Kouhsari L. The role of pathologic results and clinical presentations in predictive value of transvaginal ultrasonographic findings: differentiating benign from malignant ovarian neoplasms. Virchows Archiv. 2009;455:310-1.
186. Murala KS, Ma K, Henry RJW, Snape S, Dwivedi R. Performance of IOTA simple rules and RMI in preoperative classification of adnexal lesions in DGH setting. BJOG: An International Journal of Obstetrics and Gynaecology. 2014;121:49-50.
187. Murta EFC, da Silva CS, Gomes RAS, Tavares-Murta BM, Melo ALKO. Ultrasonographic criteria and tumor marker assay are good procedures for the diagnosis of ovarian neoplasia in preselected outpatients. European Journal of Gynaecological Oncology. 2004;25(6):707-12.

Table 11. Excluded studies: no 2 × 2 table *(Continued)*

188. Murta EFC, Nomelini RS. Early diagnosis and predictors of malignancy of adnexal masses. Current Opinion in Obstetrics & Gynecology. 2006;18(1):14-9.

189. Myers ER, Bastian LA, Havrilesky LJ, Kulasingam SL, Terplan MS, Cline KE, et al. Management of adnexal mass. Evidence Report/Technology Assessment. 2006(130):1-145.

190. Nagy B, Krasznai ZT, Csoban M, Antal-Szalmas P, Hernadi Z, Kappelmayer J. The emerging value of HE4 in the discrimination of ovarian cancer from other malignant and benign pelvic masses. Clinical Chemistry and Laboratory Medicine. 2011;49:S261.

191. Negishi Y, Iwabuchi H, Sakunaga H, Sakamoto M, Okabe K, Sato H, et al. Serum and tissue measurements of CA72-4 in ovarian cancer patients. Gynecologic Oncology. 1993;48(2):148-54.

192. Nikolova T, Dimitrov G, Antovska V, Eftimovska N, Nikolova N, Janusevska A. A comparison of HE4 with CA 125 as tumor markers in patients with ovarian cancer. International Journal of Gynecological Cancer. 2013;1):586.

193. Nogales FF, Dulcey I, Preda O. Germ cell tumors of the ovary: an update. Archives of Pathology & Laboratory Medicine. 2014;138(3):351-62.

194. Nolen BM, Lokshin AE. Multianalyte assay systems in the differential diagnosis of ovarian cancer. Expert Opinion on Medical Diagnostics. 2012;6(2):131-8.

195. Nolen BM, Lokshin AE. Biomarker testing for ovarian cancer: clinical utility of multiplex assays. Molecular Diagnosis & Therapy. 2013;17(3):139-46.

196. Ohel I, Sheiner E, Aricha-Tamir B, Piura B, Meirovitz M, Silberstein T, et al. Three-dimensional power Doppler ultrasound in ovarian cancer and its correlation with histology. Archives of Gynecology & Obstetrics. 2010;281(5):919-25.

197. Ong N, Zakaria AR. Frozen section: The gatekeeper for radical ovarian cancer surgery. Medical Journal of Malaysia. 2018;73 (Supplement 1):48.

198. Onsrud M, Shabana A, Austgulen R. Soluble tumor necrosis factor receptors and CA 125 in serum as markers for epithelial ovarian cancer. Tumor Biology. 1996;17(2):90-6.

199. Osmers RG, Osmers M, von Maydell B, Wagner B, Kuhn W. Evaluation of ovarian tumors in postmenopausal women by transvaginal sonography. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 1998;77(1):81-8.

200. Osmers RGW, Kuhn W. Early diagnosis of ovarian cancer using ultrasonography. Archives of Gynecology and Obstetrics. 1997;260(1-4):351-5.

201. Ozan H, Ozsoy S, Baykara S, Celik N. Review of 28 ovarian granulosa cell tumor cases. International Journal of Gynecological Cancer. 2011;3):S657.

202. Park Y, Lee J-H, Hong DJ, Lee EY, Kim H-S. Diagnostic performances of HE4 and CA125 for the detection of ovarian cancer from patients with various gynecologic and non-gynecologic diseases. Clinical Biochemistry. 2011;44(10-11):884-8.

203. Pitta DdR, Sarian LO, Campos EA, Andrade LLDA, Sallum LF, Braganca JF, et al. HE4 can help discriminate women with malignant ovarian tumors only if CA125 levels are elevated. International Journal of Biological Markers. 2013;28(4):e377-86.

204. Presl J, Sefrhansova L, Sedlakova I, Vlasak P, Bouda J, Bartakova A, et al. Possibilities of preoperative prediction of the biological behavior of gynecological pelvic tumors. [Czech]

205. Moznosti poedoperaeni predikce biologickeho chovani gynekologickych panevnich tumoru. Aktualni Gynekologie a Porodnictvi. 2014;6:53-9.

206. Prompeler HJ. [Ultrasonographic clarification of adnexal findings]. Radiologe. 2011;51(7):568-80.

Table 11. Excluded studies: no 2 × 2 table *(Continued)*

207. Qiao J-J, Yu J, Yu Z, Li N, Song C, Li M. Contrast-enhanced ultrasonography in differential diagnosis of benign and malignant ovarian tumors. PLoS ONE [Electronic Resource]. 2015;10(3):e0118872.

208. Quaranta M, Mehra G, Nath R and Sayasneh A. Why a benign ovarian mass can still undergo surgery by a gynaecological cancer surgeon? Australasian Journal of Ultrasound in Medicine. 2019;22 (2): 146.

209. Redman C, Duffy S, Dobson C. Improving early detection of ovarian cancer. Practitioner. 2011;255(1741):27-30, 3.

210. Reina H, Buchel J, Butenschon A, Vigo F, Schoetzau A, Heinzelmann-Schwarz V, et al. Estimating risk of malignancy in adnexal masses with ultrasound: A retrospective diagnostic accuracy study. Geburtshilfe und Frauenheilkunde Conference. 2018;78(10).

211. Riman T, Dickman PW, Nilsson S, Correia N, Nordlinder H, Magnusson CM, et al. Risk factors for invasive epithelial ovarian cancer: results from a Swedish case-control study. American journal of epidemiology. 2002;156(4):363-73.

212. Romagnolo C, Trivella G, Bonacina M, Fornale M, Maggino T, Ferrazzi E. Preoperative diagnosis of 221 consecutive ovarian masses: scoring system and expert evaluation. European Journal of Gynaecological Oncology. 2006;27(5):487-9.

213. Roman LD, Muderspach LI, Stein SM, LaiferNarin S, Groshen S, Morrow CP. Pelvic examination, tumor marker level, and gray-scale and Doppler sonography in the prediction of pelvic cancer. Obstetrics and Gynecology. 1997;89(4):493-500.

214. Ronco DA, Manahan KJ, Geisler JP. Ovarian cancer risk assessment: a tool for preoperative assessment. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2011;158(2):325-9.

215. Salem S, White LM, Lai J. Doppler sonography of adnexal masses: the predictive value of the pulsatility index in benign and malignant disease. AJR American Journal of Roentgenology. 1994;163(5):1147-50.

216. Sandhu N, Karlsen MA, Hogdall C, Laursen IA, Christensen IJ, Hogdall EVS. Stability of HE4 and CA125 in blood samples from patients diagnosed with ovarian cancer. Scandinavian Journal of Clinical & Laboratory Investigation. 2014;74(6):477-84.

217. Sasaroli D, Coukos G, Scholler N. Beyond CA125: The coming of age of ovarian cancer biomarkers. Biomarkers in Medicine. 2009;3(3):275-88.

218. Sassone AM, Timor-Tritsch IE, Artner A, Westhoff C, Warren WB. Transvaginal sonographic characterization of ovarian disease: evaluation of a new scoring system to predict ovarian malignancy. Obstetrics & Gynecology. 1991;78(1):70-6.

219. Sayasneh A, Preisler J, Stlader C, Husicka R, Naji O, Kaijser J, et al. A randomised controlled trial to compare the clinical impact of RMI versus LR2 to characterise adnexal masses: Interim analysis of phase 4 IOTA study. Bjog [Internet]. 2013 [cited CENTRAL Y/U]:[357-8 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/086/CN-01007086/frame.html http://onlinelibrary.wiley.com/store/10.1111/1471-0528.12343/asset/bjo12343.pdf?v=1&t=ic4hlb2n&s=3c47dc39d6bc29d1bd842be48b583b830e2f86c6.

220. Scholler N, Crawford M, Sato A, Drescher CW, O'Briant KC, Kiviat N, et al. Bead-based ELISA for validation of ovarian cancer early detection markers. Clinical Cancer Research. 2006;12(7 I):2117-24.

221. Schutter EM, Mijatovic V, Kok A, Van Kamp GJ, Verstraeten R, Verheijen RH. Urinary gonadotropin peptide (UGP) and serum CA 125 in gynaecologic practice, a clinical prospective study. Anticancer Research. 1999;19(6C):5551-7.

222. Schutter EM, Sohn C, Kristen P, Mobus V, Crombach G, Kaufmann M, et al. Estimation of probability of malignancy using a logistic model combining physical examination, ultrasound, serum CA 125, and serum CA 72-4 in postmenopausal women with a pelvic mass: an international multicenter study. Gynecologic Oncology. 1998;69(1):56-63.

223. Sconfienza LM, Perrone N, Delnevo A, Lacelli F, Murolo C, Gandolfo N, et al. Diagnostic value of contrast-enhanced ultrasonography in the characterization of ovarian tumors(). Journal of Ultrasound. 2010;13(1):9-15.

224. Simmons AR, Baggerly K, Bast RC, Jr. The emerging role of HE4 in the evaluation of epithelial ovarian and endometrial carcinomas. Oncology (Williston Park). 2013;27(6):548-56.

Table 11. Excluded studies: no 2 × 2 table *(Continued)*

225. Simon I, Katsaros D, de la Longrais IR, Massobrio M, Scorilas A, Kim NW, et al. B7-H4 is over-expressed in early-stage ovarian cancer and is independent of CA125 expression. Gynecologic Oncology. 2007;106(2):334-41.

226. Singh D, Lowe K, Colllins C, Cohen L, Dungan J, Shulman L, et al. Utility of symptom index in women at increased risk for ovarian cancer. Gynecologic Oncology. 2011;121(1):S61-S.

227. Sinha A, Drews F, Lim K, Pugh ND. Retrospective analysis of suspicious pelvic masses using the Pelvic Mass Index (PMI) scoring system from 2007 to 2014. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2016;201:79-84.

228. Smitz J, Ron-El R, Tarlatzis BC. The use of gonadotrophin releasing hormone agonists for in vitro fertilization and other assisted procreation techniques: Experience from three centres. Human Reproduction. 1992;7(SUPPL. 1):49-66.

229. Stanculescu R, Russu M, Vladescu T, Bausic V, Teodorescu A. The combined value of CA125 determination and echographic ovarian imaging, in preoperative diagnosis of benign, borderline and malignant ovarian tumors. Giornale Italiano di Ostetricia e Ginecologia. 2012;34(1):311-3.

230. Stavros S, Drakakis P, Haidopoulos D, Konomos G, Rodolakis A, Mesogitis S, et al. 3D image analysis; Involvement in the preoperative characterization of ovarian masses as benign or malignant. International Journal of Gynecological Cancer. 2014;2).

231. Stein SM, Laifer-Narin S, Johnson MB, Roman LD, Muderspach LI, Tyszka JM, et al. Differentiation of benign and malignant adnexal masses: relative value of gray-scale, color Doppler, and spectral Doppler sonography. AJR American Journal of Roentgenology. 1995;164(2):381-6.

232. Stieber P, Kirschenhofer A, Furst S, Nagel D, Hertlein L, Lenhard M, et al. The ROMA score and the influencing factors upon the diagnostic accuracy. Tumor Biology. 2011;32:S19.

233. Stukan M, Dudziak M, Ratajczak K, Grabowski JP. Usefulness of diagnostic indices comprising clinical, sonographic, and biomarker data for discriminating benign from malignant ovarian masses. Journal of Ultrasound in Medicine. 2015;34(2):207-17.

234. Su Z, Graybill WS, Zhu Y. Detection and monitoring of ovarian cancer. Clinica Chimica Acta. 2013;415:341-5.

235. Suh KS, Park SW, Castro A, Patel H, Blake P, Liang M, et al. Ovarian cancer biomarkers for molecular biosensors and translational medicine. Expert Review of Molecular Diagnostics. 2010;10(8):1069-83.

236. Sumpaico WW. Comparison of roma to rmi for ovarian carcinoma in asia. International Journal of Gynecology and Obstetrics. 2012;119:S248-S9.

237. Tantipalakorn C, Wanapirak C, Khunamornpong S, Sukpan K, Tongsong T. IOTA simple rules in differentiating between benign and malignant ovarian tumors. Asian Pacific Journal of Cancer Prevention: Apjcp. 2014;15(13):5123-6.

238. Terzic MM, Dotlic J, Likic I, Ladjevic N, Brndusic N, Arsenovic N, et al. Current diagnostic approach to patients with adnexal masses: which tools are relevant in routine praxis? Chinese Journal of Cancer Research. 2013;25(1):55-62.

239. Timmerman D, Ameye L, Fischerova D, Epstein E, Melis GB, Guerriero S, et al. Simple ultrasound rules to distinguish between benign and malignant adnexal masses before surgery: Prospective validation by IOTA group. Bmj. 2011;342(7788):94.

240. Timmerman D, Valentin L, Bourne TH, Collins WP, Verrelst H, Vergote I, et al. Terms, definitions and measurements to describe the sonographic features of adnexal tumors: a consensus opinion from the International Ovarian Tumor Analysis (IOTA) Group. Ultrasound in Obstetrics & Gynecology. 2000;16(5):500-5.

241. Twickler DM, Moschos E. Ultrasound and assessment of ovarian cancer risk. AJR American Journal of Roentgenology. 2010;194(2):322-9.

242. Ueland F, Goodrich S, Desimone C, Seamon L, Miller R, Podzielinski I, et al. Predicting the risk of malignancy for an ovarian tumor by combining imaging and OVA1. International Journal of Gynecological Cancer. 2011;3):S115.

Table 11. Excluded studies: no 2 × 2 table *(Continued)*

243. Utrilla-Layna J, Olartecoechea B, Auba M, Ruiz-Zambrana A, Pineda L, Alcazar JL. Predicting malignancy in entirely solid-appearing adnexal masses on gray-scale ultrasound based on additional ultrasound findings, clinical complaints and biochemical parameters: A retrospective study. Donald School Journal of Ultrasound in Obstetrics and Gynecology. 2013;7(1):80-5.

244. Vaes E, Manchanda R, Autier P, Nir R, Nir D, Bleiberg H, et al. Differential diagnosis of adnexal masses: sequential use of the risk of malignancy index and HistoScanning, a novel computer-aided diagnostic tool. Ultrasound in Obstetrics & Gynecology. 2012;39(1):91-8.

245. Valentin L, Ameye L, Jurkovic D, Metzger U, Lecuru F, Van Huffel S, et al. Which extrauterine pelvic masses are difficult to correctly classify as benign or malignant on the basis of ultrasound findings and is there a way of making a correct diagnosis? Ultrasound in Obstetrics & Gynecology. 2006;27(4):438-44.

246. Van Calster B, Valentin L, Van Holsbeke C, Zhang J, Jurkovic D, Lissoni AA, et al. A novel approach to predict the likelihood of specific ovarian tumor pathology based on serum CA-125: a multicenter observational study. Cancer Epidemiology, Biomarkers & Prevention. 2011;20(11):2420-8.

247. van Nagell JR, Jr. Early detection of ovarian cancer in symptomatic women. Lancet Oncology. 2012;13(3):223-4.

248. Van Trappen PO, Rufford BD, Mills TD, Sohaib SA, Webb JAW, Sahdev A, et al. Differential diagnosis of adnexal masses: risk of malignancy index, ultrasonography, magnetic resonance imaging, and radioimmunoscintigraphy. International Journal of Gynecological Cancer. 2007;17(1):61-7.

249. Veyer L, Marret H, Bleuzen A, Simon E, Body G, Tranquart F. Preoperative diagnosis of ovarian tumors using pelvic contrast-enhanced sonography. Journal of Ultrasound in Medicine. 2010;29(7):1041-9.

250. Vranes HS, Klaric P, Sonicki Z, Gall V, Jukic M, Vukovic A. Prediction of ovarian tumor malignancy. Collegium Antropologicum. 2011;35(3):775-9.

251. Weinberger V, Minar L. Diagnostics of malign ovarian tumors by ultrosound and CA 125-our experience. International Journal of Gynecological Cancer. 2013;1:498.

252. Wu H-H, Wang P-H, Yeh J-Y, Chen Y-J, Yen M-S, Huang R-L, et al. Serum cytokeratin-19 fragment (Cyfra 21-1) is a prognostic indicator for epithelial ovarian cancer. Taiwanese Journal of Obstetrics & Gynecology. 2014;53(1):30-4.

253. Wynants L, Timmerman D, Verbakel JY, Testa A, Savelli L, Fischerova D, et al. Clinical Utility of Risk Models to Refer Patients with Adnexal Masses to Specialized Oncology Care: Multicenter External Validation Using Decision Curve Analysis. Clinical Cancer Research. 2017; 23(17): 5082-5090.

254. Wynn ML, Chang S, Peipins LA. Temporal patterns of conditions and symptoms potentially associated with ovarian cancer. Journal of women's health (2002). 2007;16(7):971-86.

255. Yang Z, Luo Z, Zhao B, Zhang W, Zhang J, Li Z, et al. Diagnosis and preoperative predictive value of serum HE4 concentrations for optimal debulking in epithelial ovarian cancer. Oncology Letters. 2013;6(1):28-34.

256. Yasmin SY, A.; Asif, M. Frequency of benign and malignant ovarian tumours in a tertiary care hospital. J Postgrad Med Inst. 2006;20:393-7.

257. Yavuzcan A, Caglar M, Ozgu E, Ustun Y, Dilbaz S, Ozdemir I, et al. Addition of parity to the risk of malignancy index score in evaluating adnexal masses. Taiwanese Journal of Obstetrics & Gynecology. 2014;53(4):518-22.

258. Yazbek J, Raju SK, Ben-Nagi J, Holland TK, Hillaby K, Jurkovic D. Effect of quality of gynaecological ultrasonography on management of patients with suspected ovarian cancer: a randomised controlled trial. The Lancet Oncology [Internet]. 2008 [cited UPDATE (1991-2008) Y]; (2):[124-31 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/305/CN-00622305/frame.htmlhttp://ac.els-cdn.com/S1470204508700056/1-s2.0-S1470204508700056-main.pdf?_tid=b21772ca-620a-11e5-8f-bf-00000aacb362&acdnat=1443023558_910592ab86216e8621e934ba16a60306.

Table 11. Excluded studies: no 2 × 2 table *(Continued)*

259. Yilmaz Y, Demirel G, Ulu HO, Celik IH, Suna Oguz S, Erdeve O, et al. Four neonates with giant ovarian cysts: difficulties in diagnosis and decision making process. Journal of Maternal-Fetal & Neonatal Medicine. 2012;25(8):1508-10.
260. Yoruk P, Dundar O, Yildizhan B, Tutuncu L, Pekin T. Comparison of the risk of malignancy index and self-constructed logistic regression models in preoperative evaluation of adnexal masses. Journal of Ultrasound in Medicine. 2008;27(10):1469-77.
261. Yoruk P, Dundar O, Yildizhan B, Tutuncu L, Pekin T. Comparison of the risk of malignancy index and self-constructed logistic regression models in preoperative evaluation of adnexal masses. Journal of Ultrasound in Medicine. 2008;27(10):1469-77.
262. Yu S, Lee JK, Kim JH, Park H, Lee MY, et al. Diagnostic performance and establishment of reference limits of HE4 in Korean healthy women. Gynecologic Oncology. 2016;143(1): 128-134.
263. Zhang L, Cheng N, Ding J, Liu Z. The clinical value of HE4 in differential diagnosis of ovarian cancer and gynecological pelvic benign diseases. Biochimica Clinica. 2013;37:S203.
264. Zhang L, Liu Z. HE4AS specific marker for ovarian cancer: Comparison with CA125 in differential diagnosis of ovarian cancer and gynecological pelvic benign diseases. Clinical Chemistry and Laboratory Medicine. 2011;49:S282.
265. Zourou I, Beretouli E, Touplikioti P, Zioga C, Maroufidou T, Voutsas M, et al. HE4 a novel human biomarker for ovarian cancer. Cytopathology. 2014;25:69.

Table 12. Excluded studies: data not reported separately by menopausal status

Data not reported separately by menopausal status
1. Anonymous. Comparison of HE4, CA125, and risk of ovarian malignancy algorithm (ROMA) in the prediction of ovarian cancer in Korean women. International Journal of Gynecological Cancer. 2015;1:991.
2. Abdalla N, Piorkowski R, Bachanek M, Stanirowski P, Cendrowski K, Sawicki W. Does the risk of ovarian malignancy algorithm provide better diagnostic performance than HE4 and CA125 in the presurgical differentiation of adnexal tumors in Polish women? Disease Markers. 2018;2018 (no pagination)(5289804).
3. Akturk E, Karaca RE, Alanbay I, Dede M, Karasahin E, Yenen MC, et al. Comparison of four malignancy risk indices in the detection of malignant ovarian masses. Journal of Gynecologic Oncology. 2011;22(3):177-82.
4. Alcazar JL, Pascual MA, Graupera B, Auba M, Errasti T, Olartecoechea B, et al. External validation of IOTA simple descriptors and simple rules for classifying adnexal masses. Ultrasound in Obstetrics & Gynecology. 2016;48(3):397-402.
5. Arun-Muthuvel V, Jaya V. Pre-operative evaluation of ovarian tumors by risk of malignancy index, CA125 and ultrasound. Asian Pacific Journal of Cancer Prevention: Apjcp 2014;15(6):2929-32.
6. Asif N, Sattar A, Dawood MM, Rafi T, Aamir M, Anwar M. Pre-operative evaluation of ovarian mass: risk of malignancy index. Jcpsp, Journal of the College of Physicians & Surgeons - Pakistan 2004;14(3):128-31
7. Aslam N, Tailor A, Lawton F, Carr J, Savvas M, Jurkovic D. Prospective evaluation of three different models for the pre-operative diagnosis of ovarian cancer. BJOG: An International Journal of Obstetrics & Gynaecology. 2000;107(11):1347-53.
8. Borges A, Rodrigues S, Aguino J, Bernardo M, Mahomed F, Djokovic D. Performance of the IOTA ADNEX model in preoperative discrimination of adnexal formations: A Portuguese prospective multicenter pilot study. Australasian Journal of Ultrasound in Medicine. 2019;22 (2):145.
5. Bouzari Z, Yazdani S, Ahmadi MH, Barat S, Kelagar ZS, Kutenaie MJ, et al. Comparison of three malignancy risk indices and CA-125 in the preoperative evaluation of patients with pelvic masses. BMC Research Notes. 2011;4:206.

Table 12. Excluded studies: data not reported separately by menopausal status *(Continued)*

9. Campos C, Sarian LO, Jales RM, Hartman C, Araujo KG, Pitta D, et al. Performance of the Risk of Malignancy Index for Discriminating Malignant Tumors in Women With Adnexal Masses. Journal of Ultrasound in Medicine. 2016;35(1):143-52.

10. Chen H, Qian L, Jiang M, Du Q, Yuan F, Feng W. IOTA ADNEX model for evaluating adnexal masses using data from a gynecologic oncology center in China. Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2019;01.

11. Cho HY, Park SH, Park YH, Kim HB, Kang JB, Hong SH, et al. Comparison of HE4, CA125, and Risk of Ovarian Malignancy Algorithm in the Prediction of Ovarian Cancer in Korean Women. Journal of Korean Medical Science. 2015;30(12):1777-83.

12. Clarke SE, Grimshaw R, Rittenberg P, Kieser K, Bentley J. Risk of malignancy index in the evaluation of patients with adnexal masses. Journal of Obstetrics & Gynaecology Canada: JOGC. 2009;31(5):440-5.

13. Cymbaluk-Ploska A, Chudecka-Glaz A, Surowiec A, Pius-Sadowska E, Machalinski B, Menkiszak J. MMP3 in Comparison to CA 125, HE4 and the ROMA Algorithm in Differentiation of Ovarian Tumors. Asian Pacific Journal of Cancer Prevention: Apjcp. 2016;17(5):2597-603.

14. Davies AP, Jacobs I, Woolas R, Fish A, Oram D. The adnexal mass: benign or malignant? Evaluation of a risk of malignancy index. Br J Obstet Gynaecol. 1993;100(10):927-31.

15. Devey A, Lukas P, Cavalier E. Retrospective evaluation of CA125, HE4 and ROMA index values in a consecutive population of 124 women. Clinica Chimica Acta. 2019;493 (Supplement 1):S134.

16. Di Legge A, Testa AC, Ameye L, Van Calster B, Leone F, Savelli L, et al. Lesion size affects the diagnostic performance of the international ovarian tumor analysis (IOTA) logistic regression models, the IOTA simple rules and the risk of malignancy index to estimate the risk of malignancy in adnexal masses. International Journal of Gynecology and Obstetrics. 2012;119:S741.

17. Dotlic J, Terzic M, Likic I, Atanackovic J, Ladjevic N. Evaluation of adnexal masses: correlation between clinical, ultrasound and histopathological findings. Vojnosanitetski Pregled. 2011;68(10):861-6.

18. Enakpene CA, Omigbodun AO, Goecke TW, Odukogbe A-T, Beckmann MW. Preoperative evaluation and triage of women with suspicious adnexal masses using risk of malignancy index. Journal of Obstetrics & Gynaecology Research. 2009;35(1):131-8.

19. Hogdall E. Approaches to the detection of ovarian cancer. Scandinavian Journal of Clinical and Laboratory Investigation Supplement. 2016;245:S49-53.

20. Huchon C, Metzger U, Bats A-S, Bensaid C, Chatellier G, Azizi M, et al. Value of three-dimensional contrast-enhanced power Doppler ultrasound for characterizing adnexal masses. Journal of Obstetrics & Gynaecology Research. 2012;38(5):832-40.

21. Jabeen R, Khan SA, Naveed S. Risk of Malignancy index in the preoperative evaluation of patients with ovarian masses. Rawal Medical Journal 2015;40(1):78-80.

22. Jacobs IJ, Rivera H, Oram DH, Bast RC. DIFFERENTIAL-DIAGNOSIS OF OVARIAN-CANCER WITH TUMOR-MARKERS CA 125, CA 15-3 AND TAG 72-CENTER-DOT-3. British Journal of Obstetrics and Gynaecology. 1993;100(12):1120-4.

23. Jafari-Shobeiri M, Parizad M, Nazari F, Ouladsahebmadarek E, Sayyah-Melli M, Mostafa-Gharabaghi P, et al. Diagnostic value of HE4, CA125 and risk of ovarian malignancy algorithm in detecting ovarian cancer. International Journal of Women's Health and Reproduction Sciences. 2015;3(4):208-11.

24. Kader Ali Mohan GR, Jaaback K, Proietto A, Robertson R, Angstetra D. Risk Malignancy Index (RMI) in patients with abnormal pelvic mass: Comparing RMI 1, 2 and 3 in an Australian population. Australian & New Zealand Journal of Obstetrics & Gynaecology. 2010;50(1):77-80.

25. Karlsen MA, Hogdall EV, Christensen IJ, Borgfeldt C, Kalapotharakos G, Zdrazilova-Dubska L, et al. A novel diagnostic index combining HE4, CA125 and age may improve triage of women with suspected ovarian cancer - An international multicenter study in women with an ovarian mass. Gynecologic Oncology. 2015;138(3):640-6.

Table 12. Excluded studies: data not reported separately by menopausal status *(Continued)*

26. Lee HN, Lee KH, Park H. Diagnostic accuracy of risk of ovarian malignancy algorithm (ROMA) experienced in the clinical practice. Gynecologic Oncology. 2019;154 (Supplement 1):261.
27. Leone Roberti Maggiore U, Chiappa V, Ferrero S, Bogani G, Perotto S, Martinelli F, et al. Subjective ultrasound assessment and the ADNEX model to differentiate between benign and malignant ovarian tumors. Journal of Minimally Invasive Gynecology. 2017;24 (7 Supplement 1):S106.
28. Lewis J, McKnight L, Das N, Lutchman-Singh K. A prospective study comparing the performance of RMI, simple rules (IOTA) and MDT opinion for assessment of high risk masses. International Journal of Gynecological Cancer. 2018;28 (Supplement 2):752.
29. Li PL, Zhang X, Li TF, Wang LL, Du LT, Yang YM, et al. Combined detection of sialic acid and hydroxyproline in diagnosis of ovarian cancer and its comparison with human epididymis protein 4 and carbohydrate antigen 125. Clinica Chimica Acta. 2015;439:148-53.
30. Lokich E, Palisoul M, Romano N, Craig Miller M, Robison K, Stuckey A, et al. Assessing the risk of ovarian malignancy algorithm for the conservative management of women with a pelvic mass. Gynecologic Oncology. 2015;139(2):248-52.
31. Madar I, Szabo G. Evaluation of IOTA simple rules and IOTA ADNEX model in the hands of expert examiners at the diagnosis of the adnexal tumors. Australasian Journal of Ultrasound in Medicine. 2019;22 (2):147.
32. Mallari RGO, Coloma MLB. Comparison of sassone scoring and adnex model in differentiating benign and malignant ovarian tumour. International Journal of Gynecology and Obstetrics. 2018;143 (Supplement 3):306-7.
33. Manjunath AP, Pratapkumar, Sujatha K, Vani R. Comparison of three risk of malignancy indices in evaluation of pelvic masses. Gynecologic Oncology. 2001;81(2):225-9.
34. Michalak M, Gasiorowska E, Markwitz EN. Diagnostic value of CA125, HE4, ROMA and logistic regression model in pelvic mass diagnostics - our experience. Ginekologia Polska. 2015;86(4):256-61.
35. Migda M, Bartosz M, Migda MS, Kierszk M, Katarzyna G, Malenczyk M. Diagnostic value of the gynecology imaging reporting and data system (GI-RADS) with the ovarian malignancy marker CA-125 in preoperative adnexal tumor assessment. Journal of ovarian research. 2018;11(1):92.
36. Minar L, Felsinger M, Cermakova Z, Zlamal F, Bienertova-Vasku J. Comparison of the Copenhagen Index versus ROMA for the preoperative assessment of women with ovarian tumors. International Journal of Gynaecology & Obstetrics. 2018;140(2):241-6.
37. Mol BW, Boll D, De Kanter M, Heintz AP, Sijmons EA, Oei SG, et al. Distinguishing the benign and malignant adnexal mass: an external validation of prognostic models. Gynecologic Oncology. 2001;80(2):162-7.
38. Moolthiya W, Yuenyao P. The risk of malignancy index (RMI) in diagnosis of ovarian malignancy. Asian Pacific Journal of Cancer Prevention: Apjcp. 2009;10(5):865-8.
39. Moore RG, Blackman A, Miller MC, Robison K, DiSilvestro PA, Eklund EE, et al. Multiple biomarker algorithms to predict epithelial ovarian cancer in women with a pelvic mass: Can additional makers improve performance? Gynecologic Oncology. 2019;154(1):150-5.
40. Morgante G, la Marca A, Ditto A, De Leo V. Comparison of two malignancy risk indices based on serum CA125, ultrasound score and menopausal status in the diagnosis of ovarian masses. British Journal of Obstetrics & Gynaecology. 1999;106(6):524-7.
41. Nakajima T, Nagaishi M, Maebayashi A, Miyakawa Y, Ikeda Y, Sato M, et al. Evaluation of HE4 in ovarian cancer. International Journal of Gynecological Cancer. 2018;28 (Supplement 2):785.
42. Nunes N, Ambler G, Foo X, Widschwendter M, Jurkovic D. Prospective evaluation of IOTA logistic regression models LR1 and LR2 in comparison with subjective pattern recognition for diagnosis of ovarian cancer in an outpatient setting. Ultrasound in Obstetrics & Gynecology. 2018;51(6):829-35.
43. Nunes N, Ambler G, Hoo WL, Naftalin J, Foo X, Widschwendter M, et al. Prospective validation of the iota logistic regression models (LR1/LR2) by a level-2 ultrasound operator and comparison to pattern recognition for the diagnosis of ovarian cancer. International Journal of Gynecological Cancer. 2013;1):587.

Table 12. Excluded studies: data not reported separately by menopausal status *(Continued)*

44. Obeidat BR, Amarin ZO, Latimer JA, Crawford RA. Risk of malignancy index in the preoperative evaluation of pelvic masses. International Journal of Gynaecology & Obstetrics. 2004;85(3):255-8.
45. Ong C, Biswas A, Choolani M, Low JJH. Comparison of risk of malignancy indices in evaluating ovarian masses in a Southeast Asian population. Singapore Medical Journal. 2013;54(3):136-9.
46. Oranratanaphan S, Wanishpongpan S, Termrungruanglert W, Triratanachat S. Assessment of diagnostic values among CA-125, RMI, HE4, and ROMA for cancer prediction in women with nonfunctional ovarian cysts. Obstetrics and Gynecology International. 2018;2018 (no pagination)(7821574).
47. Ozbay PO, Ekinci T, Caltekin MD, Yilmaz HT, Temur M, Yilmaz O, et al. Comparative evaluation of the risk of malignancy index scoring systems (1-4) used in differential diagnosis of adnexal masses. Asian Pacific Journal of Cancer Prevention: Apjcp. 2015;16(1):345-9.
48. Priyanka V, Karthiga S, Sivanesan B, Jagadeesan N, Balasubramani L. A Comparative Study of RMI and ROMA in Women Presenting with an Adnexal Mass. Indian Journal of Gynecologic Oncology. 2018;16 (1) (no pagination)(4).
49. Qiu L, Yang F, Luo H. A preliminary study the sequential use of the risk malignancy index and contrast-enhanced ultrasonography in differential diagnosis of adnexal masses. Medicine (United States). 2018;97 (29) (no pagination)(e11536).
50. Radwan R, Falcone K, Raju S, Jones S, Gannon S, Le J, et al. A risk of ovarian malignancy algorithm (ROMA) derived from lumipulse G HE4 and lumipulse CA125II assays. Clinical Chemistry. 2016;62 (10 Supplement 1):S11.
51. Ristic A, Filimonovic D, Dzatic-Smijkovic O, Dimitrijevic D, Anicic R, Mihajlovic S, et al. Risk of malignancy index in discrimination between benign and malignant adnexal masses. European Journal of Gynaecological Oncology. 2018;39(5):733-6.
52. Sandri MT, Bottari F, Franchi D, Boveri S, Candiani M, Ronzoni S, et al. Comparison of HE4, CA125 and ROMA algorithm in women with a pelvic mass: correlation with pathological outcome. Gynecologic Oncology. 2013;128(2):233-8.
53. Sayasneh A, Ferrara L, De Cock B, Saso S, Al-Memar M, Johnson S, et al. Evaluating the risk of ovarian cancer before surgery using the ADNEX model: a multicentre external validation study. British Journal of Cancer. 2016;115(5):542-8.
54. Shimada K, Matsumoto K, Mimura T, Ishikawa T, Munechika J, Ohgiya Y, et al. Ultrasound-based logistic regression model LR2 versus magnetic resonance imaging for discriminating between benign and malignant adnexal masses: a prospective study. International Journal of Clinical Oncology. 2018;23(3):514-21
55. Shulman LP, Francis M, Bullock R, Pappas T. Clinical Performance Comparison of Two In-Vitro Diagnostic Multivariate Index Assays (IVDMIAs) for Presurgical Assessment for Ovarian Cancer Risk. Advances in Therapy. 2019.
56. Simsek HS, Tokmak A, Ozgu E, Doganay M, Danisman N, Erkaya S, et al. Role of a risk of malignancy index in clinical approaches to adnexal masses. Asian Pacific Journal of Cancer Prevention: Apjcp. 2014;15(18):7793-7.
57. Sladkevicius P, Valentin L. Interobserver agreement in describing the ultrasound appearance of adnexal masses and in calculating the risk of malignancy using logistic regression models. Clinical Cancer Research. 2015;21(3):594-601.
58. Stukan M, Badocha M, Ratajczak K. Development and validation of a model that includes two ultrasound parameters and the plasma D-dimer level for predicting malignancy in adnexal masses: An observational study. BMC Cancer. 2019;19 (1) (no pagination)(564).
59. Tingulstad S, Hagen B, Skjeldestad FE, Onsrud M, Kiserud T, Halvorsen T, et al. Evaluation of a risk of malignancy index based on serum CA125, ultrasound findings and menopausal status in the pre-operative diagnosis of pelvic masses. British Journal of Obstetrics and Gynaecology. 1996;103(8):826-31.
60. Ulusoy S, Akbayir O, Numanoglu C, Ulusoy N, Odabas E, Gulkilik A. The risk of malignancy index in discrimination of adnexal masses. International Journal of Gynaecology & Obstetrics. 2007;96(3):186-91.

Table 12. Excluded studies: data not reported separately by menopausal status *(Continued)*

61. Vanderstichele A, Busschaert P, Smeets D, Landolfo C, Van Nieuwenhuysen E, Leunen K, et al. Chromosomal instability in cell-free DNA as a highly specific biomarker for detection of ovarian cancer in women with adnexal masses. Clinical Cancer Research. 2017;23(9):2223-31.

62. Van Calster B, Timmerman D, Valentin L, McIndoe A, Ghaem-Maghami S, Testa AC, et al. Triaging women with ovarian masses for surgery: observational diagnostic study to compare RCOG guidelines with an International Ovarian Tumour Analysis (IOTA) group protocol. BJOG: An International Journal of Obstetrics & Gynaecology. 2012;119(6):662-71.

63. Van Holsbeke C, Van Calster B, Valentin L, Testa AC, Ferrazzi E, Dimou I, et al. External validation of mathematical models to distinguish between benign and malignant adnexal tumors: a multicenter study by the International Ovarian Tumor Analysis Group. Clinical Cancer Research. 2007;13(15 Pt 1):4440-7.

64. Wilailak S, Chan KK, Chen CA, Nam JH, Ochiai K, Aw TC, et al. Distinguishing benign from malignant pelvic mass utilizing an algorithm with HE4, menopausal status, and ultrasound findings. Journal of Gynecologic Oncology. 2015;26(1):46-53.

65. Winarto H, Laihad BJ, Nuranna L. Modification of cutoff values for HE4, CA125, the Risk of Malignancy Index, and the Risk of Malignancy Algorithm for ovarian cancer detection in Jakarta, Indonesia. Asian Pacific Journal of Cancer Prevention: Apjcp. 2014;15(5):1949-53.

66. Yehia M, Mansour A, Mekawy S. Human epididymis protein 4 (HE4) mRNA as a prognostic marker in ovarian tumors in relation to RMI and CA125. International Journal of Cancer Research. 2015;11(4):175-85.

67. Yamamoto Y, Yamada R, Oguri H, Maeda N, Fukaya T. Comparison of four malignancy risk indices in the preoperative evaluation of patients with pelvic masses. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2009;144(2):163-7.

68. Yamamoto Y, Tsuchida A, Ushiwaka T, Nagai R, Matsumoto M, Komatsu J, et al. Comparison of 4 risk-of-malignancy indexes in the preoperative evaluation of patients with pelvic masses: A prospective study. Clinical Ovarian and other Gynecologic Cancer 2015;7(1-2):8-12.

69. Yanaranop M, Jantarateptewan N, Tiyayon J, Nakrangsee S. Significance of Serum Human Epididymis Protein 4 and Cancer Antigen 125 in Distinguishing Type I and Type II Epithelial Ovarian Cancers. International Journal of Gynecological Cancer. 2018;28(6):1058-65.

70. Yanaranop M, Anakrat V, Siricharoenthai S, Nakrangsee S, Thinkhamrop B. Is the Risk of Ovarian Malignancy Algorithm Better Than Other Tests for Predicting Ovarian Malignancy in Women with Pelvic Masses? Gynecologic and Obstetric Investigation. 2017;82(1):47-53.

71. Yanaranop M, Tiyayon J, Siricharoenthai S, Nakrangsee S, Thinkhamrop B. Rajavithi-ovarian cancer predictive score (R-OPS): A new scoring system for predicting ovarian malignancy in women presenting with a pelvic mass. Gynecologic Oncology. 2016;141(3):479-84.

72. Yoshida A, Derchain SF, Pitta DR, Andrade LA, Sarian LO. Comparing the Copenhagen Index (CPH-I) and Risk of Ovarian Malignancy Algorithm (ROMA): Two equivalent ways to differentiate malignant from benign ovarian tumors before surgery? Gynecologic Oncology. 2016;140(3):481-5.

73. Zermeno-Nava JDJ, Martinez-Martinez MU, Ramirez-De-Avila AL, Hernandez-Arteaga AC, Garcia-Valdivieso MG, Hernandez-Cedillo A, et al. Determination of sialic acid in saliva by means of surface-enhanced Raman spectroscopy as a marker in adnexal mass patients: Ovarian cancer vs benign cases. Journal of Ovarian Research. 2018;11 (1) (no pagination)(61).

74. Zhang F, Zhang ZL. The Diagnostic Value of Transvaginal Sonograph (TVS), Color Doppler, and Serum Tumor Marker CA125, CEA, and AFP in Ovarian Cancer. Cell Biochemistry & Biophysics. 2015;72(2):353-7.

75. Zhang P, Wang C, Cheng L, Zhang P, Guo L, Liu W, et al. Comparison of HE4, CA125, and ROMA Diagnostic Accuracy: A Prospective and Multicenter Study for Chinese Women With Epithelial Ovarian Cancer. Medicine. 2015;94(52):e2402.

8851

Auekitrungreung, 2019 (not found in end note)

Table 13. Excluded studies: duplicate data reporting

Duplicate data reporting
1. Abdalla N, Bachanek M, Timorek-Lemieszczuk A, Cendrowski K, Sawicki W. Comparison of the diagnostic value of gynecologic imaging reporting and data system and Roma in the presurgical assesment of adnexal tumors. International Journal of Gynecological Cancer. 2015;1):374.
2. Ameye L, Valentin L, Testa AC, Van Holsbeke C, Domali E, Van Huffel S, et al. A scoring system to differentiate malignant from benign masses in specific ultrasound-based subgroups of adnexal tumors. Ultrasound in Obstetrics & Gynecology. 2009;33(1):92-101.
3. Antonic J, Rakar S. Validity of colour and pulsed Doppler US and tumour marker CA 125 in differentiation between benign and malignant ovarian masses. European Journal of Gynaecological Oncology. 1996;17(1):29-35.
4. Bouzari Z, Yazdani S, Shirkhani Kelagar Z, Abbaszadeh N. Risk of malignancy index as an evaluation of preoperative pelvic mass. Caspian Journal of Internal Medicine. 2011;2(4):331-5.
5. Chudecka-Glaz A, Cymbaluk-Ploska A, Jastrzebska J, Menkiszak J. Can ROMA algorithm stratify ovarian tumor patients better when being based on specific age ranges instead of the premenopausal and postmenopausal status? Tumour Biology. 2016;37(7):8879-87.
6. Chudecka-Glaz A, Cymbaluk-Ploska A, Luterek-Puszynska K, Menkiszak J. Diagnostic usefulness of the Risk of Ovarian Malignancy Algorithm using the electrochemiluminescence immunoassay for HE4 and the chemiluminescence microparticle immunoassay for CA125. Oncology Letters. 2016;12(5):3101-14.
7. Chen X, Zhou H, Chen R, He J, Wang Y, Huang L, et al. Development of a multimarker assay for differential diagnosis of benign and malignant pelvic masses. Clinica Chimica Acta. 2015;440:57-63.
8. Derchain S, Pitta DR, Sarian LO, Barreta A, Campos EA, Angelo-Andrade LL, et al. Analysis of symptoms for the preoperative prediction of malignancy of ovarian masses in brazilian women. International Journal of Gynecological Cancer. 2013;1):464.
9. Ertas S, Vural F, Tufekci EC, Ertas AC, Kose G, Aka N. Predictive Value of Malignancy Risk Indices for Ovarian Masses in Premenopausal and Postmenopausal Women. Asian Pacific journal of cancer prevention: APJCP. 2016;17(4):2177-83.
10. Fujiwara H, Suzuki M, Takeshima N, Takizawa K, Kimura E, Nakanishi T, et al. Evaluation of human epididymis protein 4 (HE4) and Risk of Ovarian Malignancy Algorithm (ROMA) as diagnostic tools of type I and type II epithelial ovarian cancer in Japanese women. Tumor Biology. 2015;36(2):1045-53.
11. Grenache DG, Heichman KA, Werner TL, Vucetic Z. Clinical performance of two multi-marker blood tests for predicting malignancy in women with an adnexal mass. Clinica Chimica Acta. 2015;438:358-63.
12. Janas L, Glowacka E, Wilczynski JR, Malinowski A, Nowak M. Evaluation of applicability of HE4 and ROMA in the preoperative diagnosis of adnexal masses. [Polish]. Ginekologia polska. 2015;86(3):193-7.
13. Krascsenits G, Balazs B, Dudnyikova A, Purcsi K, Orosz E, Pete I. Investigating the predictive value of RMI and ROMA indices in patients with ovarian tumors of uncertain dignity. [Hungarian]. Magyar onkologia. 2016;60(4):320-7.
14. Leone Roberti Maggiore U, Chiappa V, Bogani G, Perotto S, Signorelli M, Martinelli F, et al. Subjective ultrasound assessment and the adnex model to differentiate between benign and malignant ovarian tumors. International Journal of Gynecological Cancer. 2017;27 (Supplement 4):972.
15. Mahyuddin AP, Liu L, Zhao C, Kothandaraman N, Salto-Tellez M, Pang BNK, et al. Diagnostic accuracy of haptoglobin within ovarian cyst fluid as a potential point-of-care test for epithelial ovarian cancer: an observational study. BJOG: An International Journal of Obstetrics and Gynaecology. 2018;125(4):421-31.
16. Meys EMJ, Jeelof LS, Ramaekers BLT, Dirksen CD, Kooreman LFS, Slangen BFM, et al. Economic evaluation of an expert examiner and different ultrasound models in the diagnosis of ovarian cancer. European Journal of Cancer. 2018;100:55-64.

Table 13. Excluded studies: duplicate data reporting *(Continued)*

Meys EMJ, Kaijser J, Kruitwagen RFPM, Slangen BFM, Van Calster B, Aertgeerts B, et al. Subjective assessment versus ultrasound models to diagnose ovarian cancer: A systematic review and meta-analysis. European Journal of Cancer. 2016;58:17-29.
17. Nunes N, Yazbek J, Ambler G, Hoo W, Naftalin J, Jurkovic D. Prospective evaluation of the IOTA logistic regression model LR2 for the diagnosis of ovarian cancer. Ultrasound in Obstetrics & Gynecology. 2012;40(3):355-9.
18. Richards A, Herbst U, Pather S, Saidi S, Tejada-Berges T, Williams P, et al. HE4, CA125, the Risk of Malignancy Algorithm (ROMA) and the Risk of Malignancy Index (RMI) and complex pelvic masses - a prospective comparison in the preoperative evaluation of adnexal and pelvic masses in an Australian population. BJOG: An International Journal of Obstetrics and Gynaecology. 2015;2):150.
19. Sokalska A, Timmerman D, Testa AC, Van Holsbeke C, Lissoni AA, Leone FPG, et al. Diagnostic accuracy of transvaginal ultrasound examination for assigning a specific diagnosis to adnexal masses. Ultrasound in Obstetrics & Gynecology. 2009;34(4):462-70.
20. Van Den Akker PAJ, Zusterzeel PLM, Aalders AL, Snijders MPLM, Samlal RAK, Vollebergh JHA, et al. Use of risk of malignancy index to indicate frozen section analysis in the surgical care of women with ovarian tumors. International Journal of Gynecology and Obstetrics. 2016;133(3):355-8.
21. Yoshida A, Derchain SF, Pitta DR, De Angelo Andrade LAL, Sarian LO. Comparing the Copenhagen Index (CPH-I) and Risk of Ovarian Malignancy Algorithm (ROMA): Two equivalent ways to differentiate malignant from benign ovarian tumors before surgery? Gynecologic Oncology. 2016;140(3):481-5.
22. Zhang P, Wang C, Cheng L, Zhaang P, Guo L, Liu W, et al. Comparison of HE4, CA125, and ROMA diagnostic accuracy a prospective and multicenter study for Chinese women with epithelial ovarian cancer. Medicine (United States). 2015;94 (52) (no pagination)(e2402).

Table 14. Excluded studies: full text not available

Full text not available
1. Castelli M, Battaglia F, Scambia G, Benedetti Panici P, Ferrandina G, Mileo AM, et al. Immunosuppressive acidic protein and CA 125 levels in patients with ovarian cancer. Oncology. 1991;48(1):13-7.
2. Gadducci A, Ferdeghini M, Rispoli G, Prontera C, Bianchi R, Fioretti P. Comparison of tumor-associated trypsin inhibitor (TATI) with CA125 as a marker for diagnosis and monitoring of epithelial ovarian cancer. Scandinavian Journal of Clinical and Laboratory Investigation, Supplement. 1991;51(207):19-24.
3. Inoue M, Fujita M, Nakazawa A, Ogawa H, Tanizawa O. Sialyl-Tn, sialyl-Lewis Xi, CA 19-9 CA 125, carcinoembryonic antigen, and tissue polypeptide antigen in differentiating ovarian cancer from benign tumors. Obstetrics and Gynecology. 1992;79(3):434-40.
Jabeen R, Khan SA, Naveed S. Risk of Malignancy index in the preoperative evaluation of patients with ovarian masses. Rawal Medical Journal. 2015;40(1):78-80.
4. Kenemans P, Bon GG, Kessler AC, Verstraeten AA, Van Kamp GJ. Multicenter technical and clinical evaluation of a fully automated enzyme immunoassay for CA 125. Clinical Chemistry. 1992;38(8 Part 1):1466-71.
5. Lotzniker M, Pavesi F, Scarabelli M, Vadacca G, Franchi M, Moratti R. Tumour associated antigens CA 15.3 and CA 125 in ovarian cancer. International Journal of Biological Markers. 1991;6(2):115-21.
6. Paschalidis M, Tserkezoglou A, Fotiou S, Magiakos G, Voulgaris Z, Ferderigou A. CASA and CA125 in the preoperative evaluation of adnexal tumours. European Journal of Gynaecological Oncology. 1996;17(5):450.
7. PetersEngl C, Medl M, Ogris E, Leodolter S. Tumor-associated trypsin inhibitor (TATI) and cancer antigen 125 (CA125) in patients with epithelial ovarian cancer. Anticancer Research. 1995;15(6B):2727-30.

Table 14. Excluded studies: full text not available *(Continued)*

8. Shahzad N, Rashid N, Zahra S, Malik A. Role of malignancy index in prediction of malignancy in ovarian masses preoperative. Medical Forum Monthly. 2015;26(6):44-7.
9. Singer G, Guth U. Therapeutische Umschau. 2017;74(4):181-9.
10. Song HJ, Nam KM, Kim JD, Park CY, Kim YS. Looking for optimized weights of CA125 and HE4 in early screening system of ovarian cancer for Korean patients. Technology & Health Care. 2015;24 Suppl 1:S163-70.
11. Schwarzler P, Timmerman D. IOTA strategies for assessment of adnexal findings: Improving strategies for diagnosing ovarian cancer. [German]. Gynakologe. 2018;51(3):178-87.
12. Wang Q, Wu Y, Zhang H, Yang K, Tong Y, Chen L, et al. Clinical value of serum HE4, CA125, CA72-4, and ROMA index for diagnosis of ovarian cancer and prediction of postoperative recurrence. Clinical Laboratory. 2019;65(4):631-40.
13. Zapardiel I, Gorostidi M, Ravaggi A, Allende MT, Silveira M, Macuks R. Utility of human epididymis protein 4 serum marker for the detection of adnexal malignancy: A multicentric prospective study. European Journal of Cancer Prevention. 2017;26(4):346-50.

Table 15. Excluded studies: index test not applicable

Index test not applicable
1. Abdalla N, Bachanek M, Piorkowski R, Cendrowski K, Sawicki W. The role of risk of malignancy index RMI-IV, CA125 and HE4 in the preoperative assesment of adnexal masses. International Journal of Gynecological Cancer. 2015;1):1266.
2. Abdel-Azeez HA, Labib HA, Sharaf SM, Refai AN. HE4 and mesothelin: novel biomarkers of ovarian carcinoma in patients with pelvic masses. Asian Pacific Journal of Cancer Prevention: Apjcp. 2010;11(1):111-6.
3. Acharya U, Sree S, Saba L, Molinari F, Guerriero S, Suri J. Ovarian Tumor Characterization and Classification Using Ultrasound-A New Online Paradigm. Journal of Digital Imaging. 2013;26(3):544-53.
4. Acharya UR, Sree SV, Krishnan MMR, Saba L, Molinari F, Guerriero S, et al. Ovarian tumor characterization using 3D ultrasound. Technology in Cancer Research & Treatment. 2012;11(6):543-52.
5. Akturk E, Dede M, Yenen MC, Kocyitit YK, Ergun A. Comparison of nine morphological scoring systems to detect ovarian malignancy. European Journal of Gynaecological Oncology. 2015;36(3):304-8.
6. Akturk E, Karaca RE, Alanbay I, Dede M, Karasahin E, Yenen MC, et al. Comparison of four malignancy risk indices in the detection of malignant ovarian masses. Journal of Gynecologic Oncology. 2011;22(3):177-82.
7. Al-Asadi JN, Al-Maliki SK, Al-Dahhhan F, Al-Naama L, Suood F. The accuracy of risk malignancy index in prediction of malignancy in women with adnexal mass in Basrah, Iraq. Nigerian Journal of Clinical Practice. 2018;21(10):1254-9.
8. Alcazar JL, Jurado M. Using a logistic model to predict malignancy of adnexal masses based on menopausal status, ultrasound morphology, and color Doppler findings. Gynecologic Oncology. 1998;69(2):146-50.
9. Alcazar JL, Jurado M. Prospective evaluation of a logistic model based on sonographic morphologic and color Doppler findings developed to predict adnexal malignancy. Journal of Ultrasound in Medicine. 1999;18(12):837-42.
10. Alcazar JL, Pascual MA, Olartecoechea B, Graupera B, Auba M, Ajossa S, et al. IOTA simple rules for discriminating between benign and malignant adnexal masses: prospective external validation. Ultrasound in Obstetrics & Gynecology. 2013;42(4):467-71.
11. Al-Musalhi K, Al-Kindi M, Ramadhan F, Al-Rawahi T, Al-Hatali K, Mula-Abed WA. Validity of cancer antigen-125 (CA-125) and risk of malignancy index (RMI) in the diagnosis of ovarian cancer. Oman Medical Journal. 2015;30(6):428-34.

Table 15. Excluded studies: index test not applicable *(Continued)*

12. Ameye L, Timmerman D, Valentin L, Paladini D, Zhang J, Van Holsbeke C, et al. Clinically oriented three-step strategy for assessment of adnexal pathology. Ultrasound in Obstetrics & Gynecology. 2012;40(5):582-91.

13. Amonkar SD, Bertenshaw GP, Chen T-H, Bergstrom KJ, Zhao J, Seshaiah P, et al. Development and preliminary evaluation of a multivariate index assay for ovarian cancer. PLoS ONE [Electronic Resource]. 2009;4(2):e4599.

14. Amor F, Alcazar JL, Vaccaro H, Leon M, Iturra A. GI-RADS reporting system for ultrasound evaluation of adnexal masses in clinical practice: a prospective multicenter study. Ultrasound in Obstetrics & Gynecology. 2011;38(4):450-5.

15. Andersen ES, Knudsen A, Rix P, Johansen B. Risk of malignancy index in the preoperative evaluation of patients with adnexal masses. Gynecologic Oncology. 2003;90(1):109-12.

16. Antovska SV, Bashevska N, Aleksioska N. Predictive values of the ultrasound parameters, CA-125 and risk of malignancy index in patients with ovarian cancer. Klinicka Onkologie. 2011;24(6):435-42.

17. Arab M, Yaseri M, Ashrafganjoi T, Maktabi M, Noghabaee G, Sheibani K. Comparison of two ovarian malignancy prediction models based on age sonographic findings and serum CA125 measurement. Asian Pacific Journal of Cancer Prevention: Apjcp. 2012;13(8):4199-202.

18. Ashrafgangooei T, Rezaeezadeh M. Risk of malignancy index in preoperative evaluation of pelvic masses. Asian Pacific Journal of Cancer Prevention: Apjcp. 2011;12(7):1727-30.

19. Bailey J, Tailor A, Naik R, Lopes A, Godfrey K, Hatem HM, et al. Risk of malignancy index for referral of ovarian cancer cases to a tertiary center: does it identify the correct cases? International Journal of Gynecological Cancer. 2006;16 Suppl 1:30-4.

20. Bandiera E, Zanotti L, Bignotti E, Romani C, Tassi R, Todeschini P, et al. Human kallikrein 5: an interesting novel biomarker in ovarian cancer patients that elicits humoral response. International Journal of Gynecological Cancer. 2009;19(6):1015-21.

21. Barco A, Molina R, Carreras N, Rius M, Rodriguez I, Auge J, et al. Effectiveness of IOTA and ROMA algorithms in surgical management of adnexal masses. Clinica Chimica Acta. 2019;493 (Supplement 1):S134-S5.

22. Barkauskas DA, An HJ, Kronewitter SR, de Leoz ML, Chew HK, de Vere White RW, et al. Detecting glycan cancer biomarkers in serum samples using MALDI FT-ICR mass spectrometry data. Bioinformatics. 2009;25(2):251-7.

23. Barroilhet L, Vitonis A, Shipp T, Muto M, Benacerraf B. Sonographic predictors of ovarian malignancy. Journal of Clinical Ultrasound. 2013;41(5):269-74.

24. Bastani A, Asghary A, Heidari MH, Karimi-Busheri F. Evaluation of the sensitivity and specificity of serum level of prostasin, CA125, LDH, AFP, and hCG+beta in epithelial ovarian cancer patients. European Journal of Gynaecological Oncology. 2017;38(3):418-24.

25. Bedkowska GE, Gacuta E, Zajkowska M, Glazewska EK, Osada J, Szmitkowski M, et al. Plasma levels of MMP-7 and TIMP-1 in laboratory diagnostics and differentiation of selected histological types of epithelial ovarian cancers. Journal of ovarian research. 2017;10(1):39.

26. Bristow RE, Hodeib M, Smith A, Chan DW, Zhang Z, Fung ET, et al. Impact of a multivariate index assay on referral patterns for surgical management of an adnexal mass. American Journal of Obstetrics & Gynecology 2013;209(6):581.e1-8.

27. Bristow RE, Smith A, Zhang Z, Chan DW, Crutcher G, Fung ET, et al. Ovarian malignancy risk stratification of the adnexal mass using a multivariate index assay. Gynecologic Oncology 2013;128(2):252-9. [Other: Secondary publication- Bristow 2013a]

28. Bromley B, Goodman H, Benacerraf BR. Comparison between sonographic morphology and Doppler waveform for the diagnosis of ovarian malignancy. Obstetrics and gynecology. 1994;83(3):434-7.

29. Brown DL, Doubilet PM, Miller FH, Frates MC, Laing FC, DiSalvo DN, et al. Benign and malignant ovarian masses: selection of the most discriminating gray-scale and Doppler sonographic features. Radiology. 1998;208(1):103-10.

Table 15. Excluded studies: index test not applicable *(Continued)*

30. Buckshee K, Temsu I, Bhatla N, Deka D. Pelvic examination, transvaginal ultrasound and transvaginal color Doppler sonography as predictors of ovarian cancer. International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics. 1998;61(1):51-7.

31. Cao D, Guo S, Allan RW, Molberg KH, Peng Y. SALL4 is a novel sensitive and specific marker of ovarian primitive germ cell tumors and is particularly useful in distinguishing yolk sac tumor from clear cell carcinoma. American Journal of Surgical Pathology. 2009;33(6):894-904.

32. Caruso A, Caforio L, Testa AC, Ciampelli M, Panici PB, Mancuso S. Transvaginal color Doppler ultrasonography in the presurgical characterization of adnexal masses. Gynecologic Oncology. 1996;63(2):184-91.

33. Chirdchim W, Wanichsetakul P, Phinyo P, Patumanond J, Suwannarurk K, Srisomboon J. Development and Validation of a Predictive Score for Preoperative Diagnosis of Early Stage Epithelial Ovarian Cancer. Asian Pacific journal of cancer prevention: APJCP. 2019;20(4):1207-13.

34. Choi H, Kim H. Diagnostic accuracy of inflammatory markers for distinguishing malignant and benign ovarian masses. International Journal of Gynecology and Obstetrics. 2018;143 (Supplement 3):896.

35. Ciledag N, Arda K, Aktas E, Aribas BK. A pilot study on real-time transvaginal ultrasonographic elastography of cystic ovarian lesions. Indian Journal of Medical Research. 2013;137(6):1089-92.

36. Clark BZ, Beriwal S, Dabbs DJ, Bhargava R. Semiquantitative GATA-3 immunoreactivity in breast, bladder, gynecologic tract, and other cytokeratin 7-positive carcinomas. American Journal of Clinical Pathology. 2014;142(1):64-71.

37. Clayton RD, Snowden S, Weston MJ, Mogensen O, Eastaugh J, Lane G. Neural networks in the diagnosis of malignant ovarian tumours. British Journal of Obstetrics & Gynaecology. 1999;106(10):1078-82.

38. Coleman RL, Herzog TJ, Chan DW, Munroe DG, Pappas TC, Smith A, et al. Validation of a second-generation multivariate index assay for malignancy risk of adnexal masses. American Journal of Obstetrics & Gynecology. 2016;215(1):82.e1-.e11.

39.Cortesi L, De Matteis E, Toss A, Marchi I, Medici V, Contu G, et al. Evaluation of Transvaginal Ultrasound plus CA-125 Measurement and Prophylactic Salpingo-Oophorectomy in Women at Different Risk Levels of Ovarian Cancer: The Modena Study Group Cohort Study. Oncology. 2017;93(6):377-86.

40. Daemen A, Jurkovic D, Van Holsbeke C, Guerriero S, Testa AC, Czekierdowski A, et al. Effect of cancer prevalence on the use of risk-assessment cut-off levels and the performance of mathematical models to distinguish malignant from benign adnexal masses. Ultrasound in Obstetrics & Gynecology. 2011;37(2):226-31.

41. Daemen A, Valentin L, Fruscio R, Van Holsbeke C, Melis GB, Guerriero S, et al. Improving the preoperative classification of adnexal masses as benign or malignant by second-stage tests. Ultrasound in Obstetrics & Gynecology. 2011;37(1):100-6.

42. Dasari P, Catherine Leela Pannirselvan P, Sridhar MG. Ultrasonographic scoring and risk of malignancy index in preoperative prediction of ovarian malignancy. Journal of Gynecologic Surgery. 2013;29(2):61-4.

43. Dearking AC, Aletti GD, McGree ME, Weaver AL, Sommerfield MK, Cliby WA. How relevant are ACOG and SGO guidelines for referral of adnexal mass? Obstetrics and Gynecology 2007;110(4):841-8.

44. Delic R, Weber V, Stefanovic M, Krivec S. Usefulness of HE4 and CA125 in the preoperative diagnosis of adnexal masses: Does age or menopausal status matter? International Journal of Gynecological Cancer. 2015;1):1327-8.

45. DePriest PD, Shenson D, Fried A, Hunter JE, Andrews SJ, Gallion HH, et al. A morphology index based on sonographic findings in ovarian cancer. Gynecologic Oncology. 1993;51(1):7-11.

46. DePriest PD, Varner E, Powell J, Fried A, Puls L, Higgins R, et al. The efficacy of a sonographic morphology index in identifying ovarian cancer: a multi-institutional investigation. Gynecologic Oncology. 1994;55(2):174-8.

Table 15. Excluded studies: index test not applicable *(Continued)*

47. Desai D, Desai VA, Verma RN, Shrivastava A. Role of gray scale and color Doppler in differentiating benign from malignant ovarian masses. Journal of Mid-life Health. 2010;1(1):23-5.

48. Dora SK, Dandapat AB, Pande B, Hota JP. A prospective study to evaluate the risk malignancy index and its diagnostic implication in patients with suspected ovarian mass. Journal of ovarian research. 2017;10(1):55.

49. Eskander RN, Carpenter BA, Wu HG, Wolf JK. The clinical utility of an elevated-risk multivariate index assay score in ovarian cancer patients. Current Medical Research & Opinion. 2016;32(6):1161-5.

50. Fan HY, Duan DM, Liu YF. The diagnostic value of determination of serum GOLPH3 associated with CA125, CA19.9 in patients with ovarian cancer. European Review for Medical & Pharmacological Sciences. 2017;21(18):4039-44.

51. Ferrazzi E, Zanetta G, Dordoni D, Berlanda N, Mezzopane R, Lissoni G. Transvaginal ultrasonographic characterization of ovarian masses: comparison of five scoring systems in a multicenter study. Ultrasound in Obstetrics & Gynecology. 1997;10(3):192-7.

52. Firoozabadi RD, Karimi Zarchi M, Mansurian HR, Moghadam BR, Teimoori S, Naseri A. Evaluation of diagnostic value of CT scan, physical examination and ultrasound based on pathological findings in patients with pelvic masses. Asian Pacific Journal of Cancer Prevention: Apjcp. 2011;12(7):1745-7.

53. Founta C, Ratnavelu N, Hoskins T, Addison C, Bradbury M, Handley G, et al. The diagnostic value of adding HE4 to CA125 for symptomatic women investigated for ovarian cancer in primary care: Final results of a prospective pilot study. International Journal of Gynecological Cancer. 2015;1):44-5.

54. Geomini PMA, Kluivers KB, Moret E, Bremer GL, Kruitwagen RFP, Mol BWJ. Evaluation of adnexal masses with three-dimensional ultrasonography. Obstetrics & Gynecology. 2006;108(5):1167-75.

55. Geomini PMAJ, Coppus SFPJ, Kluivers KB, Bremer GL, Kruitwagen RFPM, Mol BWJ. Is three-dimensional ultrasonography of additional value in the assessment of adnexal masses? Gynecologic Oncology. 2007;106(1):153-9.

56. Goff BA, Agnew K, Neradilek MB, Gray HJ, Liao JB, Urban RR. Combining a symptom index, CA125 and HE4 (triple screen) to detect ovarian cancer in women with a pelvic mass. Gynecologic Oncology. 2017;147(2):291-5.

57. Goff BA, Mandel L, Muntz HG, Melancon CH. Ovarian carcinoma diagnosis. Cancer. 2000;89(10):2068-75.

58. Goff BA, Mandel LS, Melancon CH, Muntz HG. Frequency of symptoms of ovarian cancer in women presenting to primary care clinics. Jama. 2004;291(22):2705-12.

59. Goodrich ST, Bristow RE, Santoso JT, Miller RW, Smith A, Zhang Z, et al. The effect of ovarian imaging on the clinical interpretation of a multivariate index assay. American Journal of Obstetrics & Gynecology 2014;211(1):65.e1-65.e11. [Other: Secondary publication]

60. Gong SP, Chen YN, Zhang YD, Yao W, Chen L, Liu SS, et al. [Diagnostic value of CA125, HE4 and Copenhagen Index in differentiating benign from malignant epithelial ovarian tumors]. Nan Fang Yi Ke Da Xue Xue Bao = Journal of Southern Medical University. 2017;37(5):628-32

61. Guerriero S, Ajossa S, Piras S, Angiolucci M, Marisa O, Melis GB. Diagnosis of paraovarian cysts using transvaginal sonography combined with CA 125 determination. Ultrasound in Obstetrics & Gynecology. 2006;28(6):856-8.

62. Gupta D, Lis CG. Pretreatment serum albumin as a predictor of cancer survival: A systematic review of the epidemiological literature. Nutrition Journal. 2010;9(1):1-16.

63. Gupta P, Weathers T, Stehman FB, Schilder JM, Method MW, Hess LM. Ovarian cancer symptoms and diagnosis: Patient reported symptoms and symptoms reported in patients' medical records. International Journal of Gynecological Cancer. 2011;2):119.

64. Hafeez S, Sufian S, Beg M, Hadi Q, Jamil Y, Masroor I. Role of ultrasound in characterization of ovarian masses. Asian Pacific Journal of Cancer Prevention: Apjcp. 2013;14(1):603-6.

Table 15. Excluded studies: index test not applicable *(Continued)*

65. Hakansson F, Hogdall EVS, Nedergaard L, Lundvall L, Engelholm SA, Pedersen AT, et al. Risk of malignancy index used as a diagnostic tool in a tertiary centre for patients with a pelvic mass. Acta Obstetricia et Gynecologica Scandinavica. 2012;91(4):496-502.

66. Huang B, Lv G, Chen Q. Comparison of GI-RADS ultrasonographic stratification and rmi4 in differentiating benign from malignant adnexal masses. Ultrasound in Medicine and Biology. 2017;43 (Supplement 1):S137.

67. Im SS, Gordon AN, Buttin BM, Leath CA, 3rd, Gostout BS, Shah C, et al. Validation of referral guidelines for women with pelvic masses. Obstetrics & Gynecology. 2005;105(1):35-41.

68. Jain KA. Prospective evaluation of adnexal masses with endovaginal gray-scale and duplex and color Doppler US: correlation with pathologic findings. Radiology. 1994;191(1):63-7.

69. Jeschke U, Wiest I, Schumacher AL, Kupka M, Rack B, Stahn R, et al. Determination of MUC1 in sera of ovarian cancer patients and in sera of patients with benign changes of the ovaries with CA15-3, CA27.29, and PankoMab. Anticancer Research. 2012;32(5):2185-9.

70. Jokubkiene L, Sladkevicius P, Valentin L. Does three-dimensional power Doppler ultrasound help in discrimination between benign and malignant ovarian masses? Ultrasound in Obstetrics & Gynecology. 2007;29(2):215-25.

71. Kalghatgi-Kulkarni K, Kushtagi P. Ovarian crescent sign and sonomorphological indices in preoperative determination of malignancy in adnexal masses. Indian Journal of Medical Sciences. 2008;62(12):477-83.

72. Kappelmayer J, Antal-Szalmas P, Nagy B, Jr. Human epididymis protein 4 (HE4) in laboratory medicine and an algorithm in renal disorders. Clinica Chimica Acta. 2015;438:35-42.

73. Karadag B, Kocak M, Kayikcioglu F, Ercan F, Dilbaz B, Kose M, et al. Risk for malignant and borderline ovarian neoplasms following basic preoperative evaluation by ultrasonography, ca125 level and age. Asian Pacific Journal of Cancer Prevention: Apjcp. 2014;15(19):8489-93.

74. Karakaya BK, Ozgu E, Kansu HC, Evliyaoglu O, Sarikaya E, Coskun B, et al. Evaluation of Probably Benign Adnexal Masses in Postmenopausal Women. Revista Brasileira de Ginecologia e Obstetricia. 2017;39(5):229-34.

75. Karlsen MA, Sandhu N, Hogdall C, Christensen IJ, Nedergaard L, Lundvall L, et al. Evaluation of HE4, CA125, risk of ovarian malignancy algorithm (ROMA) and risk of malignancy index (RMI) as diagnostic tools of epithelial ovarian cancer in patients with a pelvic mass. Gynecologic Oncology. 2012;127(2):379-83.

76. Khashper A, Addley HC, Abourokbah N, Nougaret S, Sala E, Reinhold C. T2-hypointense adnexal lesions: an imaging algorithm. RadioGraphics. 2012;32(4):1047-64.

77. Khazendar S, Sayasneh A, Al-Assam H, Du H, Kaijser J, Ferrara L, et al. Automated characterisation of ultrasound images of ovarian tumours: the diagnostic accuracy of a support vector machine and image processing with a local binary pattern operator. Facts Views & Vision in Obgyn. 2015;7(1):7-15.

78. Koneczny J, Czekierdowski A, Florczak M, Poziemski P, Stachowicz N, Borowski D. The use of sonographic subjective tumor assessment, IOTA logistic regression model 1, IOTA Simple Rules and GI-RADS system in the preoperative prediction of malignancy in women with adnexal masses. Ginekologia Polska. 2017;88(12):647-53.

79. Kumazawa S, Umezu T, Kanayama Y, Kamiyama N, Suzuki S, Mizuno M, et al. Contrast-enhanced ultrasonography using Sonazoid is useful for diagnosis of malignant ovarian tumors: Comparison with Doppler ultrasound. Journal of Medical Ultrasonics. 2013;40(1):81-4.

80. Kupesic S, Vujisic S, Kurjak A, Mihaljevic D, Radosevic S. Preoperative assessment of ovarian tumors by CA 125 measurement and transvaginal color Doppler ultrasound. Acta Medica Croatica. 2002;56(1):3-10.

81. Kurjak A, Kupesic S, Anic T, Kosuta D. Three-dimensional ultrasound and power doppler improve the diagnosis of ovarian lesions. Gynecologic Oncology. 2000;76(1):28-32.

Table 15. Excluded studies: index test not applicable *(Continued)*

82. Kurjak A, Kupesic S, Sparac V, Prka M, Bekavac I. The detection of stage I ovarian cancer by three-dimensional sonography and power Doppler. Gynecologic Oncology. 2003;90(2):258-64.

83. Kurjak A, Predanic M. New scoring system for prediction of ovarian malignancy based on transvaginal color doppler sonography. Journal of Ultrasound in Medicine. 1992;11(12):631-8.

84. Kuyumcuoglu U, Guzel AI, Celik Y, Erdemoglu M, Komek H. 18F-FDG PET-CT and USG/CT in benign and malignant ovarian tumors with postoperative histopathological correlation. Ginekologia Polska. 2011;82(8):602-6.

85. Lee TS, Kim JW, Park NH, Song YS, Kang SB, Lee HP. Assessing clinical performance of gynecology residents: sonographic evaluation of adnexal masses based on morphological scoring systems. Ultrasound in Obstetrics & Gynecology. 2005;26(7):776-9.

86. Lim J, Chew G. Evaluating the use of intraoperative frozen section in the diagnosis of ovarian neoplasms: SGH experience. International Journal of Gynecological Cancer. 2015;1):486.

87. Longoria TC, Ueland FR, Zhang Z, Chan DW, Smith A, Fung ET, Munroe DG, Bristow RE. Clinical performance of a multivariate index assay for detecting early-stage ovarian cancer. American Journal of Obstetrics & Gynecology 2014;210(1):78.e1-9. [Other: Secondary publication]

88. Lu C, Van Gestel T, Suykens JA, Van Huffel S, Vergote I, Timmerman D. Preoperative prediction of malignancy of ovarian tumors using least squares support vector machines. Artificial intelligence in medicine. 2003;28(3):281-306.

89. Lucidarme O, Akakpo JP, Granberg S, Sideri M, Levavi H, Schneider A, et al. A new computer-aided diagnostic tool for non-invasive characterisation of malignant ovarian masses: Results of a multicentre validation study. European Radiology. 2010;20(8):1822-30.

90. Lv XL, Zhu Y, Liu JW, Ai H. The application value of the detection of the level of tissue polypeptide antigen, ovarian cancer antigen X1, cathepsin L and CA125 on the diagnosis of epithelial ovarian cancer. European Review for Medical & Pharmacological Sciences. 2016;20(24):5113-6.

91. Ma S, Shen K, Lang J. A risk of malignancy index in preoperative diagnosis of ovarian cancer. Chinese Medical Journal. 2003;116(3):396-9

92. Mahyuddin AP, Liu L, Zhao C, Kothandaraman N, Salto-Tellez M, Pang B, et al. Diagnostic accuracy of haptoglobin within ovarian cyst fluid as a potential point-of-care test for epithelial ovarian cancer: an observational study. BJOG: An International Journal of Obstetrics & Gynaecology. 2018;125(4):421-31.

93. Majeed H, Ramzan A, Imran F, Mahfooz ur R. Validity of resistive index for the diagnosis of malignant ovarian masses. JPMA - Journal of the Pakistan Medical Association. 2011;61(11):1104-7.

94. Malla VG, Tuteja A, Singh P. Rethinking the role of serum cancer antigen 125 and risk of malignancy index in Indian women with ovarian masses: Newer perspectives and review of literature. Journal of SAFOG. 2018;10(2):110-7.

95. Manganaro L, Anastasi E, Porpora MG, Vinci V, Saldari M, Bernardo S, et al. Biparametric Magnetic Resonance Imaging as an Adjunct to CA125 and HE4 to Improve Characterization of Large Ovarian Masses. Anticancer Research. 2015;35(11):6341-51.

96. Maturen KE, Blaty AD, Wasnik AP, Patel-Lippmann K, Robbins JB, Barroilhet L, et al. Risk Stratification of Adnexal Cysts and Cystic Masses: Clinical Performance of Society of Radiologists in Ultrasound Guidelines. Radiology. 2017;285(2):650-9.

97. Meng W, Ying W, Qichao Z, Ping L, Jie T. Clinical value of combining transvaginal contrast-enhanced ultrasonography with serum human epididymisprotein-4 and the resistance index for early-stage epithelial ovarian cancer. Saudi Medical Journal. 2017;38(6):592-7.

98. Meray O, Turkcuoglu I, Meydanli MM, Kafkasli A. Risk of malignancy index is not sensitive in detecting non-epithelial ovarian cancer and borderline ovarian tumor. Journal of the Turkishgerman Gynecological Association. 2010;11(1):22-6.

99. Mi D, Zhang YX, Wang CJ, Feng Q, Qi P, Chen SQ. Diagnostic and prognostic value of serum human epididymis protein 4 in patients with primary fallopian tube carcinoma. Journal of Obstetrics & Gynaecology Research. 2016;42(10):1326-35.

Table 15. Excluded studies: index test not applicable *(Continued)*

100. Mohammed ABF, Ahuga VK, Taha M. Validation of the Risk of Malignancy Index in primary evaluation of ovarian masses. Middle East Fertility Society Journal. 2014;19(4):324-8.
101. Moore RG, Hawkins DM, Miller MC, Landrum LM, Gajewski W, Ball JJ, et al. Combining clinical assessment and the Risk of Ovarian Malignancy Algorithm for the prediction of ovarian cancer. Gynecologic Oncology. 2014;135(3):547-51.
102. Moszynski R, Szpurek D, Smolen A, Sajdak S. Comparison of diagnostic usefulness of predictive models in preliminary differentiation of adnexal masses. International Journal of Gynecological Cancer. 2006;16(1):45-51.
103. Moszynski R, Szpurek D, Szubert S, Sajdak S. Analysis of false negative results of subjective ultrasonography assessment of adnexal masses. Ginekologia Polska. 2013;84(2):102-7.
104. Mousavi AS, Borna S, Moeinoddini S. Estimation of probability of malignancy using a logistic model combining color Doppler ultrasonography, serum CA125 level in women with a pelvic mass. International Journal of Gynecological Cancer. 2006;16:92-8.
105. Niewiadomska A, Bartczak M, Glowacka E, Zebrowska M, Swiatkowska E. Roma-the golden mean in the diagnosis of ovarian cancer? Tumor Biology. 2015;36 (Supplement 1):S74.
106. Ozcan HC, Balat O, Ugur MG, Kul S, Bozdag Z, Sucu S, et al. The management of adnexal masses in premenopausal patients: a ten-year retrospective study at a single center. European Journal of Gynaecological Oncology. 2017;38(3):372-7.
107. Paul P, Talukder S, Sangma MA, Saha PL, Begum M, Razi AZ, et al. Detection of Ovarian Tumor by Color Doppler Ultrasonography and CA-125. Mymensingh Medical Journal: MMJ. 2017;26(4):705-9.
108. Peces Rama A, Llanos Llanos MC, Sanchez Ferrer ML, Alcazar Zambrano JL, Martinez Mendoza A, Nieto Diaz A. Simple descriptors and simple rules of the International Ovarian Tumor Analysis (IOTA) Group: a prospective study of combined use for the description of adnexal masses. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2015;195:7-11.
109. Pereira PN, Sarian LO, Yoshida A, Araujo KG, Barros RHO, Baiao AC, et al. Accuracy of the ADNEX MR scoring system based on a simplified MRI protocol for the assessment of adnexal masses. Diagnostic and Interventional Radiology. 2018;24(2):63-71.
110. Petignat PG, G.; Vajda, D.; Joris, F. Ovarian Cancer: symptoms and pathology. Cases of a cantonal Cancer Registry (1989-1995). Schweizerische Medizinische Wochenschrift. 1997;127(48):1993-9.
111. Prompeler HJ, Madjar H, Sauerbrei W. Classification of adnexal tumors by transvaginal color Doppler. Gynecologic Oncology. 1996;61(3):354-63.
112. Prompeler HJ, Madjar H, Sauerbrei W, Lattermann U, Pfleiderer A. Diagnostic formula for the differentiation of adnexal tumors by transvaginal sonography. Obstetrics and Gynecology. 1997;89(3):428-33.
113. Rajendra Acharya U, Vinitha Sree S, Kulshreshtha S, Molinari F, Koh JEW, Saba L, et al. GyneScan: An improved online paradigm for screening of ovarian cancer via tissue characterization. Technology in Cancer Research and Treatment. 2014;13(6):529-40.
114. Ribeiro B, Vieira C, Bivar L, Monteiro C, Serrano P, Reis I. Surgical approach of adnexial masses - From clinical diagnosis to histological results. European Journal of Obstetrics Gynecology and Reproductive Biology. 2019;234:e44.
115. Rogulski L, Strzelczyk J. Simple ultrasound rules used by general gynecologists supplemented with Roma assessment in differentiating malignant and benign adnexal masses. International Journal of Gynecological Cancer. 2015;1):1479.
116. Rossi A, Braghin C, Soldano F, Isola M, Capodicasa V, Londero AP, et al. A proposal for a new scoring system to evaluate pelvic masses: Pelvic Masses Score (PMS). European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2011;157(1):84-8.
117. Ruiz M, Labauge P, Louboutin A, Limot O, Fauconnier A, Huchon C. External validation of the MR imaging scoring system for the management of adnexal masses. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2016;205:115-9.

Table 15. Excluded studies: index test not applicable *(Continued)*

118. Ruiz de Gauna B, Rodriguez D, Olartecoechea B, Auba M, Jurado M, Gomez Roig MD, et al. Diagnostic performance of IOTA simple rules for adnexal masses classification: a comparison between two centers with different ovarian cancer prevalence. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2015;191:10-4.
119. Sasaguri K, Yamaguchi K, Nakazono T, Mizuguchi M, Aishima S, Yokoyama M, et al. External validation of ADNEX MR SCORING system: a single-centre retrospective study. Clinical Radiology. 2019;74(2):131-9.
120. Schelling M, Braun M, Kuhn W, Bogner G, Gruber R, Gnirs J, et al. Combined transvaginal B-mode and color Doppler sonography for differential diagnosis of ovarian tumors: results of a multivariate logistic regression analysis. Gynecologic Oncology. 2000;77(1):78-86.
121. Seebacher V, Aust S, D'Andrea D, Grimm C, Reiser E, Tiringer D, et al. Development of a tool for prediction of ovarian cancer in patients with adnexal masses: Value of plasma fibrinogen. PLoS ONE [Electronic Resource]. 2017;12(8):e0182383.
122. Senel SA, Ozcam H, Ateser GB, Vatansever D. Risk of malignancy indices in differentiation of malignant adnexal masses from the benign adnexal masses. International Journal of Gynecological Cancer. 2015;1):1006.
123. Shah D, Shah S, Parikh J, Bhatt CJ, Vaishnav K, Bala DV. Doppler ultrasound: a good and reliable predictor of ovarian malignancy. Journal of Obstetrics & Gynaecology of India. 2013;63(3):186-9.
124. Shen ZY, He AQ, Xia GL, Wu MF, Li J, Ding YS. New sonographic morphology score for the differentiation of malignant from benign ovarian tumors. Journal of Obstetrics & Gynaecology Research. 2016;42(8):1000-12.
125. Simmons AR, Clarke CH, Badgwell DB, Lu Z, Sokoll LJ, Lu KH, et al. Validation of a Biomarker Panel and Longitudinal Biomarker Performance for Early Detection of Ovarian Cancer. International Journal of Gynecological Cancer. 2016;26(6):1070-7.
126. Sladkevicius P, Jokubkiene L, Valentin L. Contribution of morphological assessment of the vessel tree by three-dimensional ultrasound to a correct diagnosis of malignancy in ovarian masses. Ultrasound in Obstetrics & Gynecology. 2007;30(6):874-82.
127. Sladkevicius P, Valentin L. Interobserver agreement in describing the ultrasound appearance of adnexal masses and in calculating the risk of malignancy using logistic regression models. Clinical Cancer Research. 2015;21(3):594-601.
128. Stiekema A, Lok CAR, Kenter GG, van Driel WJ, Vincent AD, Korse CM. A predictive model combining human epididymal protein 4 and radiologic features for the diagnosis of ovarian cancer. Gynecologic Oncology. 2014;132(3):573-7.
129. Strolyte D, Celkiene I, Paskauskas S, Vaitkiene D, Gaurilcikas A. Evaluation of a transvaginal ultrasound report quality in cases of borderline ovarian tumors: A retrospective study. International Journal of Gynecological Cancer. 2015;1):1351.
130. Sutantawibul A, Ruangvutilert P, Sunsaneevithayakul P, Boriboonhirunsarn D. A model for malignancy probability prediction of adnexal masses. Journal of the Medical Association of Thailand. 2003;86(8):742-9.
131. Szpurek D, Moszyniki R, Sajdak S. Clinical value of the ultrasound Doppler index in determination of ovarian tumor malignancy. European Journal of Gynaecological Oncology. 2004;25(4):442-4.
132. Szpurek D, Moszynski R, Smolen A, Sajdak S. Artificial neural network computer prediction of ovarian malignancy in women with adnexal masses. International Journal of Gynaecology & Obstetrics. 2005;89(2):108-13.
133. Szpurek D, Moszynski R, Zietkowiak W, Spaczynski M, Sajdak S. An ultrasonographic morphological index for prediction of ovarian tumor malignancy. European Journal of Gynaecological Oncology. 2005;26(1):51-4.
134. Szubert S, Lojewski A, Moszynski R, Lisowski A, Sajdak S, Szpurek D. Ultrasonographic features and CA125 levels of hormonally active ovarian tumors. Ginekologia Polska. 2016;87(4):254-9.
135. Tailor A, Jurkovic D, Bourne TH, Collins WP, Campbell S. Sonographic prediction of malignancy in adnexal masses using multivariate logistic regression analysis. Ultrasound in Obstetrics & Gynecology. 1997;10(1):41-7.

Table 15. Excluded studies: index test not applicable *(Continued)*

136. Tailor A, Jurkovic D, Bourne TH, Natucci M, Collins WP, Campbell S. Comparison of transvaginal color Doppler imaging and color Doppler energy for assessment of intraovarian blood flow. Obstetrics & Gynecology. 1998;91(4):561-7.

137.Tajima A, Suzuki C, Kikuchi I, Kasahara H, Koizumi A, Nojima M, et al. Efficacy of the echo pattern classification of ovarian tumors 2000 in conjunction with transvaginal ultrasonography for diagnosis of ovarian masses. Journal of Medical Ultrasonics. 2016;43(2):249-55.

138. Takac I. Receiver operating characteristic curves of transvaginal Doppler blood flow measurements in benign and malignant adnexal tumors. Journal of Ultrasound in Medicine. 1998;17(10):637-42.

139.Takmaz OK, Angin AD, PiRiMoglu ZM. The worthiness of ovary symptom index, ca125 and transvaginal ultrasonography in screening malignancies of adnexial masses. Zeynep Kamil Tip Bulteni. 2018;49(1):54-8.

140.Tanaka YO, Okada S, Satoh T, Matsumoto K, Oki A, Saida T, et al. Differentiation of epithelial ovarian cancer subtypes by use of imaging and clinical data: a detailed analysis. Cancer Imaging. 2016;16:3.

141. Tay SK, Tan YY. Risk factors and a risk scoring system for the prediction of malignancy in ovarian cysts. Australian & New Zealand Journal of Obstetrics & Gynaecology. 1992;32(4):341-5.

142. Tekay A, Jouppila P. Validity of pulsatility and resistance indices in classification of adnexal tumors with transvaginal color Doppler ultrasound. Ultrasound in Obstetrics & Gynecology. 1992;2(5):338-44.

143. Tepper R, Keselbrener L, Manor M, Eyal S, Beyth Y, Zimmer Y, et al. Decay constant of Doppler flow waveform as a possible indicator of ovarian malignancy. Ultrasound in Medicine & Biology. 1997;23(8):1171-7.

144. Testa AC, Ferrandina G, Fruscella E, Van Holsbeke C, Ferrazzi E, Leone FPG, et al. The use of contrasted transvaginal sonography in the diagnosis of gynecologic diseases: a preliminary study. Journal of Ultrasound in Medicine. 2005;24(9):1267-78.

145.Timmerman D, Van Calster B, Testa A, Savelli L, Fischerova D, Froyman W, et al. Predicting the risk of malignancy in adnexal masses based on the Simple Rules from the International Ovarian Tumor Analysis group. American Journal of Obstetrics & Gynecology. 2016;214(4):424-37.

146. Timmerman D, Schwarzler P, Collins WP, Claerhout F, Coenen M, Amant F, et al. Subjective assessment of adnexal masses with the use of ultrasonography: an analysis of interobserver variability and experience. Ultrasound in Obstetrics & Gynecology. 1999;13(1):11-6.

147. Timmerman D, Verrelst H, Bourne TH, De Moor B, Collins WP, Vergote I, et al. Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses. Ultrasound in Obstetrics & Gynecology. 1999;13(1):17-25.

148.Tokmak A, Yildirim G, Oztas E, Akar S, Erkenekli K, Gulsen P, et al. Use of Neutrophil-to-Lymphocyte Ratio Combined With CA-125 to Distinguish Endometriomas From Other Benign Ovarian Cysts. Reproductive Sciences. 2016;23(6):795-802.

149.Tongsong T, Tinnangwattana D, Vichak-Ururote L, Tontivuthikul P, Charoenratana C, Lerthiranwong T. Comparison of Effectiveness in Differentiating Benign from Malignant Ovarian Masses between IOTA Simple Rules and Subjective Sonographic Assessment. Asian Pacific Journal of Cancer Prevention: Apjcp. 2016;17(9):4377-80.

150. Tongsong T, Wanapirak C, Neeyalavira V, Khunamornpong S, Sukpan K. E-flow doppler indices for prediction of benign and malignant ovarian tumors. Asian Pacific Journal of Cancer Prevention: Apjcp. 2009;10(1):139-42.

151.Torky HA, Sherif A, Abo-Louz A, Ali M, Ahmed A, Ali A. Evaluation of Serum Nidogen-2 as a Screening and Diagnostic Tool for Ovarian Cancer. Gynecologic & Obstetric Investigation. 2018;83(5):461-5.

152. Twickler DM, Forte TB, Santos-Ramos R, McIntire D, Harris P, Scott D. The Ovarian Tumor Index predicts risk for malignancy. Cancer. 1999;86(11):2280-90.

153. Ueland FR, Desimone CP, Seamon LG, Miller RA, Goodrich S, Podzielinski I, et al. Effectiveness of a multivariate index assay in the preoperative assessment of ovarian tumors. Obstetrics & Gynecology 2011;6:1289-97. [Other: Secondary publication]

Table 15. Excluded studies: index test not applicable *(Continued)*

154. Ueland FR, DePriest PD, Pavlik EJ, Kryscio RJ, van Nagell JR, Jr. Preoperative differentiation of malignant from benign ovarian tumors: the efficacy of morphology indexing and Doppler flow sonography. Gynecologic Oncology. 2003;91(1):46-50.

155. Urban RR, Pappas TC, Bullock RG, Munroe DG, Bonato V, Agnew K, et al. Combined symptom index and second-generation multivariate biomarker test for prediction of ovarian cancer in patients with an adnexal mass. Gynecologic Oncology. 2018;150(2):318-23.

156. Urban RR, Smith A, Agnew K, Bonato V, Goff BA. Evaluation of a Validated Biomarker Test in Combination with a Symptom Index to Predict Ovarian Malignancy. International Journal of Gynecological Cancer. 2017;27(2):233-8.

157. Utrilla-Layna J, Alcazar JL, Auba M, Laparte C, Olartecoechea B, Errasti T, et al. Performance of three-dimensional power Doppler angiography as third-step assessment in differential diagnosis of adnexal masses. Ultrasound in Obstetrics & Gynecology. 2015;45(5):613-7.

158. Valentin L. Gray scale sonography, subjective evaluation of the color Doppler image and measurement of blood flow velocity for distinguishing benign and malignant tumors of suspected adnexal origin. European journal of obstetrics, gynecology, and reproductive biology. 1997;72(1):63-72.

159. Valentin L. Prospective cross-validation of Doppler ultrasound examination and gray-scale ultrasound imaging for discrimination of benign and malignant pelvic masses. Ultrasound in Obstetrics & Gynecology. 1999;14(4):273-83.

160. Valentin L. Pattern recognition of pelvic masses by gray-scale ultrasound imaging: the contribution of Doppler ultrasound. Ultrasound in Obstetrics & Gynecology. 1999;14(5):338-47.

161. Valentin L. Comparison of Lerner score, Doppler ultrasound examination, and their combination for discrimination between benign and malignant adnexal masses. Ultrasound in Obstetrics & Gynecology. 2000;15(2):143-7.

162. Valentin L, Hagen B, Tingulstad S, Eik-Nes S. Comparison of 'pattern recognition' and logistic regression models for discrimination between benign and malignant pelvic masses: a prospective cross validation. Ultrasound in Obstetrics & Gynecology. 2001;18(4):357-65.

163. Valentin L, Sladkevicius P, Marsal K. Limited contribution of Doppler velocimetry to the differential diagnosis of extrauterine pelvic tumors. Obstetrics & Gynecology. 1994;83(3):425-33.

164. Van Calster B, Timmerman D, Lu C, Suykens JAK, Valentin L, Van Holsbeke C, et al. Preoperative diagnosis of ovarian tumors using Bayesian kernel-based methods. Ultrasound in Obstetrics & Gynecology. 2007;29(5):496-504.

165. Van Calster B, Timmerman D, Nabney I, Valentin L, Testa A, Van Holsbeke C, et al. Using Bayesian neural networks with ARD input selection to detect malignant ovarian masses prior to surgery. Neural Comput & Applic. 2008;17(5-6):489-500.

166. Van Calster B, Valentin L, Van Holsbeke C, Testa AC, Bourne T, Van Huffel S, et al. Polytomous diagnosis of ovarian tumors as benign, borderline, primary invasive or metastatic: development and validation of standard and kernel-based risk prediction models. BMC Medical Research Methodology. 2010;10:96.

167. van Haaften-Day C, Shen Y, Xu FJ, Yu YH, Berchuck A, Havrilesky LJ, et al. OVX1, macrophage-colony stimulating factor, and CA-125-II as tumor markers for epithelial ovarian carcinoma - A critical appraisal. Cancer. 2001;92(11):2837-44.

168. Van Holsbeke C, Van Calster B, Bourne T, Ajossa S, Testa AC, Guerriero S, et al. External validation of diagnostic models to estimate the risk of malignancy in adnexal masses. Clinical Cancer Research. 2012;18(3):815-25.

169. Van Holsbeke C, Daemen A, Yazbek J, Holland TK, Bourne T, Mesens T, et al. Ultrasound experience substantially impacts on diagnostic performance and confidence when adnexal masses are classified using pattern recognition. Gynecologic and obstetric investigation. 2010;69(3):160-8.

170. Van Holsbeke C, Van Belle V, Leone FPG, Guerriero S, Paladini D, Melis GB, et al. Prospective external validation of the 'ovarian crescent sign' as a single ultrasound parameter to distinguish between benign and malignant adnexal pathology. Ultrasound in Obstetrics & Gynecology. 2010;36(1):81-7.

Table 15. Excluded studies: index test not applicable *(Continued)*

171. Van Holsbeke C, Van Calster B, Testa AC, Domali E, Lu C, Van Huffel S, et al. Prospective internal validation of mathematical models to predict malignancy in adnexal masses: results from the international ovarian tumor analysis study. Clinical Cancer Research. 2009;15(2):684-91.
172. Wang T, Yang Y, Zhu Q. A three-parameter logistic model to characterize ovarian tissue using polarization-sensitive optical coherence tomography. Biomedical Optics Express. 2013;4(5):772-7.
173. Ware Miller, R. W, Smith, A, DeSimone, C. P, et al. Performance of the American College of Obstetricians and Gynecologists' Ovarian Tumor Referral Guidelines With a Multivariate Index Assay.
174. Watrowski R, Zeillinger R. Simple laboratory score improves the preoperative diagnosis of adnexal mass. Tumour Biology. 2016;37(4):4343-9.
175. Winarto H, Laihad BJ, Nuranna L. Modification of cutoff values for HE4, CA125, the Risk of Malignancy Index, and the Risk of Malignancy Algorithm for ovarian cancer detection in Jakarta, Indonesia. Asian Pacific Journal of Cancer Prevention: Apjcp. 2014;15(5):1949-53.
176. Yadav G, Singh P, Gothwal M, Kathuria P. Correlation between pre-operative clinical diagnosis, imaging and histopathology of adnexal masses: A cross-sectional observational study. International Journal of Gynecological Cancer. 2018;28 (Supplement 2):216.
177. Yamamoto Y, Kunimi Y, Wakikawa A, Ueno A, Nagai R, Komatsu J, et al. Comparison of benign, borderline, and malignant mucinous ovarian tumors. European Journal of Gynaecological Oncology. 2018;39(5):751-4.
178. Yazbek J, Ameye L, Testa AC, Valentin L, Timmerman D, Holland TK, et al. Confidence of expert ultrasound operators in making a diagnosis of adnexal tumor: effect on diagnostic accuracy and interobserver agreement. Ultrasound in Obstetrics & Gynecology. 2010;35(1):89-93.
179. Yazbek J, Ameye L, Timmerman D, Testa AC, Valentin L, Holland TK, et al. Use of ultrasound pattern recognition by expert operators to identify borderline ovarian tumors: a study of diagnostic performance and interobserver agreement. Ultrasound in Obstetrics & Gynecology. 2010;35(1):84-8.
200. Yoshida A, Derchain SF, Pitta DR, Crozatti N, Andrade L, da Silva RF, et al. Preoperative measurement of serum C-reactive protein: is it useful in the differential diagnosis of adnexal masses? International Journal of Biological Markers. 2017;32(1):e83-e9.
201. Zhang P, Wang C, Cheng L, Zhang P, Guo L, Liu W, et al. Development of a multi-marker model combining HE4, CA125, progesterone, and estradiol for distinguishing benign from malignant pelvic masses in postmenopausal women. Tumour Biology. 2016;37(2):2183-91.
202. Zhang Q, Wang CR, Yu JP, Ma Q, Xu WW. The Establishment of an HE4-CLIA Method and the Combined Analysis of HE4 and CA125 in Ovarian Cancer. Journal of Clinical Laboratory Analysis. 2016;30(5):709-18.
203. Zhao T, Hu W. CA125 and HE4: Measurement Tools for Ovarian Cancer. Gynecologic & Obstetric Investigation. 2016;81(5):430-5.
204. Zhang T, Li F, Liu J, Zhang S. Diagnostic performance of the Gynecology Imaging Reporting and Data System for malignant adnexal masses. International Journal of Gynecology and Obstetrics. 2017;137(3):325-31.
205. Zheng H, Tie Y, Wang X, Yang Y, Wei X, Zhao X. Assessment of the diagnostic value of using serum CA125 and GI-RADS system in the evaluation of adnexal masses. Medicine. 2019;98(7):e14577.
206. Zhu L, Xia C, Chen Y, Liu J, Nie K, Xiao Y, et al. Ultrasound/CT combined with CA125 in the diagnosis and prognosis prediction of primary ovarian tumors. Minerva Medica. 2018;109(6):489-91.
13999
Seebacher 2017 (duplicate)

Table 16. Excluded studies: no translation

No translation
1. Feng A, Zhang L, Chen YJ. The significance of serum HE4, CA125 and ROMA indexes in diagnosis of ovarian cancer to different menopausal status. [Chinese]. Journal of Xi'an Jiaotong University (Medical Sciences). 2018;39(1):78-83.
2. Gao HF, Tian H, Zhao QJ. Diagnostic value of serum human epididymis protein 4 and carbohydrate antigen 125 combined with risk of ovarian malignancy algorithm index for ovarian cancer. [Chinese]. Cancer Research and Clinic. 2018;30(11):757-62.
3. Gonzalez-Burgos OM, Alvarez-Licona NE, Lever-Rosas CD. Comparison of three ultrasound index in evaluating the risk of malignancy of adnexal tumors. [Spanish]. Ginecologia y Obstetricia de Mexico. 2018;86(8):519-29.
4. He P, Wu Q, Sun L, Wang J, Wang L, Han J, et al. Comparison of ADNEX model, simple rules risk model and risk of malignancy index in diagnosis of benign and malignant ovarian tumors. [Chinese]. Chinese Journal of Medical Imaging Technology. 2019;35(1):104-7.
5. Janas L, Glowacka E, Wilczynski JR, Malinowski A, Nowak M. [Evaluation of applicability of HE4 and ROMA in the preoperative diagnosis of adnexal masses]. Ginekologia Polska. 2015;86(3):193-7.
6. Joyeux E, Miras T, Masquin I, Duglet PE, Astruc K, Douvier S. [Before surgery predictability of malignant ovarian tumors based on ADNEX model and its use in clinical practice]. Gynecologie, Obstetrique & Fertilite. 2016;44(10):557-64.
7. Knafel A, Banas T, Nocun A, Wiechec M, Jach R, Ludwin A, et al. The Prospective External Validation of International Ovarian Tumor Analysis (IOTA) Simple Rules in the Hands of Level I and II Examiners. Ultraschall in der Medizin. 2016;37(5):516-23.
8. Liu J, Chen Q, Lyu G. Comparison of ultrasound IOTA simple rules and GI-RADS ultrasonographic stratification in diagnosis of ovarian neoplasms. [Chinese]. Chinese Journal of Medical Imaging Technology. 2017;33(5):739-42.
9. Sandal K, Polat M, Yassa M, Gunay T, Erdem GY, Guzin K. Comparision of "risk of malignancy indices" and "assesment of different neoplasia in the adnexa" (ADNEX) model as preoperative malignancy evaluation methods for adnexal masses. [Turkish]. Zeynep Kamil Tip Bulteni. 2018;49(4):324-9.
10. Tao X, Gu W, Zhou X, Zhao C. [New markers for diagnostic and prognostic evaluation of ovarian clear cell carcinoma]. Chung-Hua Ping Li Hsueh Tsa Chih - Chinese Journal of Pathology. 2015;44(9):686-8.

Table 17. Excluded studies: population not applicable

Population not applicable
1. Dual-marker algorithm better predicts risk for ovarian cancer. Contemporary OB/GYN. 2010;55(11):22-.
2. Adonakis GL, Paraskevaidis E, Tsiga S, Seferiadis K, Lolis DE. A combined approach for the early detection of ovarian cancer in asymptomatic women. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 1996;65(2):221-5.
3. Agarwal S, Bodurka DC. Symptom research in gynecologic oncology: A review of available measurement tools. Gynecologic Oncology. 2010;119(2):384-9.
4. Alcazar JL, Castillo G. Comparison of 2-dimensional and 3-dimensional power-Doppler imaging in complex adnexal masses for the prediction of ovarian cancer. American Journal of Obstetrics & Gynecology. 2005;192(3):807-12.
5. Alcazar JL, Errasti T, Zornoza A, Minguez JA, Galan MJ. Transvaginal color Doppler ultrasonography and CA-125 in suspicious adnexal masses. International Journal of Gynecology & Obstetrics. 1999;66(3):255-61.
6. Alcazar JL, Guerriero S, Pascual MA, Ajossa S, Olartecoechea B, Hereter L. Clinical and sonographic features of uncommon primary ovarian malignancies. Journal of Clinical Ultrasound. 2012;40(6):323-9.

Table 17. Excluded studies: population not applicable *(Continued)*

7. Alcazar JL, Iturra A, Sedda F, Auba M, Ajossa S, Guerriero S, et al. Three-dimensional volume off-line analysis as compared to real-time ultrasound for assessing adnexal masses. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2012;161(1):92-5.
8. Alcazar JL, Miranda D, Unanue A, Novoa E, Aleman S, Madariaga L. CA-125 levels in predicting optimal cytoreductive surgery in patients with advanced epithelial ovarian carcinoma. International Journal of Gynecology and Obstetrics. 2004;84(2):173-4.
9. Anastasi E, Manganaro L, Granato T, Benedetti Panici P, Frati L, Porpora MG. Is CA72-4 a useful biomarker in differential diagnosis between ovarian endometrioma and epithelial ovarian cancer? Disease Markers. 2013;35(5):331-5.
10. Arakawa N, Miyagi E, Nomura A, Morita E, Ino Y, Ohtake N, et al. Secretome-based identification of TFPI2, a novel serum biomarker for detection of ovarian clear cell adenocarcinoma. Journal of Proteome Research. 2013;12(10):4340-50.
11. Araujo KG, Jales RM, Pereira PN, Yoshida A, de Angelo Andrade L, Sarian LO, et al. Performance of the IOTA ADNEX model in preoperative discrimination of adnexal masses in a gynecological oncology center. Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2017;49(6):778-83.
12. Awwad Y. New screening strategy may catch ovarian cancer at early stages. Women's Health. 2013;9(6):509.
13. BA IBA. Symptoms attributed to ovarian cancer by women with the disease1997. 122, 7-8, 30 passim p.
14. Bashizadeh-Fakhar H, Rezaie-Tavirani M, Zali H, Faraji R, Nejad EK, Aghazadeh M. The diagnostic value of serum CEA, CA-125, and ROMA index in low-grade serous ovarian cancer. International Journal of Cancer Management. 2018;11 (5) (no pagination)(e63397).
15. Bensaid C, Le Frere Belda MA, Metzger U, Larousserie F, Clement D, Chatellier G, et al. Performance of laparoscopy in identifying malignant ovarian cysts. Surgical endoscopy. 2006;20(9):1410-4.
16. Blyuss O, Gentry-Maharaj A, Fourkala EO, Ryan A, Zaikin A, Menon U, et al. Serial Patterns of Ovarian Cancer Biomarkers in a Prediagnosis Longitudinal Dataset. BioMed Research International. 2015;2015:681416.
17. Bosse K, Rhiem K, Wappenschmidt B, Hellmich M, Madeja M, Ortmann M, et al. Screening for ovarian cancer by transvaginal ultrasound and serum CA125 measurement in women with a familial predisposition: A prospective cohort study. Gynecologic Oncology. 2006;103(3):1077-82.
18. Botta G, Zarcone R. Trans-vaginal ultrasound examination of ovarian masses in premenopausal women. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 1995;62(1):37-41.
19. Bourne TH, Campbell S, Reynolds KM, Whitehead MI, Hampson J, Royston P, et al. Screening for early familial ovarian cancer with transvaginal ultrasonography and colour blood flow imaging. BMJ. 1993;306(6884):1025-9.
20. Bristow RE, Nugent AC, Zahurak ML, Khouzhami V, Fox HE. Impact of surgeon specialty on ovarian-conserving surgery in young females with an adnexal mass. Journal of Adolescent Health. 2006;39(3):411-6.
21. Burger RA, Darcy KM, DiSaia PJ, Monk BJ, Grosen EA, Gatanaga T, et al. Association between serum levels of soluble tumor necrosis factor receptors/CA 125 and disease progression in patients with epithelial ovarian malignancy: a gynecologic oncology group study. Cancer. 2004;101(1):106-15.
22. Chang HW, Lee SM, Goodman SN, Singer G, Cho SKR, Sokoll LJ, et al. Assessment of plasma DNA levels, allelic imbalance, and CA 125 as diagnostic tests for cancer. Journal of the National Cancer Institute. 2002;94(22):1697-703.
23. Chen J, Chang C, Huang HC, Chung YC, Huang HJ, Liou WS, et al. Differentiating between borderline and invasive malignancies in ovarian tumors using a multivariate logistic regression model. Taiwanese Journal of Obstetrics & Gynecology. 2015;54(4):398-402.
24. Cho H-y, Kim K, Jeon Y-T, Kim Y-B, No JH. CA19-9 elevation in ovarian mature cystic teratoma: discrimination from ovarian cancer - CA19-9 level in teratoma. Medical Science Monitor. 2013;19:230-5.

Table 17. Excluded studies: population not applicable *(Continued)*

25. Cho H-Y, Kyung MS. Serum CA19-9 as a predictor of malignancy in primary ovarian mucinous tumors: a matched case-control study. Medical Science Monitor. 2014;20:1334-9.

26. Cho S, Cho H, Nam A, Kim HY, Choi YS, Park KH, et al. Neutrophil-to-lymphocyte ratio as an adjunct to CA-125 for the diagnosis of endometriosis. Fertility and Sterility. 2008;90(6):2073-9.

27. Chou CY, Chang CH, Yao BL, Kuo HC. Color Doppler ultrasonography and serum CA 125 in the differentiation of benign and malignant ovarian tumors. Journal of Clinical Ultrasound. 1994;22(8):491-6.

28. Dai S-Y, Hata K, Inubashiri E, Kanenishi K, Shiota A, Ohno M, et al. Does three-dimensional power Doppler ultrasound improve the diagnostic accuracy for the prediction of adnexal malignancy? Journal of Obstetrics and Gynaecology Research. 2008;34(3):364-70.

29. Davelaar EM, Schutter EMJ, Von Mensdorff-Pouilly S, Van Kamp GJ, Verstraeten RA, Kenemans P. Clinical and technical evaluation of the ACS:OV serum assay and comparison with three other CA125-detecting assays. Annals of Clinical Biochemistry. 2003;40(6):663-73.

30. Davies Q, Symonds IM, Perkins AC, Kerslake RW, Wastie ML, Worthington BS, et al. Magnetic resonance imaging, OC125 immunoscintigraphy and serum CA125 levels in the management of patients with suspected primary or recurrent ovarian carcinoma. Journal of Obstetrics and Gynaecology. 1996;16(2):108-16.

31. Dehaghani AS, Ghiam AF, Hosseini M, Mansouri S, Ghaderi A. Factors influencing serum concentration of CA125 and CA15-3 in Iranian healthy postmenopausal women. Pathology and Oncology Research. 2007;13(4):360-4.

32. Demidov VN, Lipatenkova J, Vikhareva O, Van Holsbeke C, Timmerman D, Valentin L. Imaging of gynecological disease (2): clinical and ultrasound characteristics of Sertoli cell tumors, Sertoli-Leydig cell tumors and Leydig cell tumors. Ultrasound in Obstetrics & Gynecology. 2008;31(1):85-91.

33. Devine PL, McGuckin MA, Quin RJ, Ward BG. Predictive value of the combination of serum markers, CA125, CASA and TPS in ovarian cancer. International Journal of Gynecological Cancer. 1995;5(3):170-8.

34. Dikensoy E, Balat O, Ugur MG, Ozkur A, Erkilic S. Serum CA-125 is a good predictor of benign disease in patients with postmenopausal ovarian cysts. European Journal of Gynaecological Oncology. 2007;28(1):45-7.

35. Epstein E, Van Calster B, Timmerman D, Nikman S. Subjective ultrasound assessment, the ADNEX model and ultrasound-guided tru-cut biopsy to differentiate disseminated primary ovarian cancer from metastatic non-ovarian cancer. Ultrasound in Obstetrics & Gynecology. 2016;47(1):110-6.

36. Fagotti A, Ludovisi M, De Blasis I, Virgilio B, Di Legge A, Mascilini F, et al. The sonographic prediction of invasive carcinoma in unilocular-solid ovarian cysts in premenopausal patients: a pilot study. Human Reproduction. 2012;27(9):2676-83.

37. Ferrazzi E, Lissoni AA, Dordoni D, Trio D, Redaelli L, Rusconi C, et al. Differentiation of small adnexal masses based on morphologic characteristics of transvaginal sonographic imaging - A multicenter study. Journal of Ultrasound in Medicine. 2005;24(11):1467-73.

38. Fishman DA, Cohen L, Blank SV, Shulman L, Singh D, Bozorgi K, et al. The role of ultrasound evaluation in the detection of early-stage epithelial ovarian cancer. American journal of obstetrics and gynecology. 2005;192(4):1214-21.

39. Fleischer AC, Lyshchik A, Jones HW, 3rd, Crispens MA, Andreotti RF, Williams PK, et al. Diagnostic parameters to differentiate benign from malignant ovarian masses with contrast-enhanced transvaginal sonography. Journal of Ultrasound in Medicine. 2009;28(10):1273-80.

40. Froyman W, Landolfo C, Bourne T, Cock BD, Testa A, Valentin L, et al. Performance of the RMI and IOTA ADNEX and Simple Rules risk model in the evaluation of adnexal masses not classifiable using the Easy Descriptors as first step. BJOG: An International Journal of Obstetrics and Gynaecology. 2016;123 (Supplement 2):83-4.

41. Froyman W, Wynants L, Landolfo C, Bourne T, Valentin L, Testa A, et al. Validation of the performance of International Ovarian Tumor Analysis (IOTA) methods in the diagnosis of early stage ovarian cancer in a non-screening population. Diagnostics. 2017;7 (2) (no pagination)(2).

Table 17. Excluded studies: population not applicable *(Continued)*

42. Ghosh E, Levine D. Recommendations for adnexal cysts: have the Society of Radiologists in Ultrasound consensus conference guidelines affected utilization of ultrasound? Ultrasound Quarterly. 2013;29(1):21-4.

43. Grab D, Flock F, Stohr I, Rieber A, Nussle S, Brambs HJ, et al. Premenopausal adnex masses: Diagnostic accuracy of sonography, magnetic resonance imaging, and positron emission tomography. [German] Diagnostische wertigkeit von sonographie, kernspintomographie und positronenemissions-tomographie zur beurteilung tier dignitat von adnextumoren bei pramenopausalen patientinnen. Geburtshilfe und Frauenheilkunde. 2000;60(11):544-52.

44. Grzybowski W, Beta J, Fritz A, Durczynski A, Bidzinski M, Grabiec M, et al. [Predictive value of CA 125 in detection of ovarian cancer in pre- and postmenopausal patients]. Ginekologia Polska. 2010;81(7):511-5.

45. Guerriero S, Ajossa S, Mais V, Risalvato A, Lai MP, Melis GB. The diagnosis of endometriomas using colour Doppler energy imaging. Human Reproduction. 1998;13(6):1691-5.

46. Guerriero S, Alcazar JL, Coccia ME, Ajossa S, Scarselli G, Boi M, et al. Complex pelvic mass as a target of evaluation of vessel distribution by color Doppler sonography for the diagnosis of adnexal malignancies: results of a multicenter European study. Journal of Ultrasound in Medicine. 2002;21(10):1105-11.

47. Hellstrom I. The HE4 (WFDC2) Protein Is a Biomarker for Ovarian Carcinoma1. Cancer Research. 2003;63:3695-700.

48. Hirai M, Hirai Y, Tsuchida T, Takada T, Iwase H, Utsugi K, et al. Stage IA ovarian cancers: comparison of sonographic findings and histopathologic types between patients with normal and elevated serum cancer antigen 125 levels. Journal of Ultrasound in Medicine. 2011;30(7):943-52.

49. Huang C-Q, Yang X-J, Yu Y, Wu H-T, Liu Y, Yonemura Y, et al. Cytoreductive surgery plus hyperthermic intraperitoneal chemotherapy improves survival for patients with peritoneal carcinomatosis from colorectal cancer: a phase II study from a Chinese center. PLoS ONE [Electronic Resource]. 2014;9(9):e108509.

50. Huhtinen K, Suvitie P, Hiissa J, Junnila J, Huvila J, Kujari H, et al. Serum HE4 concentration differentiates malignant ovarian tumours from ovarian endometriotic cysts. British Journal of Cancer. 2009;100(8):1315-9.

51. Iatrakis GM, Zervoudis S, Tsikouras P, Iatrakis DG, Bothou A, Chardavelas T, et al. A new risk malignancy index to predict ovarian cancer: A bicentric preliminary study. Journal of BUON. 2018;23(5):1380-3.

52. Ikeda Y, Hasegawa K, Kurosaki A, Miyara A, Hanaoka T, Shintani D, et al. The Risk of Ovarian Malignancy Algorithm (ROMA) as a Predictive Marker of Peritoneal Dissemination in Epithelial Ovarian Cancer Patients. Oncology research and treatment. 2015;38(6):276-81.

53. Ind T, Iles R, Shepherd J, Chard T. Serum concentrations of cancer antigen 125, placental alkaline phosphatase, cancer-associated serum antigen and free beta human chorionic gonadotrophin as prognostic markers for epithelial ovarian cancer. British Journal of Obstetrics and Gynaecology. 1997;104(9):1024-9.

54. Ionescu CA, Matei A, Navolan D, Dimitriu M, Bohaltea R, Neacsu A, et al. Correlation of ultrasound features and the Risk of Ovarian Malignancy Algorithm score for different histopathological subtypes of benign adnexal masses. Medicine (United States). 2018;97 (31) (no pagination)(e11762).

55. Jacobs I, Davies AP, Bridges J, Stabile I, Fay T, Lower A, et al. Prevalence screening for ovarian cancer in postmenopausal women by CA 125 measurement and ultrasonography. BMJ. 1993;306(6884):1030-4.

56. Jacobs IJ, Skates S, Davies AP, Woolas RP, Jeyerajah A, Weidemann P, et al. Risk of diagnosis of ovarian cancer after raised serum CA 125 concentration: a prospective cohort study. BMJ (Clinical research ed) [Internet]. 1996 [cited PENDING (UPDATE 91-08); (7069): [1355-8 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/049/CN-00251049/frame.html.

57. Johnson CC, Kessel B, Riley TL, Ragard LR, Williams CR, Xu JL, et al. The epidemiology of CA-125 in women without evidence of ovarian cancer in the Prostate, Lung, Colorectal and Ovarian Cancer (PLCO) Screening Trial. Gynecologic Oncology. 2008;110(3):383-9.

Table 17. Excluded studies: population not applicable *(Continued)*

58. Julian-Reynier C, Mancini J, Mouret-Fourme E, Gauthier-Villars M, Bonadona V, Berthet P, et al. Cancer risk management strategies and perceptions of unaffected women 5 years after predictive genetic testing for BRCA1/2 mutations. European Journal of Human Genetics. 2011;19(5):500-6.

59. Kalfa N, Sultan C. Juvenile ovarian granulosa cell tumor: a benign or malignant condition? Gynecological Endocrinology. 2009;25(5):299-302.

60. Kaymak O, Ustunyurt E, Okyay RE, Kalyoncu S, Mollamahmutoglu L. Role of transvaginal sonography and serum CA-125 values in determining the nature of the postmenopausal ovarian cysts. Journal of the Turkish German Gynecology Association. 2006;7(1):24-8.

61. Kelly PJ, Archbold P, Price JH, Cardwell C, McCluggage WG. Serum CA19.9 levels are commonly elevated in primary ovarian mucinous tumours but cannot be used to predict the histological subtype. Journal of Clinical Pathology. 2010;63(2):169-73.

62. Kim HS, Park NH, Chung HH, Kim JW, Song YS, Kang SB. Significance of preoperative serum CA-125 levels in the prediction of lymph node metastasis in epithelial ovarian cancer. Acta Obstetricia et Gynecologica Scandinavica. 2008;87(11):1136-42.

63. Kinay T, Unlubilgin E, Cirik DA, Kayikcioglu F, Akgul MA, Dolen I. The value of ultrasonographic tubo-ovarian abscess morphology in predicting whether patients will require surgical treatment. International Journal of Gynaecology & Obstetrics. 2016;135(1):77-81.

64. Kobayashi H, Yamada Y, Sado T, Sakata M, Yoshida S, Kawaguchi R, et al. A randomized study of screening for ovarian cancer: a multicenter study in Japan. International Journal of Gynecological Cancer [Internet]. 2008 [cited PENDING (UPDATE 91-08); (3): [414-20 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/881/CN-00636881/frame.html.

65. Kobayashi H, Yamada Y, Sado T, Sakata M, Yoshida S, Kawaguchi R, et al. Prevalence of ovarian cancer among women with a CA125 level of 35 U/ml or less. Gynecologic & Obstetric Investigation. 2008;65(2):133-8.

66. Koldjeski D, Kirkpatrick MK, Swanson M, Everett L, Brown S. Ovarian cancer: early symptom patterns. Oncology nursing forum. 2003;30(6):927-33.

67. Kondalsamy-Chennakesavan S, Hackethal A, Bowtell D, Australian Ovarian Cancer Study G, Obermair A. Differentiating stage 1 epithelial ovarian cancer from benign ovarian tumours using a combination of tumour markers HE4, CA125, and CEA and patient's age. Gynecologic Oncology. 2013;129(3):467-71.

68. Kurjak A, Zalud I, Alfirevic Z. Evaluation of adnexal masses with transvaginal color ultrasound. Journal of Ultrasound in Medicine. 1991;10(6):295-7.

69. Kotowicz B, Fuksiewicz M, Sobiczewski P, Spiewankiewicz B, Jonska-Gmyrek J, Skrzypczak M, et al. Clinical value of human epididymis protein 4 and the Risk of Ovarian Malignancy Algorithm in differentiating borderline pelvic tumors from epithelial ovarian cancer in early stages. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2015;194:141-6.

70. Kyung MS. Serum CA19-9 as a predictor of malignancy in primary ovarian mucinous tumors: a matched case-control study. Medical science monitor: international medical journal of experimental and clinical research. 2014;20:1334-9.

71. Lee EJ, Kim SH, Kim YH, Lee HJ. Is CA-125 an additional help to radiologic findings for differentiation borderline ovarian tumor from stage I carcinoma? Acta Radiologica. 2011;52(4):458-62.

72. Li QL, Wang CJ, Qi P, Zhang YX. Correlation of preoperative ROMA scores with clinical stage in epithelial ovarian cancer patients. Clinical & Translational Oncology: Official Publication of the Federation of Spanish Oncology Societes & of the National Cancer Institute of Mexico. 2017;19(10):1260-7.

73. Lowe KA, Andersen MR, Urban N, Paley P, Dresher CW, Goff BA. The temporal stability of the Symptom Index among women at high-risk for ovarian cancer. Gynecologic Oncology. 2009;114(2):225-30.

74. Lurie G, Wilkens LR, Thompson PJ, Matsuno RK, Carney ME, Goodman MT. Symptom presentation in invasive ovarian carcinoma by tumor histological type and grade in a multiethnic population: a case analysis. Gynecologic Oncology. 2010;119(2):278-84.

Table 17. Excluded studies: population not applicable *(Continued)*

75. Meisel M, Straube W, Weise J, Burkhardt B. A study of serum CASA and CA 125 levels in patients with ovarian carcinoma. Archives of Gynecology and Obstetrics. 1995;256(1):9-15.

76. Menon U, Jacobs I. Ovarian cancer screening in the general population. Ultrasound in Obstetrics and Gynecology. 2000;15(5):350-3.

77. Menon U, Talaat A, Jeyarajah AR, Rosenthal AN, MacDonald ND, Skates SJ, et al. Ultrasound assessment of ovarian cancer risk in postmenopausal women with CA125 elevation. British Journal of Cancer. 1999;80(10):1644-7.

78. Menon U, Talaat A, Rosenthal AN, Macdonald ND, Jeyerajah AR, Skates SJ, et al. Performance of ultrasound as a second line test to serum CA125 in ovarian cancer screening. BJOG: An International Journal of Obstetrics & Gynaecology. 2014;121 Suppl 7:35-9.

79. Moore RG, Miller MC, Steinhoff MM, Skates SJ, Lu KH, Lambert-Messerlian G, et al. Serum HE4 levels are less frequently elevated than CA125 in women with benign gynecologic disorders. American Journal of Obstetrics & Gynecology. 2012;206(4):351.e1-8.

80. Moszynski R, Szubert S, Szpurek D, Michalak S, Krygowska J, Sajdak S. Usefulness of the HE4 biomarker as a second-line test in the assessment of suspicious ovarian tumors. Archives of Gynecology & Obstetrics. 2013;288(6):1377-83.

81. Niemi RJ, Saarelainen SK, Luukkaala TH, Maenpaa JU. Reliability of preoperative evaluation of postmenopausal ovarian tumors. Journal of Ovarian Research. 2017;10 (1) (no pagination)(15).

82. Nigam PK, Jain A, Goyal P, Chitra R. Role of heat stable fraction of alkaline phosphatase as an adjunct to CA 125 in monitoring patients of epithelial ovarian carcinoma. Indian Journal of Clinical Biochemistry. 2005;20(2):43-7.

83. Nikolova T, Zivadinovic R, Nikolova N, Klisarovska V, Evtimovska N, Stanojevic M. Comparison of diagnostic performances of HE4, risk of malignancy algorithm and morphology index in discrimination of ovarian endometriosis from epithelial ovarian cancer in premenopausal women. International Journal of Gynecological Cancer. 2017;27 (Supplement 4):515.

84. Nishimura S, Ito YM, Tsuda H, Ohnishi Y, Kataoka F, Nomura H, et al. The sensitivity and specificity of a new formula to distinguish endometrioid type endometrial carcinoma from ovarian endometrial carcinoma. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2010;148(1):67-72.

85. Nohuz E, De Simone L, Chene G. Reliability of IOTA score and ADNEX model in the screening of ovarian malignancy in postmenopausal women. Journal of Gynecology Obstetrics and Human Reproduction. 2019;48(2):103-7.

86. Nunes N, Ambler G, Foo X, Naftalin J, Derdelis G, Widschwendter M, et al. Comparison of two protocols for the management of asymptomatic postmenopausal women with adnexal tumours-a randomised controlled trial of RMI/RCOG vs Simple Rules. British Journal of Cancer. 2017;116(5):584-91.

87. Olivier RI, Lubsen-Brandsma MAC, Verhoef S, van Beurden M. CA125 and transvaginal ultrasound monitoring in high-risk women cannot prevent the diagnosis of advanced ovarian cancer. Gynecologic Oncology. 2006;100(1):20-6.

88. Oltmann SC, Garcia N, Barber R, Huang R, Hicks B, Fischer A. Can we preoperatively risk stratify ovarian masses for malignancy? Journal of Pediatric Surgery. 2010;45(1):130-4.

89. Osman N, O'Leary N, Mulcahy E, Barrett N, Wallis F, Hickey K, et al. Correlation of serum CA125 with stage, grade and survival of patients with epithelial ovarian cancer at a single centre. Irish Medical Journal. 2008;101(8).

90. Paramasivam S, Tripcony L, Crandon A, Quinn M, Hammond I, Marsden D, et al. Prognostic importance of preoperative CA-125 in International Federation of Gynecology and Obstetrics stage I epithelial ovarian cancer: an Australian multicenter study. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2005;23(25):5938-42.

91. Park JW, Han M, Cho YJ. Four risk of malignancy indices in evaluation between borderline and malignant pelvic masses. International Journal of Gynecological Cancer. 2017;27 (Supplement 4):1621.

92. Pitts MK, Heywood W, Ryall R, Smith AM, Shelley JM, Richters J, et al. High prevalence of symptoms associated with ovarian cancer among Australian women. Australian & New Zealand Journal of Obstetrics & Gynaecology. 2011;51(1):71-8.

Table 17. Excluded studies: population not applicable *(Continued)*

93. Pitynski K, Sporek A, Lipinska I, Banas T, Ludwin A, Balajewicz-Nowak M. Significance of adding progesterone to the Risk of Ovarian Malignancy Algorithm for early stage ovarian cancer detection in patients with a pelvic mass: A single-center case-control study. Taiwanese Journal of Obstetrics & Gynecology. 2015;54(6):766-72.
94. Ra OS, Woo PJ, Myoungseok H, Woon BJ, Jean CY. Clinical utility of the risk of malignancy indices for preoperative differentiation between ovarian cancer and borderline ovarian tumor. International Journal of Gynecological Cancer. 2018;28 (Supplement 3):27.
95. Raza A, Mould T, Wilson M, Burnell M, Bernhardt L. Increasing the effectiveness of referral of ovarian masses from cancer unit to cancer center by using a higher referral value of the risk of malignancy index. International Journal of Gynecological Cancer. 2010;20(4):552-4.
96. Robertson DM, Cahir N, Burger HG, Mamers P, McCloud PI, Pettersson K, et al. Combined inhibin and CA125 assays in the detection of ovarian cancer. Clinical Chemistry. 1999;45(5):651-8.
97. Rogers EM, Casadiego Cubides G, Lacy J, Gerstle JT, Kives S, Allen L. Preoperative risk stratification of adnexal masses: can we predict the optimal surgical management? Journal of Pediatric & Adolescent Gynecology. 2014;27(3):125-8.
98. Rosen DG, Wang L, Atkinson JN, Yu Y, Lu KH, Diamandis EP, et al. Potential markers that complement expression of CA125 in epithelial ovarian cancer. Gynecologic Oncology. 2005;99(2):267-77.
99. Roupa Z, Faros E, Raftopoulos V, Tzavelas G, Kotrotsiou E, Sotiropoulou P, et al. Serum CA 125 combined with transvaginal ultrasonography for ovarian cancer screening. In Vivo. 2004;18(6):831-6.
100. Ruggeri G, Bandiera E, Zanotti L, Belloli S, Ravaggi A, Romani C, et al. HE4 and epithelial ovarian cancer: comparison and clinical evaluation of two immunoassays and a combination algorithm. Clinica Chimica Acta. 2011;412(15-16):1447-53.
101. Ryuko K, Iwanari O, Nakayama S, Iida K, Kitao M. Clinical evaluation of serum sialosyl-Tn antigen levels in comparison with CA 125 levels in gynecologic cancers. Cancer. 1992;69(9):2368-78.
102. Sarandakou A, Phocas I, Botsis D, Sikiotis K, Rizos D, Kalambokis D, et al. Tumour-associated antigens CEA, CA125, SCC and TPS in gynaecological cancer. European Journal of Gynaecological Oncology. 1998;19(1):73-7.
103. Saraswathi A, Malati T. Superiority of CA 125 over CA 19-9 and CEA for epithelial ovarian malignancies. Indian Journal of Clinical Biochemistry. 1995;10(1):23-8.
104. Sasaki H, Oda M, Ohmura M, Akiyama M, Liu C, Tsugane S, et al. Follow up of women with simple ovarian cysts detected by transvaginal sonography in the Tokyo metropolitan area. British Journal of Obstetrics & Gynaecology. 1999;106(5):415-20.
105. Seda HB, Selvi DG, Omer D, Levent A, Huseyin Y. A clinico-pathological evaluation of 194 patients with ovarian teratoma: 7-year experience in a single center. Ginekologia Polska. 2013;84(2):108-11.
106. Seow KM, Lin YH, Hsieh BC, Huang LW, Pan HS, Chang JZ, et al. Transvaginal three-dimensional ultrasonography combined with serum CA 125 level for the diagnosis of pelvic adhesions before laparoscopic surgery. Journal of the American Association of Gynecologic Laparoscopists. 2003;10(3):320-6.
107. Sharma A, Gentry-Maharaj A, Burnell M, Fourkala EO, Campbell S, Amso N, et al. Assessing the malignant potential of ovarian inclusion cysts in postmenopausal women within the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a prospective cohort study. BJOG: An International Journal of Obstetrics & Gynaecology. 2012;119(2):207-19.
108. Sichel F, Salaun V, Bar E, Gauduchon P, Malas JP, Goussard J, et al. Biological markers and ovarian carcinomas: galactosyltransferase, CA 125, isoenzymes of amylase and alkaline phosphatase. Clinica Chimica Acta. 1994;227(1-2):87-96.
109. Sobiczewski P, Danska-Bidzinska A, Rzepka J, Kupryjanczyk J, Gujski M, Bidzinski M, et al. Evaluation of selected ultrasonographic parameters and marker levels in the preoperative differentiation of borderline ovarian tumors and ovarian cancers. Archives of Gynecology & Obstetrics. 2012;286(6):1513-9.

Table 17. Excluded studies: population not applicable *(Continued)*

110. Sorensen SS, Mosgaard BJ. Combination of cancer antigen 125 and carcinoembryonic antigen can improve ovarian cancer diagnosis. Danish Medical Bulletin. 2011;58(11):A4331.

111. Stine JE, Suri A, Gehrig PA, Chiu M, Erickson BK, Huh WK, et al. Pre-operative imaging with CA125 is a poor predictor for granulosa cell tumors. Gynecologic Oncology. 2013;131(1):59-62.

112. Suh-Burgmann E, Hung Y-Y, Kinney W. Outcomes from ultrasound follow-up of small complex adnexal masses in women over 50. American Journal of Obstetrics & Gynecology. 2014;211(6):623.e1-7.

113. Takeshima N, Shimizu Y, Umezawa S, Hirai Y, Chen JT, Fujimoto I, et al. Combined assay of serum levels of ca125 and ca19-9 in endometrial carcinoma. Gynecologic Oncology. 1994;54(3):321-6.

114. Tan PL, Willatt JM, Lindsell D. The ability of ultrasound to detect gynaecological neoplasms and their ultrasound morphological features. Australasian Radiology. 2007;51(3):260-6.

115. Testa AC, Ajossa S, Ferrandina G, Fruscella E, Ludovisi M, Malaggese M, et al. Does quantitative analysis of three-dimensional power Doppler angiography have a role in the diagnosis of malignant pelvic solid tumors? A preliminary study. Ultrasound in Obstetrics & Gynecology. 2005;26(1):67-72.

116. Testa AC, Timmerman D, Van Belle V, Fruscella E, Van Holsbeke C, Savelli L, et al. Intravenous contrast ultrasound examination using contrast-tuned imaging (CnTI) and the contrast medium SonoVue for discrimination between benign and malignant adnexal masses with solid components. Ultrasound in Obstetrics & Gynecology. 2009;34(6):699-710.

117. Theodoridis TD, Zepiridis L, Mikos T, Grimbizis GF, Dinas K, Athanasiadis A, et al. Comparison of diagnostic accuracy of transvaginal ultrasound with laparoscopy in the management of patients with adnexal masses. Archives of Gynecology & Obstetrics. 2009;280(5):767-73.

118. Thrall MM, Paley P, Pizer E, Garcia R, Goff BA. Patterns of spread and recurrence of sex cord-stromal tumors of the ovary. Gynecologic Oncology. 2011;122(2):242-5.

119. Tingulstad S, Hagen B, Skjeldestad FE, Halvorsen T, Nustad K, Onsrud M. The risk-of-malignancy index to evaluate potential ovarian cancers in local hospitals. Obstetrics & Gynecology. 1999;93(3):448-52.

120. Tornos C, Soslow R, Chen S, Akram M, Hummer AJ, Abu-Rustum N, et al. Expression of WT1, CA 125, and GCDFP-15 as useful markers in the differential diagnosis of primary ovarian carcinomas versus metastatic breast cancer to the ovary. American Journal of Surgical Pathology. 2005;29(11):1482-9.

121. Valentin L, Ameye L, Franchi D, Guerriero S, Jurkovic D, Savelli L, et al. Risk of malignancy in unilocular cysts: a study of 1148 adnexal masses classified as unilocular cysts at transvaginal ultrasound and review of the literature. Ultrasound in Obstetrics & Gynecology. 2013;41(1):80-9.

122. Valentin L, Ameye L, Savelli L, Fruscio R, Leone FP, Czekierdowski A, et al. Unilocular adnexal cysts with papillary projections but no other solid components: is there a diagnostic method that can classify them reliably as benign or malignant before surgery? Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2013;41(5):570-81.

123. Valentin L, Ameye L, Savelli L, Fruscio R, Leone FPG, Czekierdowski A, et al. Adnexal masses difficult to classify as benign or malignant using subjective assessment of gray-scale and Doppler ultrasound findings: logistic regression models do not help. Ultrasound in Obstetrics & Gynecology. 2011;38(4):456-65.

124. Valentin L, Ameye L, Savelli L, Fruscio R, Leone FPG, Czekierdowski A, et al. Unilocular adnexal cysts with papillary projections but no other solid components: is there a diagnostic method that can classify them reliably as benign or malignant before surgery? Ultrasound in Obstetrics & Gynecology. 2013;41(5):570-81.

125. Van den Akker PAJ, Aalders AL, Snijders MPLM, Kluivers KB, Samlal RAK, Vollebergh JHA, et al. Evaluation of the Risk of Malignancy Index in daily clinical management of adnexal masses. Gynecologic Oncology. 2010;116(3):384-8.

Table 17. Excluded studies: population not applicable *(Continued)*

126. van den Akker PAJ, Zusterzeel PLM, Aalders AL, Snijders MPLM, Samlal RAK, Vollebergh JHA, et al. External validation of the adapted Risk of Malignancy Index incorporating tumor size in the preoperative evaluation of adnexal masses. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2011;159(2):422-5.
127. Van Holsbeke C, Daemen A, Yazbek J, Holland TK, Bourne T, Mesens T, et al. Ultrasound methods to distinguish between malignant and benign adnexal masses in the hands of examiners with different levels of experience. Ultrasound in Obstetrics & Gynecology. 2009;34(4):454-61.
128. van Nagell JR, Depriest PD, Reedy MB, Gallion HH, Ueland FR, Pavlik EJ, et al. The efficacy of transvaginal sonographic screening in asymptomatic women at risk for ovarian cancer. Gynecologic Oncology. 2000;77(3):350-6.
129. Vuento MH, Stenman UH, Pirhonen JP, Makinen JI, Laippala PJ, Salmi TA. Significance of a single CA 125 assay combined with ultrasound in the early detection of ovarian and endometrial cancer. Gynecologic Oncology. 1997;64(1):141-6.
130. Wang S, Johnson S. Prediction of benignity of solid adnexal masses. Archives of Gynecology & Obstetrics. 2012;285(3):721-6.
131. Wei S, Li H, Zhang B. The diagnostic value of serum HE4 and CA-125 and ROMA index in ovarian cancer. Biomedical Reports. 2016;5(1):41-4.
132. Weinberger V, Fischerova D, Semeradova I, Slama J, Dundr P, Dusek L, et al. Prospective Evaluation of Ultrasound Accuracy in the Detection of Pelvic Carcinomatosis in Patients with Ovarian Cancer. Ultrasound in Medicine & Biology. 2016;42(9):2196-202.
133. Woodward E, Sleightholme H, Considine A, Williamson S, McHugo J, Cruger D. Annual surveillance by CA125 and transvaginal ultrasound for ovarian cancer in both high-risk and population risk women is ineffective. BJOG: An International Journal of Obstetrics & Gynaecology. 2007;114(12):1500-9.
134. Yazbek J, Aslam N, Tailor A, Hillaby K, Raju KS, Jurkovic D. A comparative study of the risk of malignancy index and the ovarian crescent sign for the diagnosis of invasive ovarian cancer. Ultrasound in Obstetrics & Gynecology. 2006;28(3):320-4.
135. Yesilyurt H, Tokmak A, Guzel AI, Simsek HS, Terzioglu SG, Erkaya S, et al. Parameters for predicting granulosa cell tumor of the ovary: a single center retrospective comparative study. Asian Pacific Journal of Cancer Prevention: Apjcp. 2014;15(19):8447-50.
136. Zacharakis D, Thomakos N, Biliatis I, Rodolakis A, Simou M, Daskalakis G, et al. Ultrasonographic markers and preoperative CA-125 to distinguish between borderline ovarian tumors and stage I ovarian cancer. Acta Obstetricia et Gynecologica Scandinavica. 2013;92(3):285-92.
137. Zapardiel I, Gorostidi M, Ravaggi A, Allende MT, Silveira M, Abehsera D, et al. Utility Serum Marker HE4 for the Differential Diagnosis Between Endometriosis and Adnexal Malignancy. International Journal of Gynecological Cancer. 2016;26(1):52-5.
138. Zhang M, Zhuang G, Sun X, Shen Y, Zhao A, Di W. Risk prediction model for epithelial ovarian cancer using molecular markers and clinical characteristics. Journal of ovarian research. 2015;8:67.
139. Zygmunt A, Madry R, Markowska J, Fischer Z. Estimation of the usefulness of neoplastic markers TPS and CA 125 in diagnosis and monitoring of ovarian cancer. European Journal of Gynaecological Oncology. 1999;20(4):298-301.

Table 18. Excluded studies: publication pre-1991

Publication pre 1991
1. Chen DX, Schwartz PE, Li XG, Yang Z. Evaluation of CA 125 levels in differentiating malignant from benign tumors in patients with pelvic masses. Obstetrics and gynecology. 1988;72(1):23-7.
2. Einhorn N, Bast RC, Jr, Knapp RC, Tjernberg B, Zurawski VR, Jr. Preoperative evaluation of serum CA 125 levels in patients with primary epithelial ovarian cancer. Obstetrics and gynecology. 1986;67(3):414-6.

Table 18. Excluded studies: publication pre-1991 *(Continued)*

3. Einhorn N, Knapp RC, Bast RC, Zurawski VR. Ca 125 Assay Used in Conjunction with Ca 15–3 and Tag-72 Assays for Discrimination Between Malignant and Non-Malignant Diseases of the Ovary. Acta Oncologica. 1989;28(5):655-7.
4. FINKLER NJ, BENACERRAF B, LAVIN PT, WOJCIECHOWSKI C, KNAPP RC. Comparison of Serum CA 125, Clinical Impression, and Ultrasound in the Preoperative Evaluation of Ovarian Masses. Obstetrics & Gynecology. 1988;72(4):659-64.
5. Gadducci A, Capriello P, Bartolini T, Barale E, Cappelli N, Facchini V, et al. The association of ultrasonography and CA-125 test in the preoperative evaluation of ovarian carcinoma. European Journal of Gynaecological Oncology. 1988;9(5):373-6.
6. Hawkins RE, Roberts K, Wiltshaw E, Mundy J, McCready VR. The clinical correlates of serum CA125 in 169 patients with epithelial ovarian carcinoma. Br J Cancer. 1989;60(4):634-7.
7. Jacobs I, Oram D, Fairbanks J, Turner J, Frost C, Grudzinskas JG. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. Br J Obstet Gynaecol. 1990;97(10):922-9.
8. Malkasian GD, Jr, Knapp RC, Lavin PT, Zurawski VR, Jr, Podratz KC, Stanhope CR, et al. Preoperative evaluation of serum CA 125 levels in premenopausal and postmenopausal patients with pelvic masses: discrimination of benign from malignant disease. American journal of obstetrics and gynecology. 1988;159(2):341-6.
9. Mogensen O, Mogensen B, Jakobsen A. CA 125 in the diagnosis of pelvic masses. European Journal of Cancer and Clinical Oncology. 1989;25(8):1187-90.
10. O'Connell GJ, Ryan E, Murphy KJ, Prefontaine M. Predictive value of CA 125 for ovarian carcinoma in patients presenting with pelvic masses. Obstetrics and gynecology. 1987;70(6):930-2.
11. Patsner B, Mann WJ. The value of preoperative serum CA 125 levels in patients with a pelvic mass. American journal of obstetrics and gynecology. 1988;159(4):873-6.
12. SOPER JT, HUNTER VJ, DALY L, TANNER M, CREASMAN WT, BAST RCJ. Preoperative Serum Tumor-Associated Antigen Levels in Women With Pelvic Masses. Obstetrics & Gynecology. 1990;75(2):249-54.
13. Vasilev SA, Schlaerth JB, Campeau J, Morrow CP. Serum CA 125 levels in preoperative evaluation of pelvic masses. Obstetrics and Gynecology. 1988;71(5):751-6.
14. Yedema C, Massuger L, Hilgers J, Servaas J, Poels L, Thomas C, et al. Pre-operative discrimination between benign and malignant ovarian tumors using a combination of CA125 and CA15.3 serum assays. International journal of cancer Supplement = Journal international du cancer Supplement. 1988;3:61-7.
15. Zurawski VR, Jr, Knapp RC, Einhorn N, Kenemans P, Mortel R, Ohmi K, et al. An initial analysis of preoperative serum CA 125 levels in patients with early stage ovarian carcinoma. Gynecol Oncol. 1988;30(1):7-14.

Table 19. Excluded studies: study design

Study design
1. Aguirre A, Ardeshirpour Y, Sanders MM, Brewer M, Zhu Q. Potential role of coregistered photoacoustic and ultrasound imaging in ovarian cancer detection and characterization. Clinical & translational oncology. 2011;4(1):29-37.
2. Alongkronrusmee D, Bitterman P, Abramowicz JS, Bahr JM, Basu S, Grasso S, et al. GRP78 in association with VEGFR-2 detects early stage ovarian cancer. Cancer Research. 2013;1).
3. Alqasemi U, Kumavor P, Aguirre A, Zhu Q. Recognition algorithm for assisting ovarian cancer diagnosis from coregistered ultrasound and photoacoustic images: ex vivo study. Journal of Biomedical Optics. 2012;17(12):126003.

Table 19. Excluded studies: study design *(Continued)*

4. Anastasi E, Giovanna Marchei G, Viggiani V, Gennarini G, Frati L, Reale MG. HE4: A new potential early biomarker for the recurrence of ovarian cancer. Tumor Biology. 2010;31(2):113-9.
5. Avsar AF, Keskin HL, Catma T, Kaya B, Sivaslioglu AA. A large primary vaginal calculus in a woman with paraplegia. Journal of Lower Genital Tract Disease. 2013;17(1):61-5.
6. Balasubramaniam K, Ravn P, Larsen PV, Sondergaard J, Jarbol DE. Specific and unspecific gynecological alarm symptoms - Prevalence estimates in different age groups: A population-based study. Acta Obstetricia et Gynecologica Scandinavica. 2014;94(2):191-7.
7. Barnsfather K, Fitzpatrick CB, Wilson J, Linn CL, Brizendine E, Schilder JM. The Morphology Index: predictive value of malignancy among clinicians at various levels of training. Gynecologic Oncology. 2012;127(1):94-7.
8. Baron AT, Boardman CH, Lafky JM, Rademaker A, Liu D, Fishman DA, et al. Soluble epidermal growth factor receptor (sEGFR) [corrected] and cancer antigen 125 (CA125) as screening and diagnostic tests for epithelial ovarian cancer.[Erratum appears in Cancer Epidemiol Biomarkers Prev. 2005 Jun;14(6):1583]. Cancer Epidemiology, Biomarkers & Prevention. 2005;14(2):306-18.
9. Barrett J, Sharp DJ, Stapley S, Stabb C, Hamilton W. Pathways to the diagnosis of ovarian cancer in the UK: a cohort study in primary care. BJOG: An International Journal of Obstetrics & Gynaecology. 2010;117(5):610-4.
10. Bast RC, Jr. Early detection of ovarian cancer: new technologies in pursuit of a disease that is neither common nor rare. Transactions of the American Clinical & Climatological Association. 2004;115:233-47; discussion 47-8.
11. Begum FD, Hogdall E, Kjaer SK, Blaakaer J, Christensen IJ, Christensen L, et al. Preoperative serum tetranectin, CA125 and menopausal status used as single markers in screening and in a risk assessment index (RAI) in discriminating between benign and malignant ovarian tumors. Gynecologic Oncology. 2009;113(2):221-7.
12. Braga F, Ferraro S, Mozzi R, Panteghini M. The importance of individual biology in the clinical use of serum biomarkers for ovarian cancer. Clinical Chemistry & Laboratory Medicine. 2014;52(11):1625-31.
13. Covens AL, Dodge JE, Lacchetti C, Elit LM, Le T, Devries-Aboud M, et al. Surgical management of a suspicious adnexal mass: a systematic review. Gynecologic Oncology. 2012;126(1):149-56.
14. Dai W, Shi Y, He F, Gu C, Kong Y. Two case of androgen-secreting ovary tumor. [Chinese]. Zhongguo yi xue ke xue yuan xue bao. 1995;Acta Academiae Medicinae Sinicae. 17(4):317-20.
15. Drescher CW, Shah C, Thorpe J, O'Briant K, Anderson GL, Berg CD, et al. Longitudinal screening algorithm that incorporates change over time in CA125 levels identifies ovarian cancer earlier than a single-threshold rule. Journal of Clinical Oncology. 2013;31(3):387-92.
16. Dursun H, Albayrak F, Yildirim R, Uyanik A, Yilmaz O, Okcu N, et al. Giant mesenteric cyst can present as pseudoascites with raised Ca125. Turkish Journal of Gastroenterology. 2009;20(4):305-6.
17. Eiriksson LR, Millar HC, Lennox GK, Reade CJM, Leung F, Diamandis EP, et al. The usefulness of ovarian cancer risk scoring in the discrimination of an isolated pelvic mass. Gynecologic Oncology. 2014;133:82-3.
18. Eisenkop SM, Spirtos NM, Montag TW, Nalick RH, Wang HJ. The impact of subspecialty training on the management of advanced ovarian cancer. Gynecol Oncol. 1992;47(2):203-9.
19. Elder JW, Pavlik EJ, Long A, Miller RW, DeSimone CP, Hoff JT, et al. Serial ultrasonographic evaluation of ovarian abnormalities with a morphology index. Gynecologic Oncology. 2014;135(1):8-12.
20. Filipova I, Chakalova G. Ca 125, he 4 and roma-index in patients with epithelial ovarian cancer. International Journal of Gynecological Cancer. 2013;1):883.
21. Fritz-Rdzanek A, Grzybowski W, Beta J, Durczynski A, Jakimiuk A. HE4 protein and SMRP: Potential novel biomarkers in ovarian cancer detection. Oncology Letters. 2012;4(3):385-9.

Table 19. Excluded studies: study design *(Continued)*

22. Galgano MT, Hampton GM, Frierson Jr HF. Comprehensive analysis of HE4 expression in normal and malignant human tissues. Modern Pathology. 2006;19(6):847-53.
23. Gasiorowska E, Walkowiak G, Michalak M, Jankowska A, Nowak-Markwitz E, Spaczynski M. HE4 gene expression in ovary, fallopian tube and ovarian cancer. International Journal of Gynecological Cancer. 2013;1):480.
24. Goff B, Lowe K, Kane J, Robertson M, Gaul M, Andersen M. The safety of symptom based screening for ovarian cancer. Gynecologic Oncology. 2012;127 (1):S10.
25. Goff BA, Lowe KA, Kane JC, Robertson MD, Gaul MA, Andersen MR. Symptom triggered screening for ovarian cancer: a pilot study of feasibility and acceptability. Gynecologic Oncology. 2012;124(2):230-5.
26. Goff BA, Matthews B, Andrilla CHA, Miller JW, Trivers KF, Berry D, et al. How are symptoms of ovarian cancer managed? A study of primary care physicians. Cancer. 2011;117(19):4414-23.
27. Guadagni F, Marth C, Zeimet AG, Ferroni P, Spila A, Abbolito R, et al. Evaluation of tumor-associated glycoprotein-72 and CA 125 serum markers in patients with gynecologic diseases. American journal of obstetrics and gynecology. 1994;171(5):1183-91.
28. Hakama M, Stenman UH, Knekt P, Jarvisalo J, Hakulinen T, Maatela J, et al. CA 125 as a screening test for ovarian cancer. Journal of Medical Screening. 1996;3(1):40-2.
29. Hamilton W, Menon U. Easily missed? Ovarian cancer. British Medical Journal. 2009;339.
30. Hartge P, Hayes R, Reding D, Sherman ME, Prorok P, Schiffman M, et al. Complex ovarian cysts in postmenopausal women are not associated with ovarian cancer risk factors: preliminary data from the prostate, lung, colon, and ovarian cancer screening trial. American journal of obstetrics and gynecology [Internet]. 2000 [cited PENDING (UPDATE 91-08); (5):[1232-7 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/182/CN-00331182/frame.html.
31. Havrilesky L, Sfakianos G, Barnett J, Myers E. Comparative effectiveness of three triage strategies for women presenting to a gynecologist with a pelvic mass. Gynecologic Oncology. 2012;125:S75.
32. Hennessey A. Detection of a Unilateral Mucinous Borderline Ovarian Tumor With Sonography and Magnetic Resonance Imaging. Journal of Diagnostic Medical Sonography. 2013;29(1):30-5.
33. Hess LM, Stehman FB, Method MW, Weathers TD, Gupta P, Schilder JM. Identification of the optimal pathway to reach an accurate diagnosis in the absence of an early detection strategy for ovarian cancer. Gynecologic Oncology. 2012;127(3):564-8.
34. Holcomb K, Miller C, Vucetic Z, Knapp R. Human epididymis protein 4 increases specificity for the detection of invasiive epithelial ovarian cancer in premenopausal women presenting with an adnexal mass. Gynecologic Oncology. 2011;120:S69-S70.
35. Jacobs IJ, Skates SJ, MacDonald N, Menon U, Rosenthal AN, Davies AP, et al. Screening for ovarian cancer: a pilot randomised controlled trial. Lancet (London, England) [Internet]. 1999 [cited PENDING (UPDATE 91-08); (9160):[1207-10 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/328/CN-00162328/frame.html.
36. Kang S, Hwang J, Yoo H, Lim M, Seo S, Park S. Is serum HE4 measurement useful in early detection of ovarian cancer? Gynecologic Oncology. 2012;125:S100.
37. Kenemans P, Verstraeten AA, Van Kamp GJ, Von Mensdorff-Pouilly S. The second generation CA 125 assays. Annals of Medicine. 1995;27(1):107-13.
38. Kim J-A, Chun YK, Moon MH, Lee YH, Cho HC, Lee MS, et al. High-resolution sonographic findings of ovarian granulosa cell tumors: correlation with pathologic findings. Journal of Ultrasound in Medicine. 2010;29(2):187-93.
39. Kim KH, Zsebik GN, Straughn JM, Landen CN. Management of complex pelvic masses using a multivariate index assay: a decision analysis (Provisional abstract). Gynecologic Oncology [Internet]. 2012 [cited EED Y/U]; (3):[364-8 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/cleed/articles/NHSEED-22012030548/frame.html.

Table 19. Excluded studies: study design *(Continued)*

40. Kommoss S, Anglesio MS, Mackenzie R, Yang W, Senz J, Ho J, et al. FOXL2 molecular testing in ovarian neoplasms: diagnostic approach and procedural guidelines. Modern Pathology. 2013;26(6):860-7.

41. Kong F, Nicole White C, Xiao X, Feng Y, Xu C, He D, et al. Using proteomic approaches to identify new biomarkers for detection and monitoring of ovarian cancer. Gynecologic Oncology. 2006;100(2):247-53.

42. Kuhlmann JD, Schwarzenbach H, Otterbach F, Heubner M, Wimberger P, Worm KH, et al. Primary tumor loss of heterozygosity proximal to M6P/IGF2R locus is predictive for the presence and persistence of disseminated tumor cells in the bone marrow of ovarian cancer patients. Cancer Research. 2011;1).

43. Kupets R, Fernandes K, Miroshnichenko G, Paszat L. Are too many imaging tests being performed in women with an adnexal mass? Journal of Obstetrics & Gynaecology Canada: JOGC. 2013;35(3):246-51.

44. Lataifeh I, Marsden DE, Robertson G, Gebski V, Hacker NF. Presenting symptoms of epithelial ovarian cancer. The Australian & New Zealand journal of obstetrics & gynaecology. 2005;45(3):211-4.

45. Latifi A, Abubaker K, Castrechini N, Ward AC, Liongue C, Dobill F, et al. Cisplatin treatment of primary and metastatic epithelial ovarian carcinomas generates residual cells with mesenchymal stem cell-like profile. Journal of Cellular Biochemistry. 2011;112(10):2850-64.

46. Lazebnik N, Balog A, Bennett S, Redline R, Liu J. Ovarian dysgerminoma: a challenging clinical and sonographic diagnosis. Journal of Ultrasound in Medicine. 2009;28(10):1409-15.

47. Lu KH, Skates S, Hernandez MA, Bedi D, Bevers T, Leeds L, et al. A 2-stage ovarian cancer screening strategy using the Risk of Ovarian Cancer Algorithm (ROCA) identifies early-stage incident cancers and demonstrates high positive predictive value. Cancer. 2013;119(19):3454-61.

48. Meinhold-Heerlein I, Fehm T, Haltmeier C, Brautigam K, Neubauer H, Maass N, et al. The potential of novel molecular markers to serve as therapeutical targets or serum tumor markers of serous ovarian cancer. Onkologie. 2010;33:114-.

49. Menon U. Sensitivity and Specificity of Multimodal and Ultrasound Screening for Ovarian Cancer, and Stage Distribution of Detected Cancers: Results of the Prevalence Screen of the United Kingdom Collaborative Trial of Ovarian Cancer Screening COMMENT. Obstetrical & Gynecological Survey. 2009;64(9):592-5.

50. Mohan S, Kapoor G, Nagpal PK, Aggarwal R, Gami N. Managing adnexal masses: a medical quandary. Journal of Clinical and Diagnostic Research JCDR. 2013;7(9):1971-4.

51. Molina R, Ojeda B, Filella X, Borras G, Jo J, Mas E, et al. A prospective study of tumor markers CA 125 and CA 19.9 in patients with epithelial ovarian carcinomas. Tumor Biology. 1992;13(5-6):278-86.

52. Montagnana M, Danese E, Ruzzenente O, Bresciani V, Nuzzo T, Gelati M, et al. The ROMA (Risk of Ovarian Malignancy Algorithm) for estimating the risk of epithelial ovarian cancer in women presenting with pelvic mass: is it really useful? Clinical Chemistry & Laboratory Medicine. 2011;49(3):521-5.

53. Moore R, Miller C, DiSilvestro P, Landrum L, Gajewski W, Renneisen P, et al. Evaluation of the risk of ovarian malignancy algorithm in women with a pelvic mass presenting to general gynecologists. Gynecologic Oncology. 2011;121(1):S68-S.

54. Mozina B. Usefulness of ROMA (Risk of Ovarian Malignancy Algorithm) based on tumor markers CA125 and HE4. Tumor Biology. 2014;35:S37.

55. Murphy P, Adams M, Martin B, Horlick E, Malinowski D. Correlation of mRNA and protein expression of CA125 and HE4 in ovarian cancer tissues using real-time RT-PCR and reverse phase protein array technologies. Cancer Research. 2010;1).

56. Nishimura S, Tsuda H, Ito K, Takano M, Terai Y, Jobo T, et al. Differential expression of hypoxia-inducible protein 2 among different histological types of epithelial ovarian cancer and in clear cell adenocarcinomas. International Journal of Gynecological Cancer. 2010;20(2):220-6.

Table 19. Excluded studies: study design *(Continued)*

57. Orre M, Lotfi-Miri M, Mamers P, Rogers PA. Increased microvessel density in mucinous compared with malignant serous and benign tumours of the ovary. British Journal of Cancer. 1998;77(12):2204-9.

58. Ota T, Clayton AC, Minot DM, Hartmann LC, Shridhar V, Gilks B, et al. Mini-chromosome maintenance protein 7 as a potential theranostic biomarker in epithelial ovarian cancer. Cancer Research. 2010;1).

59. Pascual MA, Graupera B, Hereter L, Rotili A, Rodriguez I, Alcazar JL. Intra- and interobserver variability of 2D and 3D transvaginal sonography in the diagnosis of benign versus malignant adnexal masses. Journal of Clinical Ultrasound. 2011;39(6):316-21.

60. Peng D, Xu T, Mason TJ, Wu W. A study of ovarian cancer biomarker amplification using ultrasound for early stage detection. Ultrasonics. 2014;54(2):451-4.

61. Perkins G, Dukes J, Pope L, Clark J, Yap TA, Riisnaes R, et al. Prospective study of genetic mutations in matched tumor and plasma specimens in advanced cancer patients referred to phase I trials. Cancer Research. 2011;1).

62. Phocas I, Sarandakou A, Sikiotis K, Rizos D, Kalambokis D, Zourlas PA. A comparative study of serum alpha-beta A immunoreactive inhibin and tumor-associated antigens CA125 and CEA in ovarian cancer. Anticancer Research [Internet]. 1996 [cited PENDING (UPDATE 91-08); (6b):[3827-31 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/922/CN-00136922/frame.html.

63. Pinsky PF, Zhu CS. Building multi-marker algorithms for disease prediction-the role of correlations among markers. Biomark Insights. 2011;6:83-93.

64. Poncelet C, Fauvet R, Yazbeck C, Coutant C, Darai E. Impact of serum tumor marker determination on the management of women with borderline ovarian tumors: multivariate analysis of a French multicentre study. European Journal of Surgical Oncology. 2010;36(11):1066-72.

65. Rana R, Padwick M. Accuracy of imaging in adnexal masses: An audit. BJOG: An International Journal of Obstetrics and Gynaecology. 2013;120:375.

66. Ricardo S, Marcos-Silva L, Pereira D, Pinto R, Almeida R, Soderberg O, et al. Detection of glyco-mucin profiles improves specificity of MUC16 and MUC1 biomarkers in ovarian serous tumours. Molecular Oncology. 2015;9(2):503-12.

67. Ricci F, Bernasconi S, Erba E, Mangioni C, Fruscio R, Broggini M, et al. Characterization of markers associated with tumorigenicity in ovarian cancer tumors. Cytometry Part A. 2010;77A(2):200-.

68. Rohilla M, Chopra S, Aggrawal N, Suri V, Rajvanshi A, Acharya G. Diagnostic dilemma of adnexal mass in a postmenopausal woman. Journal of Obstetrics & Gynaecology. 2012;32(3):315-6.

69. Rufford B, Jacobs I, Menon U. Feasibility of screening for ovarian cancer using symptoms as selection criteria. BJOG: An International Journal of Obstetrics & Gynaecology. 2007;114(1):59-64.

70. Rychlik U, Kulpa JK, Wojcik E, Tarapacz J, Stasik Z. CA 125, HE4 in relation to inflammation in ovarian cancer patients. Tumor Biology. 2011;32:S76.

71. Sarkar M, Konar H, Raut D. Symptomatology of gynecological malignancies: experiences in the gynecology out-patient clinic of a tertiary care hospital in kolkata, India. Asian Pacific Journal of Cancer Prevention: Apjcp. 2010;11(3):785-91.

72. Sarkar M, Konar H, Raut D. Clinico-pathological features of gynecological malignancies in a tertiary care hospital in eastern India: importance of strengthening primary health care in prevention and early detection. Asian Pacific Journal of Cancer Prevention: Apjcp. 2013;14(6):3541-7.

73. Scholler N, Lowe KA, Bergan LA, Kampani A, Ng V, Forrest RM, et al. Use of yeast-secreted in vivo biotinylated recombinant antibodies (biobodies) in bead-based ELISA. Clinical Cancer Research. 2008;14(9):2647-55.

Table 19. Excluded studies: study design *(Continued)*

74. Skates SJ, Horick N, Yu Y, Xu F, Berchuck A, Havrilesky LJ, et al. Preoperative sensitivity and specificity for early-stage ovarian cancer when combining cancer antigen CA-125II, CA 15-3, CA 72-4, and macrophage colony-stimulating factor using mixtures of multivariate normal distributions. Journal of Clinical Oncology. 2004;22(20):4059-66.

75. Skates SJ, Menon U, MacDonald N, Rosenthal AN, Oram DH, Knapp RC, et al. Calculation of the risk of ovarian cancer from serial CA-125 values for preclinical detection in postmenopausal women. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2003;21(10 Suppl):206s-10s.

76. Sladkevicius P, Valentin L. Interobserver agreement in the results of Doppler examinations of extrauterine pelvic tumors. Ultrasound in Obstetrics & Gynecology. 1995;6(2):91-6.

77. Sladkevicius P, Valentin L. Intra- and interobserver agreement when describing adnexal masses using the International Ovarian Tumor Analysis terms and definitions: a study on three-dimensional ultrasound volumes. Ultrasound in Obstetrics & Gynecology. 2013;41(3):318-27.

78. Stalbovskaya V, Ifeachor EC, Van Huffel S, Timmerman D. Preoperative prediction of malignancy of ovarian tumours using modified sequential non-uniform procedure. Conference proceedings: 2007;Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference. 2007:5403-6.

79. Szkandera J, Ploner F, Bauernhofer T, Kasparek A-K, Payer F, Balic M, et al. Paraneoplastic limbic encephalitis in a patient with extragonadal choriocarcinoma--significance of onconeural antibodies. Onkologie. 2010;33(8-9):452-4.

80. Toftager-Larsen K, Hording U, Dreisler A, Daugaard S, Lund B, Bock J, et al. CA-125, placental alkaline phosphatase and tissue polypeptide antigen as preoperative serum markers in ovarian carcinoma. Gynecologic and Obstetric Investigation. 1992;33(3):177-82.

81. Ueland FR, Elder JW, Long A, Desimone CP, Miller RW, Podzielinski I, et al. Serial tumor morphology indexing predicts risk of ovarian malignancy. International Journal of Gynecological Cancer. 2012;22:E107.

82. Van Calster B, Van Belle V, Vergouwe Y, Timmerman D, Van Huffel S, Steyerberg EW. Extending the c-statistic to nominal polytomous outcomes: the Polytomous Discrimination Index. Statistics in Medicine. 2012;31(23):2610-26.

83. Van Calster B, Vergouwe Y, Looman CWN, Van Belle V, Timmerman D, Steyerberg EW. Assessing the discriminative ability of risk models for more than two outcome categories. European Journal of Epidemiology. 2012;27(10):761-70.

84. Vergote IB, Abeler VM, Bormer OP, Stigbrand T, Trope C, Nustad K. CA125 and placental alkaline phosphatase as serum tumor markers in epithelial ovarian carcinoma. Tumor Biology. 1992;13(3):168-74.

85. Visintin I, Feng Z, Longton G, Ward DC, Alvero AB, Lai Y, et al. Diagnostic markers for early detection of ovarian cancer. Clinical Cancer Research. 2008;14(4):1065-72.

86. Wang BL, Shu-Zheng; Zheng, Rong-Shou; Zhang, Fang; Chen, Wan-Qing; Sun, Xi-Bin. Time Trends of Ovarian Cancer Incidence in China. Asian Pacific Journal of Cancer Prevention. 2014;15(1):191-3.

87. Wang J, Lv F, Fei X, Cui Q, Wang L, Gao X, et al. Study on the characteristics of contrast-enhanced ultrasound and its utility in assessing the microvessel density in ovarian tumors or tumor-like lesions. International Journal of Biological Sciences [Electronic Resource]. 2011;7(5):600-6.

88. Webb PM, Purdie DM, Grover S, Jordan S, Dick ML, Green AC. Symptoms and diagnosis of borderline, early and advanced epithelial ovarian cancer. Gynecol Oncol. 2004;92(1):232-9.

89. Wojcik-Krowiranda K, Litwinska M, Bienkiewicz A. Usefulness of determination of CA125 concentration as a marker for ovarian tumours with special consideration of ovarian endometriotic cysts. [Polish] Przydatnosc oceny stezenia markera CA125 w diagnostyce guzow jajnika ze szczegolnym uwzglednieniem torbieli endometrialnych. Przeglad Menopauzalny. 2010;14(6):362-5.

90. Woolas RP, Oram DH, Jeyarajah AR, Bast Jr RC, Jacobs IJ. Ovarian cancer identified through screening with serum markers but not by pelvic imaging. International Journal of Gynecological Cancer. 1999;9(6):497-501.

Table 19. Excluded studies: study design *(Continued)*

91. Zalel Y, Tepper R, Altaras M, Beyth Y. Transvaginal sonographic measurements of postmenopausal ovarian volume as a possible detection of ovarian neoplasia. Acta Obstetricia et Gynecologica Scandinavica. 1996;75(7):668-71.
92. Zannoni L, Savelli L, Jokubkiene L, Di Legge A, Condous G, Testa AC, et al. Intra- and interobserver agreement with regard to describing adnexal masses using International Ovarian Tumor Analysis terminology: reproducibility study involving seven observers. Ultrasound in Obstetrics & Gynecology. 2014;44(1):100-8.
93. Zhu Y-F, Gao G-L, Zhang Z-D, Huang Q-S. Expression of human WFDC2 protein from patients with serous carcinomas by the two-tier system. Chinese Medical Journal. 2013;126(18):3576-7.
94. Zhu Y-F, He L-S, Zhang Z-D, Huang Q-S. Expression of serum human epididymal secretory protein E4 at low grade and high grade serous carcinomas. Asian Pacific Journal of Tropical Medicine. 2012;5(12):925-30.
95. Zouboulis CC, Achenbach A, Makrantonaki E. Acne tarda and male-pattern baldness unmasking primary ovarian insufficiency: a case and review. Dermatology. 2014;229(2):51-4.
Correspondence/opinions/reviews
1. Enhanced ovarian cancer detection. Two-test combo may identify disease early enough to cure it. Duke Medicine Health News. 2009;15(6):8-9.
2. Combining CA-125 and U/S best approach to ovarian Ca screening. Contemporary OB/GYN. 2009;54(4):22-.
3. Markers for ovarian cancer may show up years earlier. Healthcare Traveler. 2010;17(8):6.
4. Biomarkers for ovarian cancer. British Journal of Hospital Medicine (17508460). 2010;71(1):8-.
5. OVA1 test for the assessment of suspected ovarian cancer (Structured abstract). Health Technology Assessment Database [Internet]. 2011 [cited MEDLINE N]; (1). Available from: http://onlinelibrary.wiley.com/o/cochrane/clhta/articles/HTA-32011001101/frame.html.
6. New guidelines for early detection of ovarian cancer. The Lancet. 2011;377(9777):1544.
7. Researchers discover antibody that may help detect ovarian cancer in its earliest stages. NEWS-Line for Laboratory Professionals. 2011;6(9-10):4-5.
8. Ovarian cancer screening strategy may allow earlier detection. British Journal of Hospital Medicine (17508460). 2013;74(9):490-.
9. Abraham J, Ueland FR. OVA1 test for preoperative assessment of ovarian cancer. Community Oncology. 2010;7(6):249-50.
10. Acharya UR, Sree SV, Saba L, Molinari F, Guerriero S, Suri JS. Ovarian tumor characterization and classification using ultrasound-a new online paradigm. Journal of Digital Imaging. 2013;26(3):544-53.
11. Alcazar JL. Re: Intra- and interobserver reproducibility of assessment of Doppler ultrasound findings in adnexal masses. L. Zannoni, L. Savelli, L. Jokubkiene, A. Di Legge, G. Condous, A. C. Testa, P. Sladkevicius and L. Valentin. Ultrasound Obstet Gynecol 2013; 42: 93-101. Ultrasound in Obstetrics & Gynecology. 2013;42(1):4-5.
12. Amarin ZO, Jamal H. Serum CA 125 and fallopian tube carcinoma [1]. Annals of Saudi Medicine. 2001;21(1-2):130.
13. Azvolinsky A. Diagnostic lens turns to difficult-to-detect ovarian cancer. Nature Medicine. 2013;19(2):117.
14. Baron AT, Maihle N. Nadir CA125 concentration as a prognostic indicator in ovarian cancer. Nature Clinical Practice Oncology. 2005;2(6):288-9.
15. Bast Jr RC, Brewer M, Zou C, Hernandez MA, Daley M, Ozols R, et al. Prevention and early detection of ovarian cancer: mission impossible? Recent Results in Cancer Research. 2007;Fortschritte der Krebsforschung. Progres dans les recherches sur le cancer. 174:91-100.

Table 19. Excluded studies: study design *(Continued)*

16. Bast RC, Badgwell D, Lu Z, Marquez R, Rosen D, Liu J, et al. New tumor markers: CA125 and beyond. International Journal of Gynecological Cancer. 2005;15:274-81.
17. Berek JS, Bast Jr RC. Ovarian cancer screening: The use of serial complementary tumor markers to improve sensitivity and specificity for early detection. Cancer. 1995;76(10 SUPPL.):2092-6.
18. Campbell S. Ovarian cancer: role of ultrasound in preoperative diagnosis and population screening. Ultrasound in Obstetrics & Gynecology. 2012;40(3):245-54.
19. Candido-dos-Reis FJ. Ovarian Cancer risk model needs more meaningful clinical parameter... in response to VanCalster B, Van Hoorde K, Valentin L, Testa AC, Fischerova D, Van Holsbekeet C, et al: International Ovarian Tumor Analysis (IOTA) Group. Evaluating the risk of ovarian cancer before surgery using the ADNEX moded to differentiate between benign, borderline, early and advanced state invasive, and secondary metastatic tumours: prospective multicentre diagnostic study. BMJ 2014;349:g5920. (15 October.). BMJ: British Medical Journal. 2014;349(7894):g6689-2.
20. Cass I. The Search for Meaning-Symptoms and Transvaginal Sonography Screening for Ovarian Cancer Silent No More. Cancer. 2009;115(16):3606-8.
21. Chan KK, Selman TJ. Testing for ovarian cancer. Best Practice & Research: Clinical Obstetrics & Gynaecology. 2006;20(6):977.
22. Cohen CJ, Jennings TS. Screening for ovarian cancer: the role of noninvasive imaging techniques. American Journal of Obstetrics & Gynecology. 1994;170(4):1088-94.
23. Crum CP. Screening of symptomatic women for ovarian cancer. Lancet Oncology. 2012;13(4):e137; author reply e9-40.
24. Finkler NJ. Clinical utility of CA 125 in preoperative diagnosis of patients with pelvic masses. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 1993;49(1-2):105-7.
25. Gentry-Maharaj A, Kalsi J, Burnell M, Manchanda R, Menon U. Screening of symptomatic women for ovarian cancer. Lancet Oncology. 2012;13(4):e138-9; author reply e9-40.
26. Hamilton W. Computer assisted diagnosis of ovarian cancer in primary care. BMJ (Online). 2012;344(7841).
27. Hartge P. Designing early detection programs for ovarian cancer. Journal of the National Cancer Institute. 2010;102(1):3-4.
28. Hartge P. Reducing ovarian cancer death rates through screening. Cancer (0008543X). 2011;117(3):449-50.
29. Hartge P, Speyer JL. Finding ovarian cancer. JNCI: Journal of the National Cancer Institute. 2012;104(2):82-3.
30. Hellstrom I, Hellstrom KE. SMRP and HE4 as biomarkers for ovarian carcinoma when used alone and in combination with CA125 and/or each other. Ovarian Cancer: State of the Art and Future Directions in Translational Research. 2008;Advances in Experimental Medicine and Biology. 622:15-21.
31. Herzog TJ. Enhancing the needle count in the haystack: serial ultrasonography for low-to-moderate risk adnexal masses. Obstetrics & Gynecology. 2013;122(2 Pt 1):198-200.
32. Javitt MC. Risk assessment for ovarian carcinoma: hope or hype? AJR American Journal of Roentgenology. 2010;194(2):308.
33. Jones HW, III. [Commentary on] Does three-dimensional power Doppler ultrasound help in discrimination between benign and malignant ovarian masses? Obstetrical & Gynecological Survey. 2007;62(5):308-9.
34. Jordan SJ, Coory MD, Webb PM. Re: predictive value of symptoms for early detection of ovarian cancer. JNCI: Journal of the National Cancer Institute. 2010;102(20):1599-601.
35. Jurkovic D. Re: Intra- and interobserver agreement when describing adnexal masses using the International Ovarian Tumor Analysis terms and definitions: a study on three-dimensional ultrasound volumes. P. Sladkevicius and L. Valentin. Ultrasound Obstet Gynecol 2013; 41: 318-327. Ultrasound in Obstetrics & Gynecology. 2013;41(3):245-6.

Table 19. Excluded studies: study design *(Continued)*

36. Jurkovic D, Alfirevic Z. DISQ 4: Serum CA 125 and benign looking adnexal masses. Ultrasound in Obstetrics and Gynecology. 2006;27(4):466.
37. Kaijser J, Sayasneh A, Van Calster B, Timmerman D, Bourne T. Preoperative identification of a suspicious adnexal mass. Gynecologic Oncology. 2012;127(1):260-2.
38. Kaijser J, Van Gorp T, Sayasneh A, Vergote I, Bourne T, Van Calster B, et al. Differentiating stage I epithelial ovarian cancer from benign disease in women with adnexal tumors using biomarkers or the ROMA algorithm. Gynecologic Oncology. 2013;130(2):398-9.
39. Karlan BY, McIntosh M. The quest for ovarian cancer's holy grail: Can CA-125 still be the chalice of early detection? Journal of Clinical Oncology. 2007;25(11):1303-4.
40. Kenemans P, Yedema CA, Bon GG, Von Mensdorff-Pouilly S. CA 125 in gynecological pathology - A review. European Journal of Obstetrics Gynecology and Reproductive Biology. 1993;49(1-2):115-24.
41. Kim YH, Kim SC. Recent advances in the biomarkers for epithelial ovarian cancer. Journal of Gynecologic Oncology. 2011;22(4):219-21.
42. Kim YT. A hospital-based case-control study of identifying ovarian cancer using symptom index. Journal of Gynecologic Oncology. 2010;21(1):65.
43. Ko HS, Kim N, Park YG. Re: interobserver agreement in describing adnexal masses using the International Ovarian Tumor Analysis simple rules in a real-time setting and using three-dimensional ultrasound volumes and digital clips. Ultrasound in Obstetrics & Gynecology. 2015;45(2):238.
44. Kondalsamy-Chennakesavan S, Obermair A. Differentiating stage I epithelial ovarian cancer from benign disease in women with adnexal tumours using biomarkers or the ROMA algorithm. Gynecologic Oncology. 2013;130(2):400.
45. Langmar Z, Nemeth M, Szekely B, Borgulya G. The performance of the risk of ovarian malignancy algorithm. British Journal of Cancer. 2011;105(1):185-6; author reply 7-8.
46. Leahy Y. Are serum protein biomarkers effective in detecting ovarian cancer in its early stages? Clinical Journal of Oncology Nursing. 2009;13(4):443-5.
47. Mano A, Godinho I, Falcao AC. CA 125 half-life breakpoint between a "good" and "poor" prognosis in patients with ovarian cancer. International Journal of Gynecology and Obstetrics. 2005;88(3):333-5.
48. Markman M. Limitations to the use of the CA-125 antigen level in ovarian cancer. Current Oncology Reports. 2003;5(4):263-4.
49. Markowska J, Manys G, Kubaszewska M. Value of CA 125 as a marker of ovarian cancer. European Journal of Gynaecological Oncology. 1992;13(4):360-5.
50. McGuire WP. CA-125 and early ovarian cancer: does this help the clinician or further muddy the water? Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2005;23(25):5862-4.
51. Melichar B. Human epididymis protein 4: the start of a post-ROMAn era? Clinical Chemistry and Laboratory Medicine. 2012;50(12):2069-73.
52. Mourits MJ, de Bock GH. Symptoms are not early signs of ovarian cancer. BMJ: British Medical Journal (Overseas & Retired Doctors Edition). 2009;339:b3955-b.
53. Murta EFC. Tumour markers, ultrasonography, and ovarian cancer diagnosis. European Journal of Gynaecological Oncology. 2014;35(2):111.
54. Nossov V, Amneus M, Su F, Lang J, Janco JM, Reddy ST, et al. The early detection of ovarian cancer: from traditional methods to proteomics. Can we really do better than serum CA-125? American Journal of Obstetrics & Gynecology. 2008;199(3):215-23.

Table 19. Excluded studies: study design *(Continued)*

55. Olaitan A. NICE on ovarian cancer. Recommendations for detection in primary care are flawed. BMJ. 2011;342:d3022.
56. Olovsson M. Immunological aspects of endometriosis. Acta Obstetricia et Gynecologica Scandinavica. 2012;91:24-5.
57. Pedersen A. Higher levels of HE4 may be sign of ovarian cancer, study reveals. Diagnostics & Imaging Week. 2007;10(51):3-.
58. Plebani M, Melichar B. ROMA or death: advances in epithelial ovarian cancer diagnosis. Clinical Chemistry & Laboratory Medicine. 2011;49(3):443-5.
59. Rasool R, Shah ZA, Salahuddin M, Bashir M. CA-125 tumour marker for ovarian malignancies. JK Practitioner. 2003;10(3):224-5.
60. Rosenthal AN, Jacobs IJ. The role of CA 125 in screening for ovarian cancer. International Journal of Biological Markers [Internet]. 1998 [cited PENDING (UPDATE 91-08); (4):[216-20 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/182/CN-00487182/frame.html.
61. Sanchez Munoz A, Gonzalez Martin A, Mendiola Fernandez C. Lights and shadows of the tumoral marker CA-125 in ovarian cancer. Clinical & translational oncology: official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico. 2008;10(8):449-52.
62. Sundar S, O'Byrne KJ. CA-125 criteria for response evaluation in ovarian cancer. Gynecologic Oncology. 2005;98(3):520-1.
63. Timmerman D, Van Calster B, Vergote I, Van Hoorde K, Van Gorp T, Valentin L, et al. Performance of the American College of Obstetricians and Gynecologists' ovarian tumor referral guidelines with a multivariate index assay. Obstetrics & Gynecology. 2011;118(5):1179-81; author reply 81.
64. Timmerman D, Verrelst H, Collins WP, Bourne TH, Vergote I. Re: Mol et al. Distinguishing the benign and malignant adnexal mass: an external validation of prognostic models. Gynecol Oncol 2001;80:162-7. Gynecol Oncol. 2001;83(1):166-8.
65. Tone AA, Huntsman DG, Miller DM. Screening of symptomatic women for ovarian cancer. Lancet Oncology. 2012;13(4):e137-8; author reply e9-40.
66. Tummala MK, McGuire WP. Ovarian cancer, CA-125 addiction, and informed confusion: Much ado about less. Journal of Clinical Oncology. 2007;25(24):3570-1.
67. Urban N, McIntosh MW, Andersen M, Karlan BY. Ovarian cancer screening. Hematology - Oncology Clinics of North America. 2003;17(4):989-1005, ix.
68. Van Calster B, Wynants L, Kaijser J, Van Huffel S, Timmerman D. The independent effect of tumor size in predicting ovarian malignancy. European Journal of Obstetrics Gynecology and Reproductive Biology. 2012;162(2):237-8.
69. Wacholder S, Han SS, Weinberg CR. Inference from a multiplicative model of joint genetic effects or ovarian cancer risk. JNCI: Journal of the National Cancer Institute. 2011;103(2):82-3.
70. Whitehouse C, Solomon E. Current status of the molecular characterization of the ovarian cancer antigen CA125 and implications for its use in clinical screening. Gynecologic Oncology. 2003;88(1 Pt 2):S152-7.
71. Wilson B, Nielsen HO. The carcinoid syndrome caused by retroperitoneal carcinoid tumors. Neuroendocrinology. 2011;94:49-50.
72. Wiwanitkit V. HE4, CA-125, and cystic ovarian mass. Journal of Gynecologic Oncology. 2012;23(2):133.
Developmental model superseded by model validation (N=4)
1. Hillaby K, Aslam N, Salim R, Lawrence A, Raju KS, Jurkovic D. The value of detection of normal ovarian tissue (the 'ovarian crescent sign') in the differential diagnosis of adnexal masses. Ultrasound in Obstetrics & Gynecology. 2004;23(1):63-7.

Table 19. Excluded studies: study design *(Continued)*

2. Timmerman D, Bourne TH, Tailor A, Collins WP, Verrelst H, Vandenberghe K, et al. A comparison of methods for preoperative discrimination between malignant and benign adnexal masses: the development of a new logistic regression model. American journal of obstetrics and gynecology. 1999;181(1):57-65.
3. Timmerman D, Testa AC, Bourne T, Ameye L, Jurkovic D, Van Holsbeke C, et al. Simple ultrasound-based rules for the diagnosis of ovarian cancer. Ultrasound in Obstetrics & Gynecology. 2008;31(6):681-90.
4. Timmerman D, Testa AC, Bourne T, Ferrazzi E, Ameye L, Konstantinovic ML, et al. Logistic regression model to distinguish between the benign and malignant adnexal mass
Case control study with healthy controls
1. Abdalla N, Winiarek J, Bachanek M, Cendrowski K, Sawicki W. Clinical, ultrasound parameters and tumor marker-based mathematical models and scoring systems in pre-surgical diagnosis of adnexal tumors. Ginekologia polska. 2016;87(12):824-9.
2. Abdel-Hay NH, Mansour AA, El Sherbini MA, Motaleb FIA, El-Shalakany AH. Forkhead box protein 3 (FoxP3) mRNA as a diagnostic marker in ovarian tumors: a pilot study. European Journal of Gynaecological Oncology. 2017;38(2):251-6.
3. Abramowicz JS, Timmerman D. Ovarian mass-differentiating benign from malignant: the value of the International Ovarian Tumor Analysis ultrasound rules. American Journal of Obstetrics & Gynecology. 2017;217(6):652-60.
4. Alcazar JL. Ultrasound-based IOTA simple rules allow accurate malignancy risk estimation for adnexal masses. Evidence Based Medicine. 2016;21(5):197.
5. Anastasi E, Capoccia D, Granato T, Viggiani V, Tartaglione S, Manganaro L, et al. Assessing the association between 25-OH vitamin D levels and ROMA score in a population of obese women. Journal of Biological Regulators & Homeostatic Agents. 2016;30(4):1165-71.
6. Aydin DS, Turkyilmaz E, Goksedef BP, Aydin S, Aslan O, Ateser G, et al. 1,138 women with adnexal mass: pathologic findings according to age. European Journal of Gynaecological Oncology. 2017;38(1):102-5.
7. Anastasi E, Granato T, Falzarano R, Storelli P, Ticino A, Frati L, et al. The use of HE4, CA125 and CA72-4 biomarkers for differential diagnosis between ovarian endometrioma and epithelial ovarian cancer. Journal of ovarian research. 2013;6(1):44.
8. Ankita S, Shrivastav BR, Kumar SA. Clinical utility of CA125, HE4 and risk of ovarian malignancy algorithm (ROMA) for early detection of ovarian carcinoma. International Research Journal of Pharmacy. 2017;8(6):39-45.
9. Bendifallah S, Body G, Darai E, Ouldamer L. Diagnostic and prognostic value of tumor markers, scores (clinical and biological) algorithms, in front of an ovarian mass suspected of an epithelial ovarian cancer: Article drafted from the French Guidelines in oncology entitled "Initial management of patients with epithelial ovarian cancer" developed by FRANCOGYN, CNGOF, SFOG, GINECO-ARCAGY under the aegis of CNGOF and endorsed by INCa. Gynecologie Obstetrique Fertilite et Senologie. 2019;47(2):134-54.
10. Bian J, Li B, Kou X-J, Liu T-Z, Ming L. Clinical significance of combined detection of serum tumor markers in diagnosis of patients with ovarian cancer. Asian Pacific Journal of Cancer Prevention: Apjcp. 2013;14(11):6241-3.
11. Broder JC, Jimenez JJ, Flye CW. R-SCAN: Follow-Up for Adnexal Cysts. Journal of the American College of Radiology. 2017;14(7):944-6.
12. Bromley B. Pelvic ultrasound: a powerful tool in managing the patient with an adnexal mass. Menopause. 2017;24(6):599-601.
13. Carvalho VP, Grassi ML, Palma CS, Carrara HHA, Faca VM, Candido Dos Reis FJ, et al. The contribution and perspectives of proteomics to uncover ovarian cancer tumor markers. Translational Research: The Journal Of Laboratory & Clinical Medicine. 2019;206:71-90.
14. Choi JI, Park SB, Han BH, Kim YH, Lee YH, Park HJ, et al. Imaging features of complex solid and multicystic ovarian lesions: proposed algorithm for differential diagnosis. Clinical Imaging. 2016;40(1):46-56

Table 19. Excluded studies: study design *(Continued)*

15. Cui R, Wang Y, Li Y. Clinical value of ROMA index in diagnosis of ovarian cancer: Meta-analysis. Cancer Management and Research. 2019;11:2545-51.
16. Chudecka-Glaz AM. ROMA, an algorithm for ovarian cancer. Clinica Chimica Acta. 2015;440:143-51.
17. Dayyani F, Uhlig S, Colson B, Simon K, Rolny V, Morgenstern D, et al. Diagnostic Performance of Risk of Ovarian Malignancy Algorithm Against CA125 and HE4 in Connection With Ovarian Cancer: A Meta-analysis. International Journal of Gynecological Cancer. 2016;26(9):1586-93.
18. Dochez V, Caillon H, Vaucel E, Dimet J, Winer N, Ducarme G. Biomarkers and algorithms for diagnosis of ovarian cancer: CA125, HE4, RMI and ROMA, a review. Journal of Ovarian Research. 2019;12 (1) (no pagination)(28).
19. El Bairi K, Kandhro AH, Gouri A, Mahfoud W, Louanjli N, Saadani B, et al. Emerging diagnostic, prognostic and therapeutic biomarkers for ovarian cancer. Cellular Oncology. 2017;40(2):105-18.
20. Fakhar HB, Rezaie-Tavirani M, Zali H, Faraji Darkhaneh R, Nejad leili EK, Aghazadeh MH. Comparison of Serum Human Epididymis Protein (HE4), Carbohydrate Antigen 125(CA125) and Risk of Ovarian Malignancy Algorithm (ROMA) as Markers in Ovarian Cancer: A Systematic Review and a Meta-analysis. Indian Journal of Gynecologic Oncology. 2018;16 (1) (no pagination)(10).
21. Falzarano R, Viggiani V, Michienzi S, Colaprisca B, Longo F, Frati L, et al. CLEIA CA125 evidences: good analytical performance avoiding "Hook effect". Tumour Biology. 2013;34(1):387-93.
22. Fawzy A, Mohamed MR, Ali MA, Abd El-Magied MH, Helal AM. Tissue CA125 and HE4 Gene Expression Levels Offer Superior Accuracy in Discriminating Benign from Malignant Pelvic Masses. Asian Pacific Journal of Cancer Prevention: Apjcp. 2016;17(1):323-33.
23. Friedman GD, Skilling JS, Udaltsova NV, Smith LH. Early symptoms of ovarian cancer: a case-control study without recall bias. Family practice. 2005;22(5):548-53.
24. Froyman W, Landolfo C, Amant F, Van den Bosch T, Vergote I, Coosemans A, et al. Morcellation and risk of malignancy in presumed ovarian fibromas/fibrothecomas. Lancet Oncology. 2016;17(3):273-4.
25. Gao Y, Hernandez C, Yuan HX, Lilly J, Kota P, Zhou H, et al. Ultrasound molecular imaging of ovarian cancer with CA-125 targeted nanobubble contrast agents. Nanomedicine. 2017;13(7):2159-68.
26. Ghasemi N, Ghobadzadeh S, Zahraei M, Mohammadpour H, Bahrami S, Ganje MB, et al. HE4 combined with CA125: favorable screening tool for ovarian cancer. Medical Oncology. 2014;31(1):808.
27. Giampaolino P, Della Corte L, Foreste V, Vitale SG, Chiofalo B, Cianci S, et al. Unraveling a difficult diagnosis: the tricks for early recognition of ovarian cancer. Minerva medica. 2019;06.
28. Havrilesky LJ, Whitehead CM, Rubatt JM, Cheek RL, Groelke J, He Q, et al. Evaluation of biomarker panels for early stage ovarian cancer detection and monitoring for disease recurrence. Gynecologic Oncology. 2008;110(3):374-82.
29. Hidalgo Mora JJ, Rams Llop N, Ros Bernal F, Alcazar Zambrano JL. Classification systems and prediction of risk of malignancy of the adnexal masses. Clinica e Investigacion en Ginecologia y Obstetricia. 2018;45(1):24-31.
30. Jacob F, Meier M, Caduff R, Goldstein D, Pochechueva T, Hacker N, et al. No benefit from combining HE4 and CA125 as ovarian tumor markers in a clinical setting. Gynecologic Oncology. 2011;121(3):487-91.
31. Jeong TD, Cho EJ, Ko DH, Lee W, Chun S, Kwon HJ, et al. A new strategy for calculating the risk of ovarian malignancy algorithm (ROMA). Clinical Chemistry and Laboratory Medicine. 2017;55(8):1209-14.
32. Jia LT, Zhang YC, Li J, Tian Y, Li JF. The role of human epididymis protein 4 in the diagnosis of epithelial ovarian cancer. Clinical & Translational Oncology: Official Publication of the Federation of Spanish Oncology Societes & of the National Cancer Institute of Mexico. 2016;18(3):233-9.

Table 19. Excluded studies: study design *(Continued)*

33. Jiang ZH, Li KT, Tian JW, Ren M. An overview of the development and application of the sonographic scoring system: differentiation of malignant from benign ovarian tumors. Archives of Gynecology & Obstetrics. 2016;293(2):303-10.

34. Karimi-Zarchi M, Behtash N, Mousavi A, Maghami FG, Gilani MM, Chiti Z, et al. A survey on the role of cancer antigen 125 (CA125), human epididymis protein 4 (HE4), risk of ovarian malignancy algorithm (ROMA), and risk of malignancy index (RMI) in pelvic mass. International Journal of Cancer Management. 2018;11 (12) (no pagination)(e79189).

36. Kim B, Park Y, Kim B, Ahn HJ, Lee KA, Chung JE, et al. Diagnostic performance of CA 125, HE4, and risk of Ovarian Malignancy Algorithm for ovarian cancer. Journal of Clinical Laboratory Analysis. 2019;33(1):e22624.

37. Lang S, Armstrong N, Deshpande S, Ramaekers B, Grimm S, de Kock S, et al. Clinically inappropriate post hoc exclusion of study participants from test accuracy calculations: the ROMA score, an example from a recent NICE diagnostic assessment. Annals of Clinical Biochemistry. 2019;56(1):72-81.

38. Lennox GK, Eiriksson LR, Reade CJ, Leung F, Mojtahedi G, Atenafu EG, et al. Effectiveness of the risk of malignancy index and the risk of ovarian malignancy algorithm in a cohort of women with ovarian cancer: does histotype and stage matter? International Journal of Gynecological Cancer. 2015;25(5):809-14.

39. Mapelli P, Incerti E, Fallanca F, Gianolli L, Picchio M. Imaging biomarkers in ovarian cancer: the role of 18F-FDG PET/CT. The Quarterly Journal of Nuclear Medicine. 2016;60(2):93-102.

40. Montagnana M, Benati M, Danese E. Circulating biomarkers in epithelial ovarian cancer diagnosis: From present to future perspective. Annals of Translational Medicine. 2017;5 (13) (no pagination)(276).

41. Meys EMJ, Jeelof LS, Ramaekers BLT, Dirksen CD, Kooreman LFS, Slangen BFM, et al. Economic evaluation of an expert examiner and different ultrasound models in the diagnosis of ovarian cancer. European Journal of Cancer. 2018;100:55-64.

42. Meys EM, Kaijser J, Kruitwagen RF, Slangen BF, Van Calster B, Aertgeerts B, et al. Subjective assessment versus ultrasound models to diagnose ovarian cancer: A systematic review and meta-analysis. European Journal of Cancer. 2016;58:17-29.

43. Meys EMJ, Kaijser J, Kruitwagen RFPM, Slangen BFM, Van Calster B, Aertgeerts B, et al. Subjective assessment versus ultrasound models to diagnose ovarian cancer: A review and meta-analysis. International Journal of Gynecological Cancer. 2015;1):33-4.

44. Mills P, Court S, Giamougiannis P, Daines L. Is the risk of malignancy (RMI) score useful in deciding management when below 250? A 2-year retrospective surgical study. BJOG: An International Journal of Obstetrics and Gynaecology. 2015;2):144-5.

45. Nandamuru D, Collins MK, Tarney CM. Evaluation and management of adnexal masses in postmenopausal women. Current Women's Health Reviews. 2019;15(1):41-5.

46. Ohler A, Pawlowski M, Dudziak M, Sznurkowski JJ. The new FIGO staging system and the review of current methods of detection of ovarian cancer. Onkologia i Radioterapia. 2015;32(2):9-18.

47. Petrovic-Simovic V. Comparing ROMA score for the same two periods in 2015 and 2018 in patients on observation. Clinica Chimica Acta. 2019;493 (Supplement 1):S116-S7.

48. Prskalo ZS, Gace M, Dobrijevic S, Mayer L. Benefits human epidydimis protein (HE4) compared to traditional used tumor markers in gynecological oncology. Libri Oncologici. 2015;43(1-3):9-14.

49. Puljiz M, Mayer L. Multidisciplinary approach to diagnostics, treatment and follow up of patients with ovarian cancer: Potential of he4 and roma index. Libri Oncologici. 2015;43(1-3):1.

50. Sadowski EA, Robbins JB, Rockall AG, Thomassin-Naggara I. A systematic approach to adnexal masses discovered on ultrasound: the ADNEx MR scoring system. Abdominal Radiology. 2018;43(3):679-95.

51. Shetty M. Imaging and Differential Diagnosis of Ovarian Cancer. Seminars in Ultrasound, CT and MRI. 2019.

52. Skates SJ. EPIC Early Detection of Ovarian Cancer. Clinical Cancer Research. 2016;22(18):4542-4.

Table 19. Excluded studies: study design *(Continued)*

53. Soletormos G, Duffy MJ, Othman Abu Hassan S, Verheijen RH, Tholander B, Bast RC, Jr, et al. Clinical Use of Cancer Biomarkers in Epithelial Ovarian Cancer: Updated Guidelines From the European Group on Tumor Markers. International Journal of Gynecological Cancer. 2016;26(1):43-51.
54. Stukan M, Dudziak M, Ratajczak K, Grabowski JP. Usefulness of diagnostic indices comprising clinical, sonographic, and biomarker data for discriminating benign from malignant ovarian masses. Journal of Ultrasound in Medicine. 2015;34(2):207-17.
55. Tcherkassova J, Abramovich C, Moro R, Chen C, Schmit R, Gerber A, et al. Combination of CA125 and RECAF biomarkers for early detection of ovarian cancer. Tumour Biology. 2011;32(4):831-8.
56. Terry KL, Schock H, Fortner RT, Husing A, Fichorova RN, Yamamoto HS, et al. A Prospective Evaluation of Early Detection Biomarkers for Ovarian Cancer in the European EPIC Cohort. Clinical Cancer Research. 2016;22(18):4664-75.
57. Thakur M, Timmerman D. Imaging of Adnexal Masses. Clinical Obstetrics and Gynecology. 2017;60(1):38-45.
58. Thomassin-Naggara I, Fedida B, Kermarrec E. Adnexal masses: Characterization of benign adnexal masses. Medical Radiology. 2019:273-85.
59. Van Calster B. External validation of ADNEX model for diagnosing ovarian cancer: evaluating performance of differentiation between tumor subgroups. Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2017;50(3):406-7.
60. van Nagell JR, Jr, Miller RW. Evaluation and Management of Ultrasonographically Detected Ovarian Tumors in Asymptomatic Women. Obstetrics & Gynecology. 2016;127(5):848-58.
61. Walker M, Sobel M. Diagnosing ovarian cancer. Cmaj. 2018;190(42):E1259.
62. Westwood M, Ramaekers B, Lang S, Grimm S, Deshpande S, de Kock S, et al. Risk scores to guide referral decisions for people with suspected ovarian cancer in secondary care: a systematic review and cost-effectiveness analysis. Health Technology Assessment (Winchester, England). 2018;22(44):1-264.
63. Zhang Z, Barnhill SD, Zhang H, Xu F, Yu Y, Jacobs I, et al. Combination of multiple serum markers using an artificial neural network to improve specificity in discriminating malignant from benign pelvic masses. Gynecologic Oncology. 1999;73(1):56-61.

Table 20. Excluded studies: test positivity threshold

Test positivity threshold
1. Banu SA, Khatun S, Shamsuddin L. Assesment of adnexal masses by transvaginal sonography and serum CA125 assay in pre- and postmenopausal women. Bangladesh Journal of Obstetrics and Gynecology. 2009;24(2):56-62.
2. Byler T, Nsouli I. Incidental CT bladder wall abnormalities: Harbinger or herring? Urology. 2014;1:S193-S4.
3. Bouzari Z, Rahimi H, Gholinia H, Yazdani S, Hajian-Tilaki K, Soleimani MJ. Cancer antigen 125 (CA125), human epididymis protein 4 (HE4), risk of malignancy index (RMI), and risk of ovarian malignancy algorithm (ROMA) as diagnostic tests in ovarian cancer. International Journal of Cancer Management. 2019;12 (1) (no pagination)(e59395).
4. Cradic KW, Lasho MA, Algeciras-Schimnich A. Validation of the Cut-points Recommended for ROMA Using the Roche Elecsys CA125 and HE4 Assays. Annals of Clinical & Laboratory Science. 2018;48(1):90-3.
5. Ferdeghini M, Gadducci A, Prontera C, Malagnino G, Annicchiarico C, Prato B, et al. COMBINED EVALUATION OF SERUM CA-125 AND CAM-29 IN PATIENTS WITH EPITHELIAL OVARIAN-CANCER. Tumor Biology. 1992;13(5-6):287-93.
6. Franchi M, Beretta P, Ghezzi F, Zanaboni F, Goddi A, Salvatore S. DIAGNOSIS OF PELVIC MASSES WITH TRANSABDOMINAL COLOR DOPPLER, CA-125 AND ULTRASONOGRAPHY. Acta Obstetricia et Gynecologica Scandinavica. 1995;74(9):734-9.

Table 20. Excluded studies: test positivity threshold *(Continued)*

7. Kadayifci A, Simsek H, Savas MC. Serum CA 125 levels in patients with ovarian cancer and liver cirrhosis. Medical Principles and Practice. 1997;6(3):137-41.

8. Katsyuba M, Khasanov R, Usmanova G, Ratner E, Muratova G, Nigmatulina N. Can copenhagen index replace risk of ovarian malignancy algorithm (ROMA) in a triage of patients with pelvic mass? International Journal of Gynecological Cancer. 2017;27 (Supplement 4):303.

9. Le Page C, Ouellet V, Madore J, Hudson TJ, Tonin PN, Provencher DM, et al. From gene profiling to diagnostic markers: IL-18 and FGF-2 complement CA125 as serum-based markers in epithelial ovarian cancer. International Journal of Cancer. 2006;118(7):1750-8.

10. Lenhard M, Stieber P, Hertlein L, Kirschenhofer A, Furst S, Mayr D, et al. The diagnostic accuracy of two human epididymis protein 4 (HE4) testing systems in combination with CA125 in the differential diagnosis of ovarian masses. Clinical Chemistry & Laboratory Medicine. 2011;49(12):2081-8.

11. Mansour GM, El-Lamie IK, El-Sayed HM, Ibrahim AM, Laban M, Abou-Louz SK, et al. Adnexal mass vascularity assessed by 3-dimensional power Doppler: does it add to the risk of malignancy index in prediction of ovarian malignancy?: four hundred-case study. International journal of gynaecological Cancer. 2009;19(5):867-72.

12. Moore RG, Brown AK, Miller MC, Skates S, Allard WJ, Verch T, et al. The use of multiple novel tumor biomarkers for the detection of ovarian carcinoma in patients with a pelvic mass. Gynecologic Oncology. 2008;108(2):402-8.

13. Moore RG, Jabre-Raughley M, Brown AK, Robison KM, Miller MC, Allard WJ, et al. Comparison of a novel multiple marker assay vs the Risk of Malignancy Index for the prediction of epithelial ovarian cancer in patients with a pelvic mass. American Journal of Obstetrics & Gynecology. 2010;203(3):228.e1-6.

14. Nolen B, Velikokhatnaya L, Marrangoni A, De Geest K, Lomakin A, Bast RC, Jr, et al. Serum biomarker panels for the discrimination of benign from malignant cases in patients with an adnexal mass. Gynecologic Oncology. 2010;117(3):440-5.

15. Shabana A, Onsrud M. Tissue polypeptide-specific antigen and CA 125 as serum tumor markers in ovarian carcinoma. Tumor Biology. 1994;15(6):361-7.

16. Shah CA, Lowe KA, Paley P, Wallace E, Anderson GL, McIntosh MW, et al. Influence of ovarian cancer risk status on the diagnostic performance of the serum biomarkers mesothelin, HE4, and CA125. Cancer Epidemiology Biomarkers and Prevention. 2009;18(5):1365-72.

17. Timmerman D, Van Calster B, Jurkovic D, Valentin L, Testa AC, Bernard J-P, et al. Inclusion of CA-125 does not improve mathematical models developed to distinguish between benign and malignant adnexal tumors. Journal of Clinical Oncology. 2007;25(27):4194-200.

18. Vinokurov VL, Dudarev AL, Jurkova LE, Lapchenkov VI, Barbanel EJ. Tumor marker CA 125 in diagnosis, monitoring management and follow-up of patients with ovarian tumors. European Journal of Gynaecological Oncology. 1992;13(2):205-8.

19. Zheng H, Gao Y. Serum HE4 as a useful biomarker in discriminating ovarian cancer from benign pelvic disease. International Journal of Gynecological Cancer. 2012;22(6):1000-5.

APPENDICES

Appendix 1. Search strategies 2015

1. OVARIAN CANCER – ULTRASOUND/IOTA

Database: MEDLINE (Ovid) 1946 to April Week 3 2015

1 exp Ovarian Neoplasms/di
2 exp Adnexal Diseases/di
3 ((borderline or border line) adj4 ovar$).tw.

4 exp Fallopian Tube Neoplasms/di
5 exp Peritoneal Neoplasms/di
6 exp Pelvic Neoplasms/di
7 ((ovar$ or adnexal or fallopian or peritoneal$ or pelvic) adj3 (cancer$ or carcinoma$ or malignan$ or mass or masses or cyst or cysts or neoplas$ or tumour$ or tumor$)).tw.
8 ((epithelial or germ cell) adj5 ovar$).tw.
9 or/1-8
10 exp ultrasonography/
11 ultraso$.tw.
12 (transvagina$ adj2 sonogra$).tw.
13 or/10-12
14 9 and 13
15 limit 14 to (human and yr=1991-2015)
16 IOTA.tw.
17 International Ovarian Tumor Analysis.tw.
18 ((ovarian or epithelial or adnex$ or fallopian or peritoneal or pelvic) adj3 (model$ or regress$ or rule$ or score$ or algorithm$ or term $ or definition$ or measure$)).ti,ab.
19 or/16-18
20 9 and 19
21 limit 20 to human
22 15 or 21

Database: MEDLINE (Ovid) In-Process & Other Non-Indexed Citations 27 April 2015

1 ((borderline or border line) adj4 ovar$).tw.
2 ((ovar$ or adnexal or fallopian or peritoneal$ or pelvic) adj3 (cancer$ or carcinoma$ or malignan$ or mass or masses or cyst or cysts or neoplas$ or tumour$ or tumor$)).tw.
3 ((epithelial or germ cell) adj5 ovar$).tw.
4 or/1-3
5 ultraso$.tw.
6 (transvagina$ adj2 sonogra$).tw.
7 or/5-6
8 4 and 7
9 limit 8 to yr="1991-2015"
10 IOTA.tw.
11 International Ovarian Tumor Analysis.tw.
12 ((ovarian or epithelial or adnex$ or fallopian or peritoneal or pelvic) adj3 (model$ or regress$ or rule$ or score$ or algorithm$ or term $ or definition$ or measure$)).ti,ab.
13 or/10-12
14 4 and 13
15 9 or 14

Database: Embase (Ovid) 1974 to 27 April 2015

1 ((borderline or border line) adj4 ovar$).tw.
2 uterine tube tumor/di [Diagnosis]
3 peritoneum tumor/di [Diagnosis]
4 pelvis tumor/di [Diagnosis]
5 ovary tumor/di [Diagnosis]
6 adnexa disease/di [Diagnosis]
7 ((ovar$ or adnexal or fallopian or peritoneal or pelvic) adj3 (cancer$ or carcinoma$ or malignan$ or mass or masses or cyst or cysts or neoplas$ or tumour$ or tumor$)).tw.
8 ((epithelial or germ cell) adj5 ovar$).tw.
9 or/1-8
10 ultraso$.tw.
11 (transvagina$ adj2 sonogra$).tw.
12 ultrasound/
13 or/10-12
14 9 and 13
15 limit 14 to (humans and yr="1991-2015")
16 IOTA.tw.
17 International Ovarian Tumor Analysis.tw.

18 ((ovarian or epithelial or adnex$ or fallopian or peritoneal or pelvic) adj3 (model$ or regress$ or rule$ or score$ or algorithm$ or term $ or definition$ or measure$)).tw.
19 or/16-18
20 9 and 19
21 15 or 20
22 limit 21 to humans

Database: Cochrane Library (Wiley) 27 April 2015 CENTRAL, CDSR Issue 4 of 12, HTA DARE Issue 2 of 4 2015

#1 borderline near/4 ovar*
#2 "border line" near/4 ovar*
#3 MeSH descriptor: [Fallopian Tube Neoplasms] explode all trees and with qualifier(s): [Diagnosis - DI]
#4 MeSH descriptor: [Pelvic Neoplasms] explode all trees and with qualifier(s): [Diagnosis - DI]
#5 MeSH descriptor: [Ovarian Neoplasms] explode all trees and with qualifier(s): [Diagnosis - DI]
#6 MeSH descriptor: [Adnexal Diseases] explode all trees and with qualifier(s): [Diagnosis - DI]
#7 (ovar* or adnexal or fallopian or peritoneal or pelvic) near/3 (cancer* or carcinoma* or malignan* or mass or masses or cyst or cysts or neoplasm* or tumor* or tumour*)
#8 (epithelial or "germ cell") near/5 (ovar*)
#9 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8
#10 ultraso*
#11 MeSH descriptor: [Ultrasonography] explode all trees
#12 transvagina* near/2 sonogra*
#13 #10 or #11 or #12
#14 #9 and #13 Publication Year from 1991 to 2015
#15 IOTA
#16 "International Ovarian Tumor Analysis"
#17 (ovarian or epithelial or adnex* or fallopian or peritoneal or pelvic*) near/3 (model* or regress* or rule* or score* or algorithm* or term* or definition* or measure*)
#18 #15 or #16 or #17
#19 #9 and #18
#20 #14 or #19

Database: CINAHL (EBSCO) 1960 – 27 April 2015

S1 (borderline or border-line) N4 (ovar*)
S2 (MH “Fallopian Tube Diseases+/DI)
S3 (MH “Peritoneal Neoplasms+/DI)
S4 (MH “Pelvic Neoplasms/DI”)
S5 (MH “Ovarian Neoplasms+/DI”
S6 (MH “Adnexal Diseases/DI”
S7 (ovar* or adnexal or fallopian or peritoneal or pelvic) N3 (cancer* or carcinoma* or malignan* or mass or masses or cyst or cysts or neoplasm* or tumour* or tumor*)
S8 (epithelial or germ cell) N1 (ovar*)
S9 S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8
S10 “ultraso*”
S11 (MH”Ultrasonography+)
S12 transvagina* N2 sonogra*
S13 S10 or S11 or S12
S14 S9 and S13 Limiters – Publication Year: 1991 – 2015
S15 “IOTA” or “international ovarian tumor analysis”
S16 (ovarian or epithelial or adnex*) N5 (model* or regress* or rule* or score* or algorithm* or term* or definition* or measure*)
S17 S15 or S16
S18 S17 or S14

Database: Science Citation Index (Web of Science) 1900 to 23 April 2015

#1 TS=(borderline ovar* or border line ovar*)
#2 TS=((ovar* or adnexal or fallopian or peritoneal or pelvic) N3 (cancer* or carcinoma* or malignan* or mass or masses or cyst or cysts or neoplasm* or tumour* or tumor*)
#3 TS=(((epithelial or “germ cell”)) near/1 (ovar*)
#4 #3 or #2 or #1
#5 TS=ultraso*
#6 TS=(transvagina* near/2 sonogra*)

#7 TS=#5 or #6
#8 TS=IOTA
#9 TS=(ovarian or epithelial or adnex*) near/2 (model* or regress* or rule* or score* or algorithm* or term* or definition* or measure*)
#10= #8 or #9
#11 #4 and #7 Indexes= SCI-EXPANDED Timespan= 1991-2015
#12 #10 and #4 Indexes= SCI-EXPANDED Timespan= 1991-2015
#13 #11 or #12

Database: Conference Proceedings Citation Index (CPCI) (Web of Science) 1900 to 24 April 2015

As Science Citation Index above. Searched 24 April 2015

2. OVARIAN CANCER SYMPTOM SCORES

Database: MEDLINE (Ovid) 1946 to March Week 4 2015

1 exp ovarian neoplasms/di
2 exp adnexal diseases/di
3 ((ovar$ or adnexal or fallopian or peritoneal$ or pelvic) adj3 (cancer$ or carcinoma$ or malignan$ or mass or masses or cyst or cysts or neoplas$ or tumor$ or tumour$)).tw.
4 ((borderline or border line) adj4 ovar$).tw.
5 exp Fallopian Tube Neoplasms/di
6 exp Peritoneal Neoplasms/di
7 exp pelvic neoplasms/di
8 ((epithelial or germ cell) adj5 ovar$).tw.
9 or/1-8
10 exp "Signs and Symptoms"/
11 symptom$.ti,ab.
12 exp early diagnosis/ or exp Diagnosis/
13 exp "Early Detection of Cancer"/
14 (early adj (sign$ or symptom$)).tw.
15 (abdom$ adj3 (pressure or pain$ or swelling$ or hard)).tw.
16 (bowel irregularit$ or bloat$ or fullness or satiet$ or gastro$).tw.
17 (fatigue or weight loss$ or weight gain$ or constipat$ or diarrhoea or diarrhea or gas).tw.)
18 (nausea$ or indigestion).tw.
19 ((loss or lack) adj3 (energ$ or appetite$)).tw.
20 (urin$ adj3 (frequenc$ or urgenc$)).tw.
21 ((leg$ or ankle$) adj2 (swell$ or swollen)).tw.
22 ((abnormal or irrregular or postmenopausal) adj1 vaginal adj (bleed$ or discharge$)).tw.
23 (pelvic discomfort$ or pelvic pain$ or chest pain$ or respirator$ difficult$ or lower back pain$).tw.
24 or/10-22
25 9 and 24
26 (index or risk$ or score$ or scoring or checklist$ or rule$ or indices or tool$ or instrument$ or survey$ or questionnaire$ or interview$).tw.
27 25 and 26
28 limit 27 to (humans and yr="2009 - 2015")

Database: MEDLINE (Ovid) In-Process & Other Non-Indexed Citations 20 March 2015

1 ((borderline or border line) adj4 ovar$).tw.
2 ((ovar$ or fallopian or peritoneal$ or pelvic) adj3 (cancer$ or carcinoma$ or malignan$ or mass or masses or cyst or cysts or neoplas$ or tumour$ or tumor$)).tw.
3 ((epithelial or germ cell) adj5 ovar$).tw.
4 or/1-3
5 (symptom$ or sign$).tw.
6 (early adj2 (sign$ or detect$ or diagnos$)).tw.
7 (abdom$ adj3 (pressure or pain$ or swelling$ or hard)).tw.
8 (bowel irregularit$ or bloat$ or fullness or satiet$ or gastro$).tw.
9 (fatigue or weight loss or weight gain$ or constipat$ or diarrhoea or diarrhea or gas).tw.
10 nausea$ or indigestion.tw.
11 ((lack or loss) adj3 (energ$ or appetite$)).tw.
12 (urin$ adj3 (frequenc$ or urgenc$)).tw.
13 ((leg$ or ankle$) adj2 (swell$ or swollen)).tw.
14 ((abnormal or irregular$ or postmenopausal) adj1 vaginal adj (bleed$ or discharge$)).tw.

15 (pelvic discomfort$ or pelvic pain$ or chest pain$ or respirator$ difficult$ or lower back pain$).tw.
16 or/5-15
17 (index or risk$ or score$ or scoring or checklist$ or rule$ or indices or tool$ or instrument$ or survey$ or questionnaire$ or interview$).tw.
18 4 and 16 and 17
19 limit 18 to yr="2009 - 2015"

Database: Embase (Ovid) 1974 to 27 March 2015

1 ((ovar$ or adnexal or fallopian or peritoneal or pelvic) adj3 (cancer$ or carcinoma$ or malignan$ or mass or masses or cyst or cysts or neoplas$ or tumour$ or tumor$)).tw.
2 ((epithelial or germ cell) adj ovar$).tw.
3 ((borderline or border line) adj4 ovar$).tw.
4 uterine tube tumor/di
5 peritoneum tumor/di
6 pelvis tumor/di
7 ovary tumor/di [Diagnosis]
8 adnexa disease/di
9 or/1-8
10 symptom/ or symptom$.tw.
11 early diagnosis/
12 diagnosis/
13 (early adj (sign$ or symptom$)).tw.
14 (abdom$ adj3 (pressure or pain$ or swelling$ or hard)).tw.
15 (bowel irregularit$ or bloat$ or fullness or satiet$ or gastro$).tw.
16 (fatigue or weight loss$ or weight gain$ or constipat$ or diarrhoea or diarrhea or gas).tw.
17 nausea$.mp. or indigestion.tw.
18 ((loss or lack) adj3 (energ$ or appetit$)).tw.
19 (urin$ adj3 (frequenc$ or urgenc$)).tw.
20 ((leg$ or ankle$) adj2 (swell$ or swollen)).tw.
21 ((abnormal or irregular or postmenopausal) adj1 vaginal adj (bleed$ or discharge$)).tw.
22 (pelvic discomfort$ or pelvic pain$ or chest pain$ or respirator$ difficult$ or lower back pain$).tw.
23 or/10-22
24 9 and 23
25 (index or risk$ or score$ or scoring or checklist$ or rule$ or indices or tool$ or instrument$ or survey$ or questionnaire$ or interview$).tw.
26 24 and 25
27 limit 26 to (human and yr="2009 - 2015")

Database: Cochrane Library (Wiley) 23 February 2015 CENTRAL, CDSR Issue 1 of 12 HTA DARE Issue 1 of 4 2015

#1 borderline near/4 ovar*
#2 "border line" near/4 ovar*
#3 MeSH descriptor: [Fallopian Tube Neoplasms] explode all trees and with qualifier(s): [Diagnosis - DI]
#4 MeSH descriptor: [Pelvic Neoplasms] explode all trees and with qualifier(s): [Diagnosis - DI]
#5 MeSH descriptor: [Ovarian Neoplasms] explode all trees and with qualifier(s): [Diagnosis - DI]
#6 MeSH descriptor: [Adnexal Diseases] explode all trees and with qualifier(s): [Diagnosis - DI]
#7 (ovar* or adnexal or fallopian or peritoneal or pelvic) near/3 (cancer* or carcinoma* or malignan* or mass or masses or cyst or cysts or neoplasm* or tumor* or tumour*)
#8 (epithelial or "germ cell") next (ovar*)
#9 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8
#10 MeSH descriptor: [Signs and Symptoms] explode all trees
#11 MeSH descriptor: [Early Diagnosis] explode all trees
#12 early near/1 (sign* or symptom*)
#13 (abdom*) near/3 (pressure* or pain* or swelling or hard)
#14 bloat* or fullness or satiet* or gastro*
#15 bowel next irregular*
#16 fatigue or "weight loss" or "weight gain" or constipat* or diarrhoea or diarrhea or gas or nausea* or indigestion
#17 (loss or lack) near/3 (appetit*)
#18 (urin*) near/3 (frequenc* or urgenc*)
#19 Leg* or ankle* near/2 (swell* or swollen)
#20 (loss or lack) near/3 (energy)
#21 (abnormal or irregular or postmenopausal) near/1 (vaginal) near/1 (bleed* or discharge*)
#22 "pelvic discomfort" or "pelvic pain" or "chest pain*" or "respirator* difficult*" or "lower back pain"
#23 #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #20 or #21 or #22

#24 #9 and #23
#25 index* or risk* or score* or scoring or checklist* or rule* or indices or tool* or instrument* or survey* or questionnaire* or interview*
#26 #24 and #25 Publication Year from 2009 to 2015

Database: CINAHL (EBSCO) 1960 – 23 February 2015

S1 (borderline or border-line) N4 (ovar*)
S2 (MH "Fallopian Tube Diseases+/DI)
S3 (MH "Peritoneal Neoplasms+/DI)
S4 (MH "Pelvic Neoplasms/DI")
S5 (MH "Ovarian Neoplasms+/DI"
S6 (MH "Adnexal Diseases/DI"
S7(ovar* or adnexal or fallopian or peritoneal or pelvic) N3 (cancer* or carcinoma* or malignan* or mass or masses or cyst or cysts or neoplasm* or tumour* or tumor*)
S8 (epithelial or germ cell) N1 (ovar*)
S9 S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8
S10 (MH "Symptoms")
S11 (MH "Early Diagnosis+")
S12 (MM"Diagnosis")
S13 early warning sign*
S14 (abdom*) N5 (pressure or pain* or swelling or hard*)
S15 bowel irregularit* or bloat* or fullness or satiet* or gastro*
S16 fatigue or weight loss* or weight gain* or constipat* or diarrhoea or gas or nausea* or indigestion
S17 loss N1 appetit*
S18 Lack N1 energy
S19 urin* N3 (frequenc* or urgenc*)
S20 Leg N2 (swell* or swollen)
S21 (abnormal or irregular or postmenopausal) N1 (vaginal bleed*) or (vaginal discharge*)
S22 pelvic discomfort* or pelvic pain* or chest pain* or respirator* difficult* or lower back pain
S23 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
S24 S9 and S23
S25 index or risk* or score* or scoring or checklist* or rule* or indices or tool or instrument* or survey* or questionnaire* or interview*
S26 S24 and S25
S27 S24 and S25 Limiters – Publication Year: 2009-2015

Science Citation Index (Web of Science) 1900 to 23 February 2015

#1 TS=(borderline ovar* or border line ovar*)
#2 TS=((ovar* or adnexal or fallopian or peritoneal or pelvic) N3 (cancer* or carcinoma* or malignan* or mass or masses or cyst or cysts or neoplasm* or tumour* or tumor*)
#3 TS=(((epithelial or "germ cell")) near/1 (ovar*)
#4 #3 or #2 or #1
#5 TS=symptom*
#6 TS="early diagnosis"
#7 TS="early warning sign*"
#8 TS=(((abdom*) near/5 (pressure* or pain* or swelling* or hard)))#9 TS=((bowel irregularit* or bloat* or fullness or satiet* or gastro*))
#10 TS=((fatigue or weight loss or weight gain or constipat* or diarrhoea or gas or nausea or indigestion))
#11 TS=((loss near/1 appetit*))
#12 TS=((lack near/1 energ*))
#13 TS=((urin*) near/3 (frequenc* or urgenc*))
#14 TS=((leg) near/2 (swell* or swollen)
#15 TS=(("pelvic discomfort" or "pelvic pain" or "chest pain" or respirator* difficult* or "lower back pain"))
#16 TS=((index or risk* or score* or scoring or checklist* or rule* or indices or tool* or instrument* or survey* or questionnaire* or interview*))
#17 #15 or #14 or #13 or #12 or #11 or #10 or #9 or #8 or #7 or #6 or #5
#18 #17 and #16 and #4 Limited: 2009-2015

Database: Conference Proceedings Citation Index (CPCI) (Web of Science) 1900 to 23 February 2015

As Science Citation Index above.

3. OVARIAN CANCER BIOMARKERS

Database: MEDLINE (Ovid) 1946 to April Week 3 2015

1 exp Ovarian Neoplasms/di
2 exp Adnexal Diseases/di
3 ((ovar$ or adnexal or fallopian or peritoneal$ or pelvic) adj3 (cancer$ or carcinoma$ or malignan$ or mass or masses or cyst or cysts or neoplas$ or tumour$ or tumor$)).tw.
4 ((borderline or border line) adj4 ovar$).tw.
5 exp Fallopian Tube Neoplasms/di
6 exp Peritoneal Neoplasms/di
7 exp Pelvic Neoplasms/di
8 ((epithelial or germ cell) adj5 ovar$).tw.
9 or/1-8
10 exp Tumor Markers, Biological/
11 exp Biological Markers/
12 Proteomics/
13 Genetic Markers/
14 Metabolomics/
15 multiplex$.tw.
16 multivariate.tw.
17 (CA125 or CA-125 or HE4 or OVA 1 or OVA1 or HCG or LDH or AFP).mp. or CEA.tw. [
18 CA-125 Antigen/
19 Chorionic Gonadotropin/
20 L-Lactate Dehydrogenase/
21 alpha-Fetoproteins/
22 Carcinoembryonic Antigen/
23 or/10-22
24 9 and 23
25 limit 24 to (humans and yr="1991-2015")

Database: MEDLINE (Ovid) In-Process & Other Non-Indexed Citations April 23, 2015

1 ((borderline or border line) adj4 ovar$).tw.
2 ((ovar$ or adnexal or fallopian or peritoneal$ or pelvic) adj3 (cancer$ or carcinoma$ or malignan$ or mass or masses or cyst or cysts or neoplas$ or tumour$ or tumor$)).tw.
3 ((epithelial or germ cell) adj5 ovar$).tw.
4 or/1-3
5 ((genetic or protein$) adj1 assay$).ti,ab.
6 multiplex.ti,ab.
7 ((multivariate or multimarker$) adj2 assay$).ti,ab.
8 (biomarker$ or marker$ or metabolomic$ or proteomic$ or lipomic$ or kallikrein$ or genomic$).ti,ab.
9 (CA125 or CA-125 or HE4 or OVA1 or OVA 1 or HCG or LDH or AFP).mp. or CEA.tw.
10 CA-125 antigen.tw.
11 chorionic gonadotropin.tw.
12 L-lactate dehydrogenase.tw.
13 alpha-fetoprotein$.tw.
14 carcinoembryonic antigen$.tw.
15 or/5-14
16 4 and 15
17 limit 16 to yr="1991 - 2015"

Database: EMBASE (Ovid) 1974 to 23 April 2015

1 ((borderline or border line) adj4 ovar$).tw.
2 uterine tube tumor/di [Diagnosis]
3 peritoneum tumor/di [Diagnosis]
4 pelvis tumor/di [Diagnosis]
5 ovary tumor/di [Diagnosis]
6 adnexa disease/di [Diagnosis]
7 ((ovar$ or adnexal or fallopian or peritoneal or pelvic) adj3 (cancer$ or carcinoma$ or malignan$ or mass or masses or cyst or cysts or neoplas$ or tumour$ or tumor$)).tw.
8 ((epithelial or germ cell) adj5 ovar$).tw.
9 or/1-8
10 multiplex$.tw.

11 ((multivariate or multimarker$) adj2 assay$).ti,ab.
12 exp tumor marker/
13 exp biological marker/
14 exp proteomics/
15 exp genetic marker/
16 exp metabolomics/
17 (CA125 or CA-125 or HE4 or OVA1 or OVA 1 or HCG or LDH or AFP).mp. or CEA.tw.
18 or/10-17
19 9 and 18
20 limit 19 to (humans and yr="1991-2015")

Database: Cochrane Library (Wiley) 23 April 2015 CENTRAL, CDSR Issue 4 of 12 HTA, DARE, Issue 2 of 4 2015

#1 borderline near/4 ovar*
#2 border next line near/4 ovar*
#3 MeSH descriptor: [Fallopian Tube Neoplasms] explode all trees and with qualifier(s): [Diagnosis - DI]
#4 MeSH descriptor: [Peritoneal Neoplasms] explode all trees
#5 MeSH descriptor: [Pelvic Neoplasms] explode all trees and with qualifier(s): [Diagnosis - DI]
#6 MeSH descriptor: [Ovarian Neoplasms] explode all trees and with qualifier(s): [Diagnosis - DI]
#7 MeSH descriptor: [Adnexal Diseases] explode all trees and with qualifier(s): [Diagnosis - DI]
#8 (ovar* or adnexal or fallopian or peritoneal or pelvic) near/2 (cancer* or carcinoma* or malignan* or mass or masses or cyst or cysts or neoplasm* or tumor* or tumour*)
#9 (epithelial or "germ cell") next (ovar*)
#10 #1 or #2 or #3 or #5 or #6 or #7 or #8 or #9
#11 biomarker*
#12 marker*
#13 metabolomics*
#14 genetic next assay*
#15 protein* next assay*
#16 proteomic*
#17 lipomic*
#18 multiplex
#19 multivariate or multimarker near/2 assay*
#20 kallikrein*
#21 genomic*
#22 MeSH descriptor: [Biological Markers] explode all trees
#23 MeSH descriptor: [Proteomics] explode all trees
#24 MeSH descriptor: [Kallikreins] explode all trees
#25 MeSH descriptor: [Genomics] explode all trees
#26 CA125 or CA-125 or HE4 or OVA 1 or OVA1 or HCG or LDH or AFP or CEA
#27 MeSH descriptor: [CA-125 Antigen] explode all trees
#28 MeSH descriptor: [Chorionic Gonadotropin] explode all trees
#29 MeSH descriptor: [alpha-Fetoproteins] explode all trees
#30 MeSH descriptor: [Carcinoembryonic Antigen] explode all trees
#31 #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30
#32 #10 and #31 Publication Year from 1991 to 2015

Database: CINAHL (EBSCO) 1960 to 23 April 2015

S1 (borderline or border-line) N4 (ovar*)
S2 (MH “Fallopian Tube Diseases+/DI)
S3 (MH “Peritoneal Neoplasms+/DI)
S4 (MH “Pelvic Neoplasms/DI”)
S5 (MH “Ovarian Neoplasms+/DI”
S6 (MH “Adnexal Diseases/DI”
S7(ovar* or adnexal or fallopian or peritoneal or pelvic) N3 (cancer* or carcinoma* or malignan* or mass or masses or cyst or cysts or neoplasm* or tumour* or tumor*)
S8 (epithelial or germ cell) N1 (ovar*)
S9 S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8
S10 multiplex
S11 (multivariate or multimarker*) N2 (assay*)
S12 (MH “Biological Markers+”)
S13 (MH “Tumor Markers, Biological+”)

S14 (MM “Proteomics”)
S15 (MM “Genetic Markers”)
S16 “metabolomic*” or CA125 or CA-125 or HE4 or OVA1 or OVA1 or HCG or LDH or AFP or CEA
S17 S10 or S11 or S12 or S13 or S14 or S15 or S16
S18 S9 and S17 Limiters – Publication Year: 1991 – 2015

Database: Science Citation Index (Web of Science) 1900 to 23 April 2015

#1 TS=(borderline ovar* or border line ovar*)
#2 TS=((ovar* or adnexal or fallopian or peritoneal or pelvic) N3 (cancer* or carcinoma* or malignan* or mass or masses or cyst or cysts or neoplasm* or tumour* or tumor*)
#3 TS=(((epithelial or “germ cell”)) near/1 (ovar*)
#4 #3 or #2 or #1
#5 TS=multiplex
#6 TS=((((multivariate or multimarker*)) near/2 (assay*)))
#7 TS=(((tumor* or tumour* or genetic*) near/2 (marker*)))
#8 TS=(metabolom* or proteiomic*) or (CA125 or CA-125 or HE4 or OVA1 or OVA 1 or HCG or LDH or AFP or CEA)
#9 TS=((((genetic* or protein*)) near/1 (assay*)))
#10 #5 or #6 or #7 or #8 or #9
#11 #4 and #10 Indexes= SCI-EXPANDED Timespan= 1991-2015

Database: Conference Proceedings Citation Index (CPCI) (Web of Science) 1900 to 24 April 2015

As Science Citation Index above. Searched 24 April 2015

Appendix 2. Search strategies 2019

Database: Ovid MEDLINE(R) and In-Process, In-Data-Review & Other Non-Indexed Citations (1946 to 21 June 2019)

1 exp Ovarian Neoplasms/di
2 exp Adnexal Diseases/di
3 ((borderline or border line) adj4 ovar$).tw.
4 exp Fallopian Tube Neoplasms/di
5 exp Peritoneal Neoplasms/di
6 exp Pelvic Neoplasms/di
7 ((ovar$ or adnexal or fallopian or peritoneal$ or pelvic) adj3 (cancer$ or carcinoma$ or malignan$ or mass or masses or cyst or cysts or neoplas$ or tumour$ or tumor$)).tw.
8 ((epithelial or germ cell) adj5 ovar$).tw.
9 or/1-8
10 exp ovarian neoplasms/
11 "Neoplasms, Glandular and Epithelial"/
12 exp ovary/
13 10 or 11 or 12
14 9 or 13 (245101)
15 exp ultrasonography/
16 ultraso$.tw.
17 (transvagina$ adj2 sonogra$).tw.
18 15 or 16 or 17
19 IOTA.tw. (2231)
20 International Ovarian Tumor Analysis.tw.
21 ((ovarian or epithelial or adnex$ or fallopian or peritoneal or pelvic) adj3 (model$ or regress$ or rule$ or score$ or algorithm$ or term $ or definition$ or measure$)).ti,ab.
22 19 or 20 or 21
23 exp Tumor Markers, Biological/
24 exp Biological Markers/
25 *Proteomics/
26 *Genetic Markers/
27 *Metabolomics/
28 multiplex$.tw.
29 multivariate.tw.
30 (CA125 or CA-125 or HE4 or OVA 1 or OVA1 or HCG or LDH or AFP).mp. or CEA.tw.
31 CA-125 Antigen/
32 Chorionic Gonadotropin/

33 L-Lactate Dehydrogenase/
34 alpha-Fetoproteins/
35 Carcinoembryonic Antigen/
36 23 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35
37 exp "Signs and Symptoms"/
38 exp early diagnosis/ or exp Diagnosis/
39 exp "Early Detection of Cancer"/
40 symptom$.ti,ab.
41 (early adj (sign$ or symptom$)).tw.
42 (abdom$ adj3 (pressure or pain$ or swelling$ or hard)).tw.
43 (bowel irregularit$ or bloat$ or fullness or satiet$ or gastro$).tw.
44 (fatigue or weight loss$ or weight gain$ or constipat$ or diarrhoea or diarrhea or gas).tw.
45 (nausea$ or indigestion).tw.
46 ((loss or lack) adj3 (energ$ or appetite$)).tw.
47 (urin$ adj3 (frequenc$ or urgenc$)).tw.
48 ((leg$ or ankle$) adj2 (swell$ or swollen)).tw.
49 ((abnormal or irregular or postmenopausal) adj1 vaginal adj (bleed$ or discharge$)).tw.
50 (pelvic discomfort$ or pelvic pain$ or chest pain$ or respirator$ difficult$ or lower back pain$).tw.
51 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50
52 (index or risk$ or score$ or scoring or checklist$ or rule$ or indices or tool$ or instrument$ or survey$ or questionnaire$ or interview$).tw.
53 (LR2 or RMI or ROMA or ADNEX).mp.
54 51 and 52
55 18 or 22 or 36 or 53 or 54
56 14 and 55
57 limit 56 to (humans and yr="2015 - 2019")

Database: Embase (1974 to 21 June 2019)

1 exp Ovary cancer/di
2 exp Adnexal Diseases/di
3 ((borderline or border line) adj4 ovar$).tw.
4 exp uterine cancer/di
5 exp Peritoneum tumor/di
6 exp Pelvis tumor/di
7 ((ovar$ or adnexal or fallopian or peritoneal$ or pelvic) adj3 (cancer$ or carcinoma$ or malignan$ or mass or masses or cyst or cysts or neoplas$ or tumour$ or tumor$)).tw.
8 ((epithelial or germ cell) adj5 ovar$).tw.
9 or/1-8
10 exp ovary cancer/
11 "Neoplasms, Glandular and Epithelial"/
12 exp ovary/
13 10 or 11 or 12
14 9 or 13
15 echography/
16 ultraso$.tw.
17 (transvagina$ adj2 sonogra$).tw.
18 15 or 16 or 17
19 IOTA.tw.
20 International Ovarian Tumor Analysis.tw.
21 ((ovarian or epithelial or adnex$ or fallopian or peritoneal or pelvic) adj3 (model$ or regress$ or rule$ or score$ or algorithm$ or term $ or definition$ or measure$)).ti,ab.
22 19 or 20 or 21
23 *Biological Marker/
24 *Proteomics/
25 *Genetic Marker/
26 *Metabolomics/
27 multiplex$.tw.
28 multivariate.tw.
29 (CA125 or CA-125 or HE4 or OVA 1 or OVA1 or HCG or LDH or AFP).mp. or CEA.tw.
30 CA-125 Antigen/
31 Chorionic Gonadotropin/
32 L-Lactate Dehydrogenase/

33 alpha-Fetoproteins/
34 Carcinoembryonic Antigen/
35 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34
36 symptom$.ti,ab.
37 early diagnosis.tw.
38 (early adj (sign$ or symptom$)).tw.
39 (abdom$ adj3 (pressure or pain$ or swelling$ or hard)).tw.
40 (bowel irregularit$ or bloat$ or fullness or satiet$ or gastro$).tw.
41 (fatigue or weight loss$ or weight gain$ or constipat$ or diarrhoea or diarrhea or gas).tw.
42 (nausea$ or indigestion).tw.
43 ((loss or lack) adj3 (energ$ or appetite$)).tw.
44 (urin$ adj3 (frequenc$ or urgenc$)).tw.
45 ((leg$ or ankle$) adj2 (swell$ or swollen)).tw.
46 ((abnormal or irregular or postmenopausal) adj1 vaginal adj (bleed$ or discharge$)).tw.
47 (pelvic discomfort$ or pelvic pain$ or chest pain$ or respirator$ difficult$ or lower back pain$).tw.
48 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47
49 (index or risk$ or score$ or scoring or checklist$ or rule$ or indices or tool$ or instrument$ or survey$ or questionnaire$ or interview$).tw.
50 (LR2 or RMI or ROMA or ADNEX).mp.
51 48 and 49
52 18 or 22 or 35 or 50 or 51
53 14 and 52
54 limit 53 to (human and yr="2015 - 2019")

Appendix 3. QUADAS-2

DOMAIN 1: PATIENT SELECTION

PATIENT SELECTION	
A. Risk of bias	
Describe methods of patient selection:	
a) Was a consecutive or random sample of patients enrolled?	Yes/No/Unclear
b) Was a case-control design (using healthy controls) avoided?	Yes/No/Unclear
c) Did the study avoid inappropriate exclusions?	Yes/No/Unclear
a) include all ages and regardless of menopausal status or justify restrictions	
b) include all stages of ovarian cancer	
c) include comorbidities such as infertility and endometriosis	
Could the selection of patients have introduced bias?	**RISK: LOW/HIGH/UNCLEAR**
Low: a) and b) and c) 'YES'	
High: a) or b) or c) 'NO'	
Unclear: not 'High' and a) or b) or c) 'UNCLEAR'	

PATIENT SELECTION	
B. Concerns regarding applicability	

(Continued)

Describe included patients (prior testing, presentation, intended use of index test and setting):	
a) Are all or some patients symptomatic	Yes /No/Unclar
b) Prior tests: self-reported symptoms OR self-reported symptoms PLUS one or more of biochemical markers and ultrasound by non-specialist sonographers (in primary or secondary care)	Yes/No/Unclear
Is there concern that the included patients do not match the review question? Low: a) and b) Yes High: a) or b) No Unclear: not High and a) or b) Unclear	**CONCERN: LOW/HIGH/UNCLEAR**

DOMAIN 2: INDEX TEST(S)

(If more than one index test was used, please complete for each test).

INDEX TEST **A. Risk of Bias**	
Describe the index test and how it was conducted and interpreted:	
a) Was the index test or testing strategy result interpreted without knowledge of the results of the reference standard?	Yes / No / Unclear
b) If a threshold was used, was it pre-specified?	Yes / No / Unclear
c) Were all components and thresholds of multivariable models pre-specified before their application?	Yes / No / Unclear
d) Were all components of multivariable models defined and assessed ina similar way for all patients (eg in the same healthcare setting)?	Yes / No / Unclear
Could the conduct or interpretation of the index test have introduced bias? High: a) or b) or c) or d) No Low: a) and b) and c) and d) Yes Unclear: not 'high' and a) or b) or c) or d) Unclear	**RISK: LOW/HIGH/UNCLEAR**

INDEX TEST **B. Concerns regarding applicability**	
a) Was USS performed in all patients by non-specialised sonographers	Yes/No/Unclear

(Continued)

b) Was USS/clinical examination performed with knowledge of symptoms/signs/biomarkers	Yes/No/Unclear
Is there concern that the index test, its conduct or interpretation differ from the review question? High: a) and b) No Low: a) and b) Yes Unclear: a) or b) Unclear	**CONCERN: LOW/HIGH/UNCLEAR**

DOMAIN 3: REFERENCE STANDARD

REFERENCE STANDARD **A. Risk of bias**	
Describe the reference standard and how it was conducted and interpreted:	
a) Were the reference standard results interpreted without knowledge of the index test?	Yes/No/Unclear
b) Is the reference standard likely to correctly classify the target condition? -Index test +ve: Histology following laparoscopy or laparotomy -Index test -ve: Histology following laparoscopy or laparotomy OR clinical follow-up for = > 12 months	Yes/No/Unclear
Could the reference standard, its conduct or its interpretation have introduced bias High: a) or b) No Low: a) and b) Yes Unclear: not 'High' and a) or b) Unclear	**RISK: LOW/HIGH/UNCLEAR**

REFERENCE STANDARD **B. Concerns regarding applicability**	
Can borderline tumours be grouped with primary ovarian cancer for analysis?	Yes/No/Unclear
Can metastatic tumours be disagregated for analysis?	Yes/No/Unclear
Is there concern that the target condition as defined by the reference standard does not match the review question? High: a) and b) No Low: a) and b) Yes	**CONCERN: Yes/No/Unclear**

(Continued)

Unclear: not 'High' and a) or b) Unclear

DOMAIN 4: FLOW AND TIMING

FLOW AND TIMING	
A. Risk of bias	
Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2 × 2 table (refer to study flow diagram):	
Describe the time interval and any interventions between index test(s) and reference standard:	
a) Was there less than 3 months' interval between application of each index test and application of the reference standard?	Yes/No/Unclear
b) Did all patients receive a reference standard?	Yes/No/Unclear
c) Did all index test -ve patients receive the same reference standard?	Yes/No/Unclear
d) Were all patients who underwent testing included in the analysis?	Yes/No/Unclear
Could the patient flow have introduced bias?	**RISK: LOW/HIGH/UNCLEAR**
LOW: a) and b) and c) and d) – Yes	
HIGH: a) or b) or c) or d) – No	
UNCLEAR: not 'high' AND a) or b) or c) or d) – Unclear	

COMPARATIVE DOMAIN (if applicable)

COMPARATIVE DOMAIN	
A. Risk of bias	
Describe the selection process for participants to receive one or other index test or index testing strategy	
Describe the time interval and any interventions between index test(s) for within-person test comparisons	
a) *For studies comparing two or more index tests or testing strategies in **different** patient populations were the selection criteria for participants receiving one or other index test or testing strategy the same?*	*Yes/No/Unclear/NA*
b) *For within-study comparisons of index tests:* *- was the interval between application of each index test < 3 months*	*Yes/No/Unclear/NA*
c) *For within-study comparisons of individual index tests:* *- were index tests interpreted blind to the results of other index test results*	*Yes/No/Unclear/NA*
Could the conduct of the comparative study have introduced bias?	**RISK: LOW/HIGH/UNCLEAR**

(Continued)

LOW: a) OR (b) and c)) – Yes

HIGH: a) OR (b) and c)) – No

UNCLEAR: a) OR (b) or c)) – Unclear

B. Concerns regarding applicability	
Describe included patients (prior testing, presentation, intended use of index test and setting):	
Is there concern that included patients have been selected in a different way to participants in non-comparative studies	**CONCERN: LOW/HIGH/UNCLEAR**

Appendix 4. Tables of excluded studies with reasons

Table 11

Table 12

Table 13

Table 14

Table 15

Table 16

Table 17

Table 18

Table 19

Table 20

Appendix 5. Quality assessment tables for studies grouped by index test

RMI Figure 13

Figure 13. Risk of bias and applicability concerns summary: Risk of Malignancy Index I. Review authors' judgements about each domain for each included study.

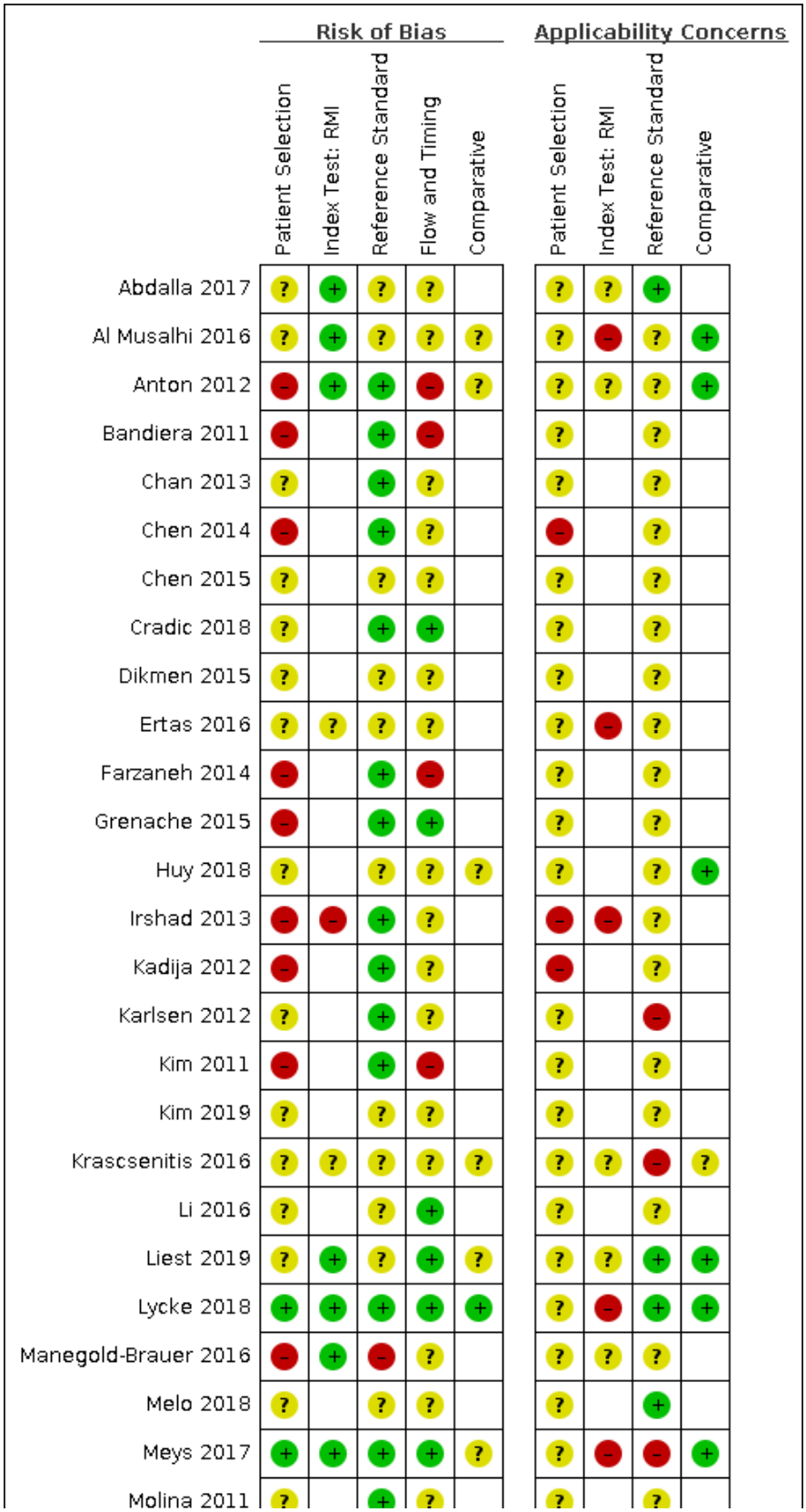


Figure 13. (Continued)

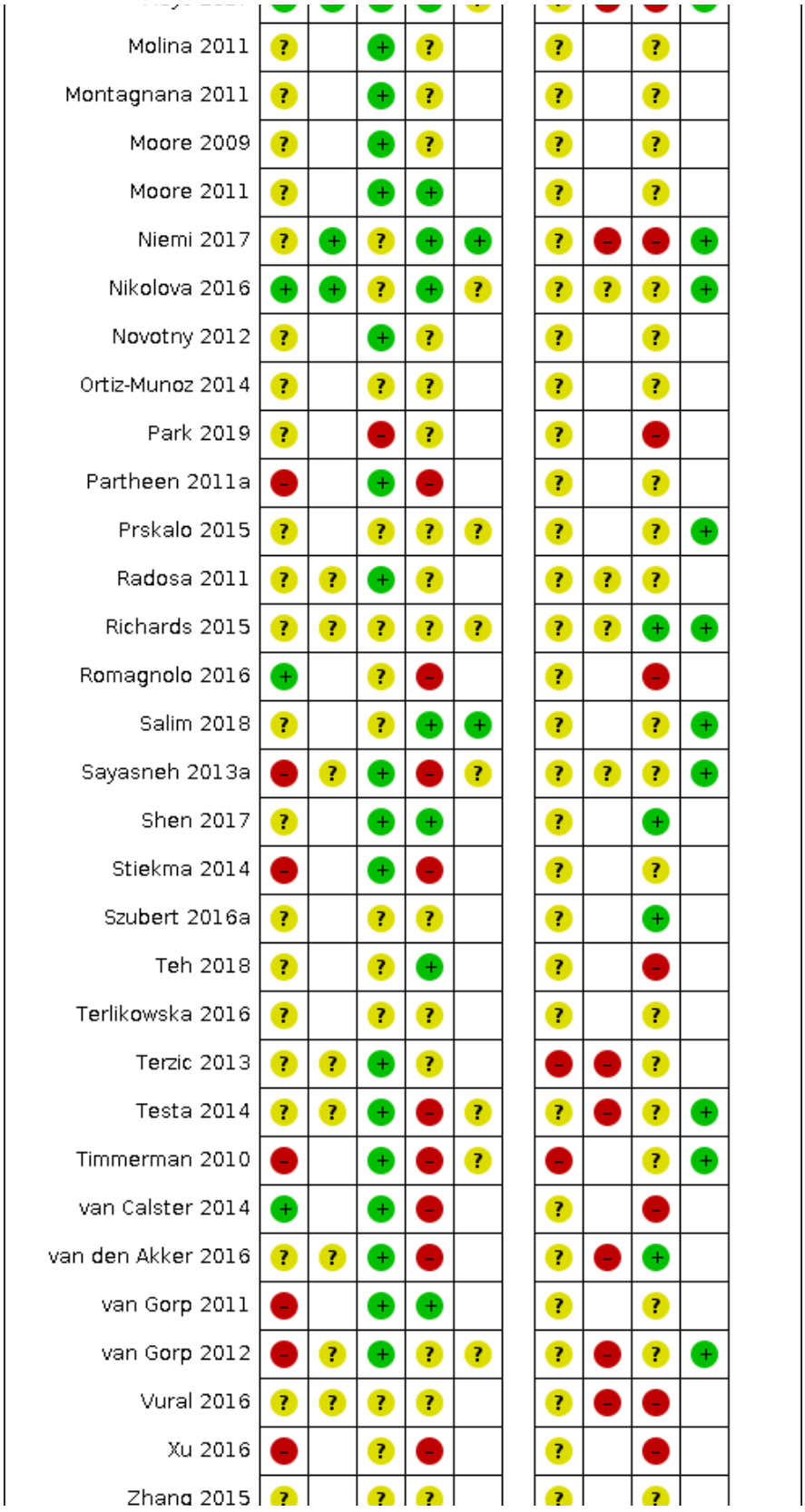
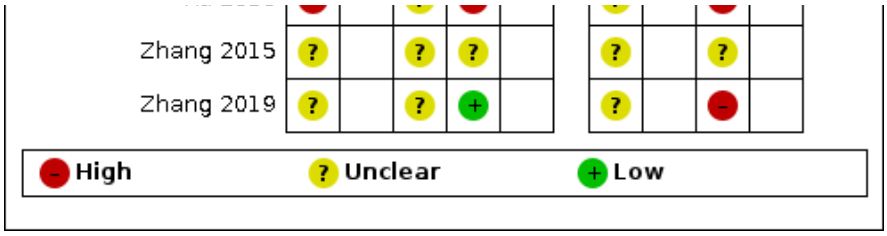


Figure 13. (Continued)



ROMA Figure 14

Figure 14. Risk of bias and applicability concerns summary: Risk of Ovarian Malignancy Algorithm. Review authors' judgements about each domain for each included study.

Figure 14. (Continued)

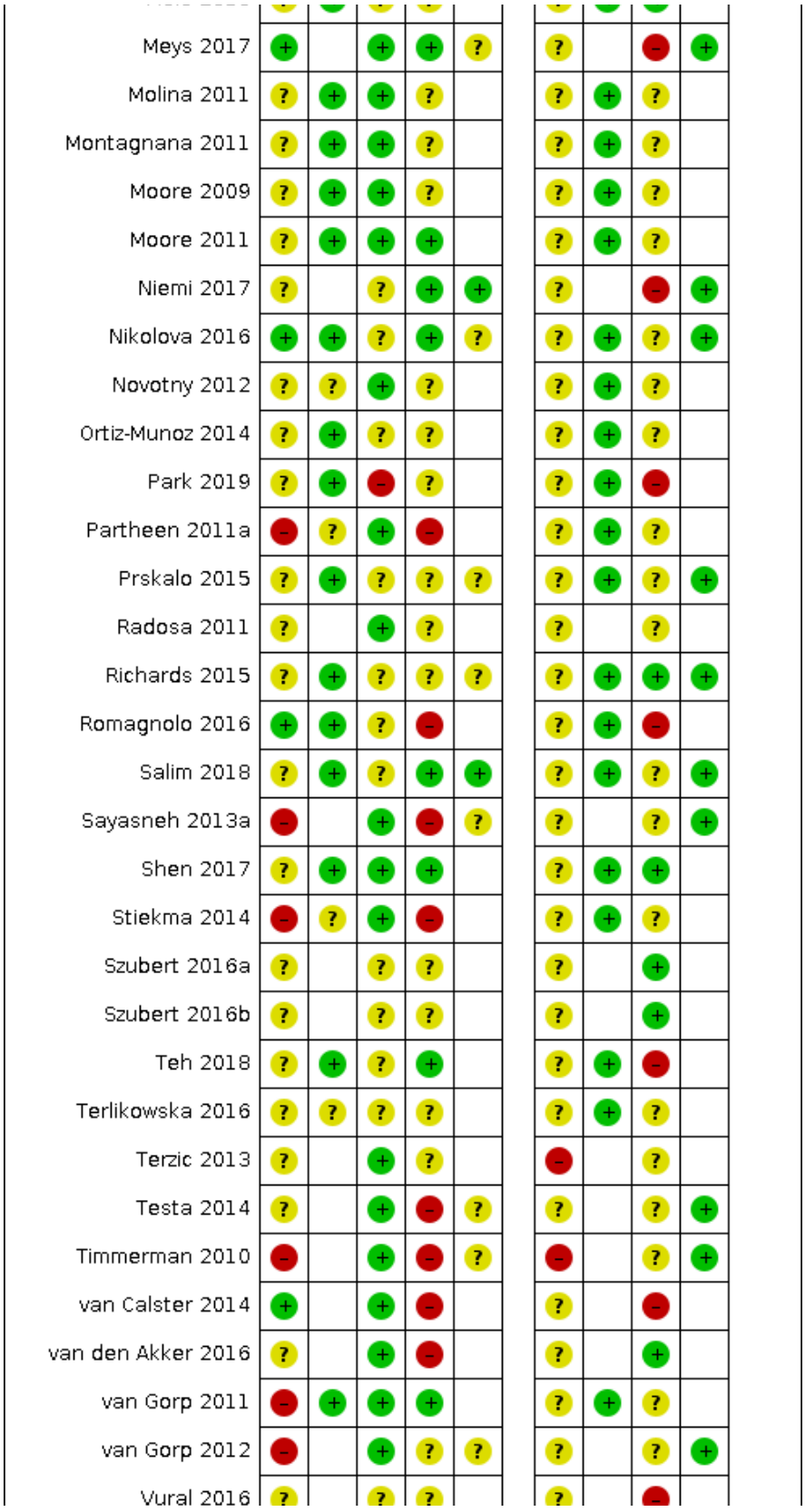
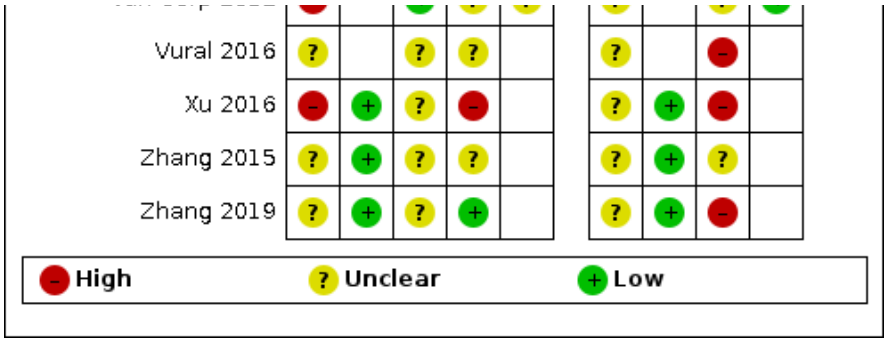


Figure 14. (Continued)



LR2 Figure 15

Figure 15. Risk of bias and applicability concerns summary: Logistic Regression 2 model. Review authors' judgements about each domain for each included study.

Figure 15. (Continued)

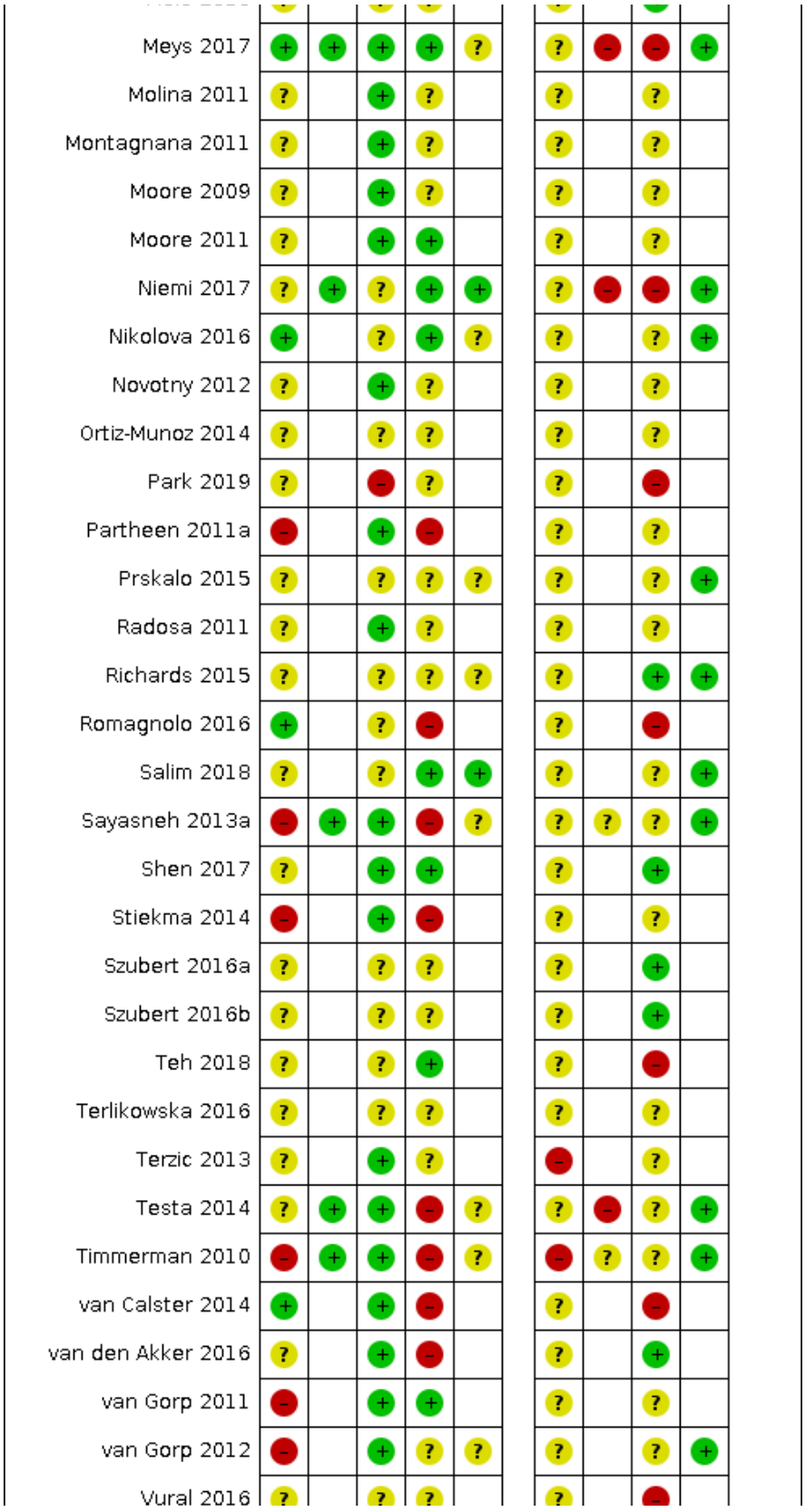
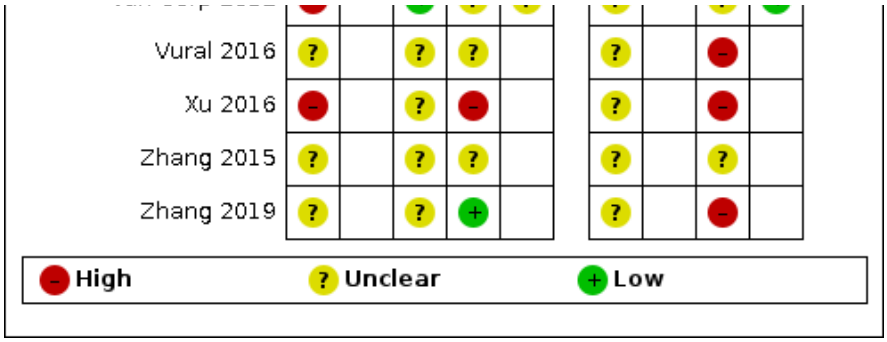


Figure 15. (Continued)



ADNEX Figure 16

Figure 16. Risk of bias and applicability concerns summary: Assessment of Different NEoplasias in the adneXa model. Review authors' judgements about each domain for each included study.

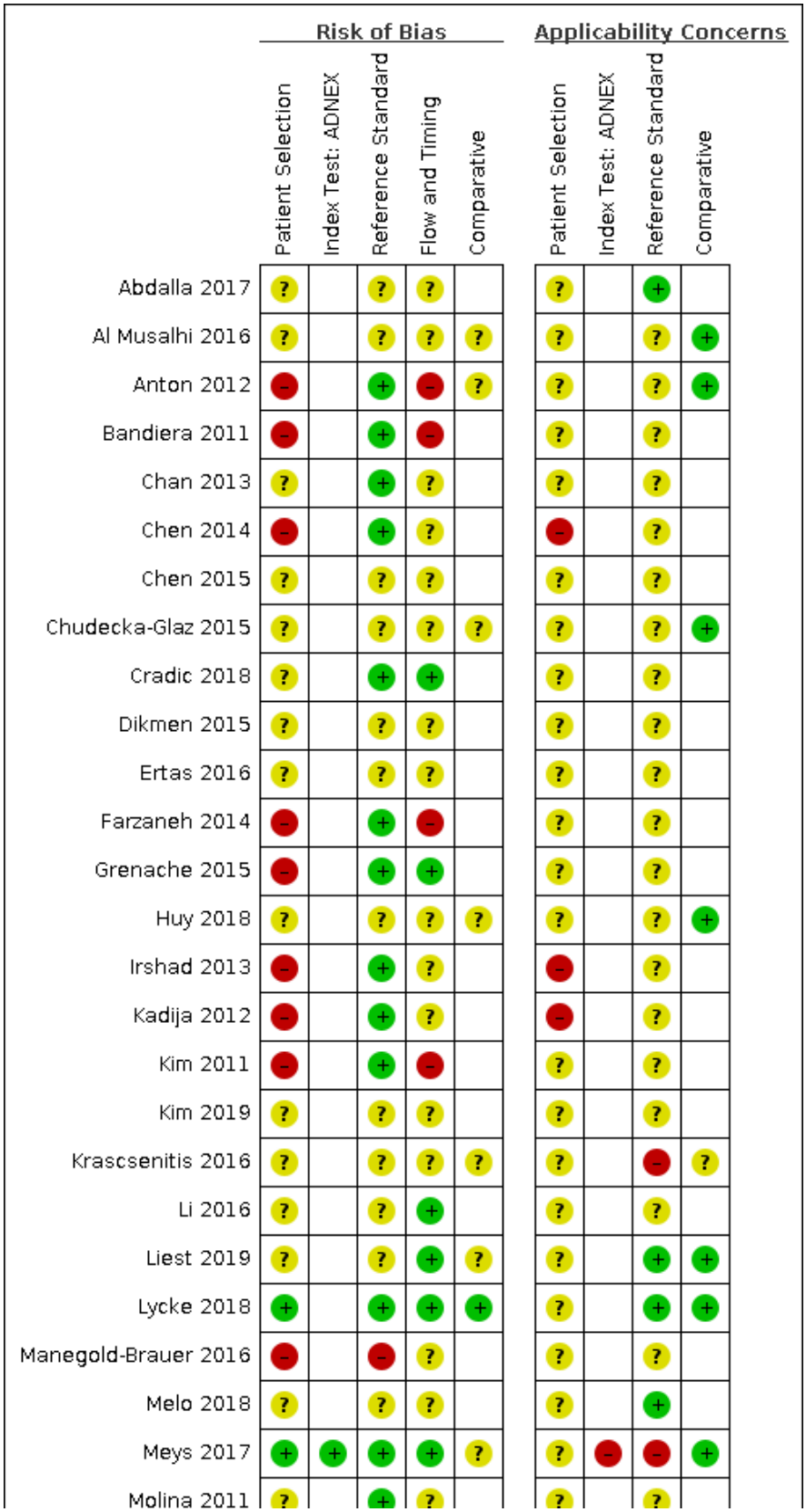


Figure 16. (Continued)

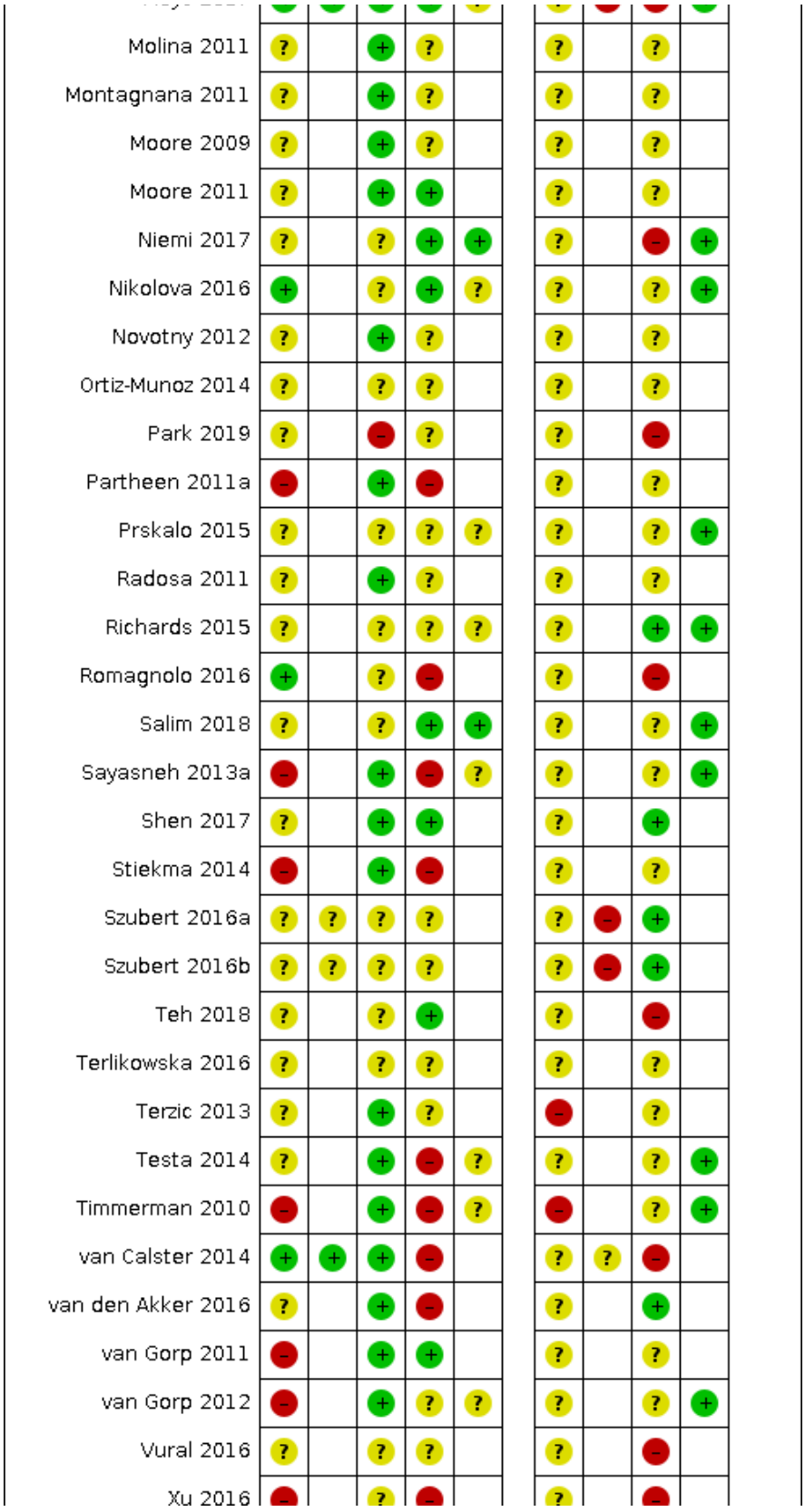
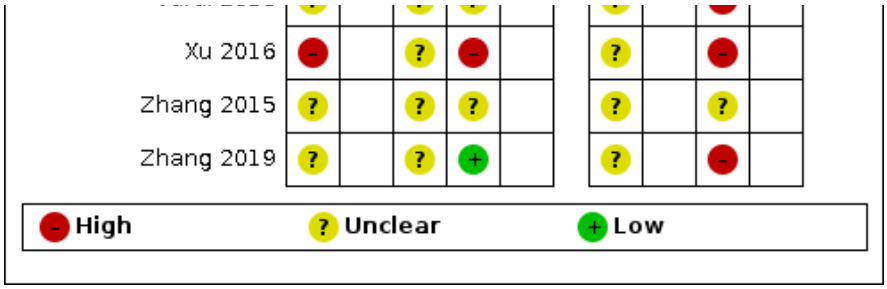


Figure 16. (Continued)



Appendix 6. List of systematic reviews and guidelines included for reference checking

List of systematic reviews and guidelines (25 studies)

1. Multianalyte testing for the evaluation of adnexal masses (Structured abstract). Health Technology Assessment Database [Internet]. 2012 [cited HTA Y/U]; (1). Available from: http://onlinelibrary.wiley.com/o/cochrane/clhta/articles/HTA-32013000454/frame.html.

2. Alcazar JL, Jurado M. Three-dimensional ultrasound for assessing women with gynecological cancer: a systematic review. Gynecologic Oncology. 2011;120(3):340-6.

3. Brun JL, Fritel X, Aubard Y, Borghese B, Bourdel N, Chabbert-Buffet N, et al. Management of presumed benign ovarian tumors: updated French guidelines. European Journal of Obstetrics & Gynecology and Reproductive Biology. 2014;183:52-8.

4. Dodge JE, Covens AL, Lacchetti C, Elit LM, Le T, Devries-Aboud M, et al. Management of a suspicious adnexal mass: a clinical practice guideline. Current Oncology. 2012;19(4):e244-57.

5. Duffy MJ, Bonfrer JM, Kulpa J, Rustin GJS, Soletormos G, Torre GC, et al. CA125 in ovarian cancer: European Group on Tumor Markers guidelines for clinical use. International Journal of Gynecological Cancer. 2005;15(5):679-91.

6. Ebell MH, Culp MB, Radke TJ. A Systematic Review of Symptoms for the Diagnosis of Ovarian Cancer. American Journal of Preventive Medicine.50(3):384-94.

7. Ferraro S, Braga F, Lanzoni M, Boracchi P, Biganzoli EM, Panteghini M. Serum human epididymis protein 4 vs carbohydrate antigen 125 for ovarian cancer diagnosis: a systematic review. Journal of Clinical Pathology. 2013;66(4):273-81.

8. Fischerova D. [Recommended guidelines of diagnosis for women with an ovarian cyst or tumour]. Ceska Gynekologie. 2014;79(6):477-86.

9. Geomini P, Kruitwagen R, Bremer GL, Cnossen J, Mol BW. The accuracy of risk scores in predicting ovarian malignancy: a systematic review. Obstetrics and gynecology. 2009;113(2 Pt 1):384-94.

10. Harris RD, Javitt MC, Glanc P, Brown DL, Dubinsky T, Harisinghani MG, et al. ACR Appropriateness Criteria clinically suspected adnexal mass. Ultrasound Quarterly. 2013;29(1):79-86.

11. Hayes, Inc. Ca 125 for ovarian cancer screening in average-risk women (Structured abstract). Health Technology Assessment Database [Internet]. 2005 [cited PENDING (UPDATE 91-08)- FT NOT FOUND (2). Available from: http://onlinelibrary.wiley.com/o/cochrane/clhta/articles/HTA-32006000089/frame.html.

12. Kaijser J, Sayasneh A, van Hoorde K, Ghaem-Maghami S, Bourne T, Timmerman D, et al. Presurgical diagnosis of adnexal tumours using mathematical models and scoring systems: a systematic review and meta-analysis. Human Reproduction Update. 2014;20(3):449-62.

13. Karlsen NS, Karlsen MA, Hogdall CK, Hogdall EVS. HE4 tissue expression and serum HE4 levels in healthy individuals and patients with benign or malignant tumors: a systematic review. Cancer Epidemiology, Biomarkers & Prevention. 2014;23(11):2285-95.

(Continued)
14. Kinkel K, Hricak H, Lu Y, Tsuda K, Filly RA. US characterization of ovarian masses: a meta-analysis (Structured abstract). Radiology [Internet]. 2000 [cited DARE Y/U]; (3):[803-11 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/cldare/articles/DARE-12000002350/frame.html.

15. Kinkel K, Lu Y, Mehdizade A, Pelte M-F, Hricak H. Indeterminate ovarian mass at US: incremental value of second imaging test for characterization - meta-analysis and Bayesian analysis. Radiology. 2005;236(1):85-94.

16. Le T, Giede C, Salem S, Lefebvre G, Rosen B, Bentley J, et al. Initial evaluation and referral guidelines for management of pelvic/ovarian masses. Journal of Obstetrics & Gynaecology Canada: JOGC. 2009;31(7):668-80.

17. Li F, Tie R, Chang K, Wang F, Deng S, Lu W, et al. Does risk for ovarian malignancy algorithm excel human epididymis protein 4 and CA125 in predicting epithelial ovarian cancer: a meta-analysis. BMC Cancer. 2012;12:258.

18. Lin J, Qin J, Sangvatanakul V. Human epididymis protein 4 for differential diagnosis between benign gynecologic disease and ovarian cancer: a systematic review and meta-analysis. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2013;167(1):81-5.

19. Mol BW, Boll D, De Kanter M, Heintz AP, Sijmons EA, Oei SG, et al. Distinguishing the benign and malignant adnexal mass: an external validation of prognostic models. Gynecologic Oncology. 2001;80(2):162-7.

20. Nunes N, Ambler G, Foo X, Naftalin J, Widschwendter M, Jurkovic D. Use of IOTA simple rules for diagnosis of ovarian cancer: meta-analysis. Ultrasound in Obstetrics & Gynecology. 2014;44(5):503-14.

21. Reed N, Millan D, Verheijen R, Castiglione M, Group EGW. Non-epithelial ovarian cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of Oncology. 2010;21 Suppl 5:v31-6.

22. Wang J, Gao J, Yao H, Wu Z, Wang M, Qi J. Diagnostic accuracy of serum HE4, CA125 and ROMA in patients with ovarian cancer: a meta-analysis (Provisional abstract). Tumor Biology [Internet]. 2014 [cited DARE N]; (6):[6127-38 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/cldare/articles/DARE-12014021188/frame.html.

23. Yang Z, Wei C, Luo Z, Li L. Clinical value of serum human epididymis protein 4 assay in the diagnosis of ovarian cancer: a meta-analysis.[Erratum appears in Onco Targets Ther. 2014;7:135]. OncoTargets and therapy. 2013;6:957-66.

24. Zhe S, Jun Z, Huamei M, Qiuli C, Minlian D, Yanhong L, et al. Four cases of ovarian adrenal rest tumors in Chinese girls with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. Hormone Research in Paediatrics Conference: 53rd Annual Meeting of the European Society for Paediatric Endocrinology, ESPE. 2014;82(pp 331).

25. Zhen S, Bian LH, Chang LL, Gao X. Comparison of serum human epididymis protein 4 and carbohydrate antigen 125 as markers in ovarian cancer: A meta-analysis. Molecular and Clinical Oncology. 2014;2(4):559-66.

HISTORY

Protocol first published: Issue 12, 2015

CONTRIBUTIONS OF AUTHORS

- Guarantor of the review: SS, JD, CD.
- Conceiving the idea: SS, CD, JD.
- Designing and co-ordinating the review: NR, CD, SS, JD.
- Designing search strategies: SB, NR, CD, SS, RN.
- Screening, data extraction, quality assessment: NR, RC, CD, PSh, PSa.
- Obtaining and screening data on unpublished studies: NR, RC, PSh, PSa.
- Data management of the review: NR, PSh, PSa, CD.
- Analysis and interpretation of data: SM, KS, NR, SS, CD, JD.
- Writing the review: CD, NR, SS, PSh.
- Providing general advice on the review: CD, SS, JD.
- Securing funding for the review: SS, CD, JD.

DECLARATIONS OF INTEREST

This review and participation of all authors in it has been funded as part of a programme of research (ROCkeTS – Refining Ovarian Cancer Test Accuracy Scores).

CD: received funding from the NIHR HTA to support this review in a methodological capacity.

NR: my participation in this review is funded by the NIHR grant listed.

PSh: none known.

JD: this work is a funded project, funded by the NIHR HTA Commissioning Board.

SB: none known.

SM: co-applicant on one funded government funded grant (mpMRI imaging) and one recently submitted grant (circulating DNA) for the diagnosis of ovarian cancer.

PSa: none known

RC: none known.

SB: none known.

KS: my participation in this review is funded by the NIHR grant listed.

SS: none known.

SOURCES OF SUPPORT

Internal sources

- None, Other

External sources

- National Institute for Health Research (HTA programme: 13/13/01), UK

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

Search strategy

We did not restrict our searches to English Language publications but we were unable to consider non-English publications due to time and resource limitations. The volume of non-English publications not considered by this review is explicit in the results of the search strategy. For pragmatic reasons, we conducted searches for the period 2015 to 2019 in a restricted number of bibliographic databases. We did not search the US Food and Drug Administration (FDA) and European Medicines Agency (EMA) websites for the period 2015 to 2019 as part of the search update for this test combination review; these literature resources were originally checked in 2015 as part of a generic protocol covering four reviews, specifically for a review of biomarkers for the diagnosis of ovarian cancer (OC).

Type of studies

Case-control studies where healthy controls could not be disaggregated from women with benign ovarian pathology were excluded. Studies concerned only with the development of multivariable models were excluded. Where papers reported data on both the development and validations of a multivariable model, we extracted only the validation data.

Index test

We did not include all thresholds reported in each study. For each index test version within an individual study we extracted up to four thresholds. We prioritised extraction of results in the following order: 1. from prespecified thresholds, 2. thresholds commonly used in clinical guidelines, 3. thresholds commonly used in the published literature and 4. thresholds reported as main outcomes in studies included in this review.

Target condition

We excluded studies reporting exclusively on metastatic disease to the ovary or recurrent OC. We disaggregated data to exclude cancers metastatic to the ovary and recurrent OCs in studies where possible; studies where the these data were unavailable or the information was available but could not be disaggregated was downgraded as unclear or high, respectively, for reference standard applicability

Data extraction

A single review author (NR or PSh or PSa) extracted study characteristic data and a second review author (RC) independently checked 30% of studies. Any differences were resolved by discussion.

A single review author (NR or PSh or PSa) extracted methodological quality data, and a second review author (RC) independently checked 30% of studies. Any differences were resolved by discussion.

Quality assessment

A separate domain for multivariable models was not considered necessary, particularly as we did not include studies only reporting development of multivariate models. Instead, we added two questions to the participant domain of QUADAS-2 drawing on the PROBAST (prediction model risk of bias assessment) tool for diagnostic and prediction models (Wolf 2019): 1. Prespecification of thresholds and 2. comparable assessment of all model/test components.

Statistical analysis

We compared test accuracy in pre- and postmenopausal women by adding a covariate in the bivariate model and calculating differences and 95% confidence intervals using non-linear estimating methods, taking advantage of advances in analysis methods compared to simple testing of differences using likelihood ratio tests. We presented the impact of using tests and test comparisons using absolute numbers of average women in a hypothetical population at a range of clinically relevant prevalence, representative of primary care and a range of specialist settings instead of restricting to a single prevalence representative of a primary care setting. This approach was adopted to illustrate the clinical utility of index tests in multiple settings, reflecting their potential use in clinical practice.

Heterogeneity and sensitivity analyses

We were unable to carry out the following planned heterogeneity analyses due to insufficient studies with differences in the relevant study characteristics or with these study characteristics reported: generalist (primary care/community/family practice) versus specialist setting (cancer unit/cancer centre/gynaecological oncology); histological subtype, reference standard QUADAS-2 domain risk of bias (high/unclear versus low); case-control study versus other study designs; 12 months' follow-up versus less than 12 months' follow-up for study participants not receiving surgery initially following a negative index test result.

We did not carry out sensitivity analyses leaving out highly influential studies as this was not considered necessary; including only studies with low concern about applicability in the patient selection domain of QUADAS-2 as there were insufficient studies; or classification of borderline tumours as malignant or benign as this proved too simple an approach given the heterogeneity in approach to management and reporting of borderline tumours in included studies. Instead, where data allowed, we compared estimates of the test accuracy of each index test for studies using an inappropriate grouping (studies excluding borderline ovarian tumours and studies where the management of borderline ovarian tumours was unclear) with studies using an appropriate grouping (studies combining borderline ovarian tumours with malignant ovarian tumours) using the hierarchical summary receiver operating characteristic (HSROC) model.

INDEX TERMS

Medical Subject Headings (MeSH)

Biomarkers; Carcinoma, Ovarian Epithelial; Cross-Sectional Studies; Menopause; *Ovarian Neoplasms [diagnostic imaging]; Sensitivity and Specificity

MeSH check words

Female; Humans