Biochemical markers and combination testing for the diagnosis of ovarian cancer in women with symptoms or signs suspicious of ovarian cancer

By

Dr. Nirmala Rai Talapadi

A thesis submitted to the University of Birmingham

for the degree of

Doctor of Medicine (MD)

School of Cancer Sciences

University of Birmingham

March 2020

ABSTRACT

Ovarian cancer (OC) has the highest mortality of all gynaecological cancers. A significant contributing factor to the high mortality in OC is delayed diagnosis. Currently, there is no consensus regarding the best test for early diagnosis.

A review of existing systematic reviews about symptoms, biochemical markers and US test used alone or in combination for the diagnosis of OC in symptomatic women demonstrated that existing reviews were variable in quality, applicability and limited by poor reporting. I attempted to address these deficiencies in two reviews on the accuracy of biomarkers alone and symptoms, biomarkers or US in combination for the diagnosis of OC in symptomatic women in generalist settings in pre and postmenopausal women separately.

My thesis finds key methodological issues, e.g., literature is not applicable to generalist settings as studies included women typical of tertiary healthcare settings, some studies excluded borderline tumours which inflates estimates of sensitivity, important differences exist in test performance between pre and postmenopausal women. Main results are 1) reviews not applicable to primary care settings – more research is needed. 2) for biomarkers i) HE4 at the threshold of 60-80pMol/L and 130-150pMol/L is recommended in pre and postmenopausal women for low prevalence settings ii) ROMA or LR2 in premenopausal women to replace RMI in secondary/tertiary setting; continue with RMI for postmenopausal women as it shows comparable accuracy to ROMA and LR2

ACKNOWLEDGEMENTS

This MD was funded by NIHR, without which I would not have been able to complete this research.

I would like to express my gratitude to my supervisors, Professor. S. Sundhar and Dr C. Davenport, I am glad that I had the opportunity to work with you. You have built up my confidence and propped me up when I was down – which was more than a few times! I cannot thank you enough for the support and guidance you have offered.

I would also like to thank the following people I have been lucky enough to work with: Professor. J. Deeks, Dr. S. Mallet and Dr. Y. Takwoingi who have been generous with their time and advice and also supported me during the course of my work.

I would also like to thank all staff in test accuracy team and Cochrane who have helped, supported and contributed to my work: S. Bayliss, R. Champanaria, A. Walker and S. Stevens who have helped me with the systematic review.

To my family, thank you for all your love and support over the years.

Finally, to all my friends – and there are too many to name here - I am so lucky to have you all in my life and for believing in me. Without your love, concern and good wishes, this journey would not have been possible. I cannot thank you all enough.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ADNEX	Assessment of Different NEoplasias in the adneX
AMSTAR	Assessment of Multiple SysTemAtic Review
ASR	Age standardised rate
AFP	Alpha feto protein
ACOG	American College of Obstetricians and Gynaecologists
ASCO	American Society of Clinical Oncology
BOTS	Borderline ovarian tumours
CA125	Cancer Antigen 125
CEA	Carcinoembryonic Antigen
CI	Confidence Interval
DOvE	Diagnosing ovarian cancer early
EFSUMB	European Federation of Societies for Ultrasound in Medicine and Biology
DTA	Diagnostic accuracy test
EBOTs	Endometrioid borderline ovarian tumours
EOCs	Epithelial ovarian cancers
ESGO	European Society of Gynaecological Oncology
FDA	U.S. Food and Drugs administration
FOC	Familial Ovarian cancer
FN	False negative
FP	False positive
GCT	Germ cell tumour
HCG	Human chorionic gonadotrophin
HE4	Human Epididymis protein 4
HGSOC	High grade serous ovarian cancer
HNPCC	Hereditary nonpolyposis colorectal cancer
HSROC	Hierarchical Summary Receiver Operating Characteristic
ICTRP	WHO International Clinical Trials Registry Platform
IGCS	International Gynaecologic Cancer Society
IOTA	International ovarian tumour analysis
LDH	Lactate dehydrogenase
LGSOC	Low grade serous ovarian cancer

LR	Logistic Regression (1 and 2)
MBOTS	Mucinous borderline ovarian tumours
MMT	Mixed mesodermal tumours
NICE	National Institute of Clinical Excellence
OC	Ovarian cancer
PPC	Primary peritoneal cancer
PROBAST	Prediction model risk of bias assessment
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCOG	Royal College of Obstetricians and Gynaecologists
RMI	Risk of Malignancy Index (1,2,3 and 4)
ROB	Risk of Bias
ROC	Receiver operating characteristic
ROMA	Risk of malignancy algorithm
RRBSO	Risk reducing bilateral Salpingo-oophorectomy
SAS	Statistical analysis system
SBOTs	Serous borderline ovarian tumours
SN	Sensitivity
SP	Specificity
SGO	Society of Gynecologic Oncology
TN	True negative
TP	True positive
SR	Systematic reviews
QUADAS	Quality assessment of diagnostic accuracy studies
US	Ultrasound
UKCRN	UK Clinical Research Network

CHAPTER 1: INTRODUCTION

Ovarian tumours are a group of heterogeneous masses that develop in the ovary. Worldwide about 7% of women have ovarian masses in their lifetime,[2] with up to 10% of women having some form of surgery for an ovarian mass.[3] The epidemiology of ovarian masses is unclear due to inconsistent reporting and spontaneous resolution of many ovarian masses. Causes of ovarian masses range from normal physiological processes to abnormal growth of cells leading to malignancy. They are commonly classified into 3 main subcategories based on their malignant potential: 1) benign; 2) borderline, and 3) malignant tumours. They can occur at any age, but 2/3rd of malignancy occurs in the 40-65 years age group.[4]

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 cysts are caused by benign conditions such as endometriosis, fibromas and cystadenomas which may require surgical intervention; however, the risk of malignancy is low, 1/1000.[3] The risk of malignancy increases in postmenopausal women, 3/1000 at age 50, [3] and is highest in women aged 55-64 years, median age 63 years.[5]

Borderline ovarian cysts are tumours of low malignant potential which are more common in the reproductive age group and typically affect women in their 40s. The incidence of borderline ovarian tumours (BOTs) is low, 1.5 - 2.5 per 100,000 women per year[6] with approximately 30% of these appearing in women younger than 40 years.[7]

A small proportion of ovarian masses can be malignant and are responsible for high case fatality ratio. Ovarian cancers can arise *de novo* or in the background of borderline or benign ovarian tumours.

Asymptomatic smaller benign tumours can be managed conservatively. Surgical management is recommended in symptomatic ovarian tumours, borderline and malignant ovarian tumours; the route, extent and place of surgery are guided by the diagnosis and the need for fertility preservation. Unfortunately, there is no non-invasive method to make a definitive preoperative diagnosis, and optimal management is guided by diagnostic tests providing risk thresholds that allow women classified as ‘low risk’ to be managed in secondary care with fertility preserving and less radical treatment options including laparoscopic surgery or surveillance whilst triaging

women at 'high-risk' for treatment at a cancer specialist centre for appropriate surgery as required.

1.1 Epidemiology of ovarian cancer

1.1.1 Incidence and mortality

Ovarian cancer (OC) is the 8th most common cancer in the women worldwide, and nearly 300,000 new cases were diagnosed in 2018.[8, 9] Incidence rates are lower in less developed countries, Sub-Saharan Africa and Asia, 5 and 5.8 ASR /100,000 respectively and highest in North America and Europe, exceeding 8.4 and 9.5 per 100,000 respectively.[8, 9] It is the commonest cause of death from gynaecological malignancies in the developed world. OC is the eighth most common cause of death in women with 184,799 deaths in 2018.[9]

In the United Kingdom, 7270 new cases in 2015 and 4227 deaths in 2016 from OC were reported.[10] In the United States 22,240 new cases and 14,070 deaths per year per 100,000 women were reported in 2018.[5, 11, 12] The overall survival for OC is poor as nearly 2/3rd of the disease are diagnosed at an advanced stage.

Age-Standardised rates (ASR) for OC for one, five and ten year net survival in women (aged 15 - 99), in England and Wales for 2010 - 2011 are 72.4, 46.2 and 34.5 respectively. Upper control limit for the ASR for one, five and ten year survival for OC are 72.5, 46.4 and 35.3 respectively and lower control limit are 72.4, 45.9 and 33.8 respectively.

1.1.2 Distribution

The incidence of OC increases with age, with a median age of diagnosis at 63 years; more than 75% of OC is diagnosed in women aged 55 years and over.[10, 13] The majority of ovarian cancers are sporadic, but 10-15% of the ovarian cancers are due to inherited mutations in cancer susceptibility genes. OC in women with BRCA mutations tends to develop at a younger age, approximately a decade earlier in comparison to sporadic OC.[14] BRCA mutations are more common in Ashkenazi Jews compared to general population.[15, 16]

1.2 Hypotheses of development of OC

Ovarian cancer development is now categorised into 2 broad categories based on molecular genetic pathways. Type I or low-grade indolent tumours develop from a precursor lesion and show step wise growth from benign, borderline into malignancy; this includes low grade serous OC (LGSOC), endometrioid cancers, clear cell cancer, mucinous cancers and Brenner tumours. Type II or high-grade aggressive tumours arise *de novo* without precursor lesions; high-grade serous OC (HGSOC) being the most common and others including undifferentiated and malignant mixed mesodermal tumours/carcinosarcoma (MMT). The low-grade tumours demonstrate a different genetic profile demonstrating more BRAS, KRAF, β-catenin and PTEN mutations in comparison to HGSOC that demonstrate ubiquitous p53 mutations (Table 1.1). Further morphological, molecular and genetic advancements have shed further light to explain the diversity seen in ovarian tumours. Piek et al. in 2003 described lesions in the fallopian tube that closely resembled HGSOC[17] which was substantiated in subsequent studies.[17-19] It is widely accepted now that serous ovarian tumours arise from the fallopian tube, whereas endometrioid and clear cell tumours arise from endometriosis and mucinous and Brenner tumours arise from intestinal/tubo-mesothelial junction.[20-23]

Table 1. 1: Precursors and molecular genetic alterations of type I and II tumours of the ovary.

Type I tumours	Precursors*	Known molecular genetic alterations
Low grade serous ca	Serous cystadenoma / adenofibroma Atypical proliferative serous tumour Non-invasive low-grade serous ca	67% - BRAF and KRAS mutations
Mucinous ca	Mucinous cystadenoma Atypical proliferative mucinous tumour Intraepithelial ca	>60% - KRAS mutations
Endometrioid ca	Endometriosis Endometrioid adenofibroma Atypical proliferative endometrioid tumour Intraepithelial ca	20% - LOH or PTEN mutations 16-54% β-catenin mutations 4-5 % KRAS mutations 13-50% - microsatellite instability
Clear cell ca	Endometriosis Clear cell adenofibroma Atypical proliferative clear cell tumour Intraepithelial ca	5- 16% - KRAS mutations ≈ 13% - Microsatellite 66% - TGF- β RII instability
Malignant Brenner tumour	Brenner tumour Atypical proliferative Brenner tumour	Unidentified yet
Type II tumours	Precursors	Known molecular genetic alterations
High grade serous ca	Unidentified yet	50-80% - p53 mutations 10-20% - amplification and overexpression ofHER2/neu gene 10-17% - Inactivation of p16 gene
Undifferentiated ca	Unidentified yet	Unidentified yet
Malignant mixed mesodermal tumour	Unidentified yet	>90% - p53 mutations

*atypical proliferative and non-invasive tumours are also called BOTS

1.3 Histological subtypes, molecular, cytogenetics and clinical features of ovarian tumours

1.3.1 Benign ovarian tumours

Ovarian cysts are fluid or solid filled pockets that can be physiological, benign or malignant. They can develop at all ages and stages of life. The increased use of imaging has increased the rate of detection of adnexal masses,[24] many of which are asymptomatic cysts that are largely physiological or benign. Most cysts that appear during the reproductive years are due to hormonal activity and disappear on their own. A nested case control study in USA noted simple

cysts in 23.8% and 13.4% in women <50 and >50 years respectively.[25] Benign ovarian cysts are broadly classified into functional, inflammatory/infectious and benign neoplastic tumours.

Functional ovarian cysts

The pituitary gonadotrophin menstrual feedback pathway regulates the ovarian function. The ovaries form follicles monthly, one of which ruptures to release an egg and form the corpus luteal cyst. These follicular cysts can persist or get large due to hormonal imbalance (polycystic ovaries) and or normal or increased hormonal activity.[26] They are usually asymptomatic and resolve spontaneously unless complicated by rupture, torsion or haemorrhage that can present with symptoms of abdominal pain, distension nausea or vomiting or can be picked up incidentally on ultrasound, commonly as simple cysts.

Follicular cysts can develop in response to intrinsic or extrinsic gonadotrophic stimulation. However, it is unclear if the persistence is either due to failure of an immature follicle to undergo atresia or failure of a dominant follicle to rupture.[27] They are typically 2 - 3 cm and not usually larger than 5 cm. Corpus luteal cysts develop following ovulation and cease to function by undergoing luteolysis at the end of a nonfertilised menstrual cycle.

It can persist as corpus luteum of pregnancy if fertilised. It can grow to a larger size and can rupture either in pregnant / non-pregnant states, causing pain and bleeding. Theca lutein cysts develop due to hypertrophy of theca lutein cells or luteinised granulosa cells or in response to excessive human chorionic gonadotrophin in gestational trophoblastic disease, multiple pregnancies or exogenous hormonal stimulation. Theca lutein cysts are often bilateral and occasionally increase to a very large size called hyperreactio luteinalis.[28] However most of these cysts are unilocular and do not grow over 7 cm.[29] Many of these are seen on ultrasound as simple cysts and unless symptomatic need no treatment.[25]

Inflammatory/infectious ovarian cysts

Inflammatory and infective adnexal masses may also present as complex ovarian masses. They often present acutely with pain and signs of infection and raised inflammatory markers; however, they can have a chronic presentation, e.g. tuberculosis and actinomycosis where the symptoms and signs of acute inflammation and infection are not present. Instead, they could be incidentally detected or have slow, insidious symptoms similar to ovarian cancer such as

abdominal pain, distension, loss of appetite and failure to thrive. They are clinically difficult to distinguish from ovarian malignancies as they have similar symptoms, appear as complex cysts +/- free fluid +/- peritoneal implants on USS and may have associated raised CA125.

Benign neoplastic ovarian tumours

Ovarian neoplastic tumours can grow to an excessive size and have an abnormal growth without showing dysplasia. They are also a group of heterologous tumours due to the pluripotent capacity of ovarian tissue and vary in histology as given in Table 1.2. Nearly 60% of all ovarian tumours are benign.[30] The normal growth control is dysregulated, leading to uncontrolled growth of ovarian tissue but without the capacity to breach stroma or invade other tissues.[31] The most common benign tumours are serous and mucinous cystadenomas.[32] They can also be detected incidentally on ultrasound (US) as they can grow to a large size without causing symptoms or present with symptoms of abdominal pain (due to haemorrhage, rupture, infection, torsion), abdominal distension, vomiting (due to torsion) or urinary or bowel symptoms (due to pressure). They can be seen on US as simple or complex cysts. Occasionally, these cysts have associated solid components or free fluid (which are features seen more with malignancy) due to ovarian accident and inflammation.

Benign neoplastic ovarian tumours can be broadly classified into epithelial, germ cell and stromal cell tumour depending on the tissue of origin. Germ cell tumours are most commonly seen in the adolescent age group, with peak incidence at 18 years. Sex cord tumours are often seen in the 4th and 5th decade and can be associated with feminising or virilising signs due to hormonal production. The various subtypes are detailed in Table 1.2 and Table 1.3.

Epithelial cell tumours

Serous cystadenoma and adenofibroma and cyst adenofibromas: Benign serous tumours account for 53.39% and the commonest of the benign tumours. The peak incidence is in the 3rd, 4th and 5th decades with increasing incidence with advanced age. 12-23% are bilateral, may be unilocular or multilocular, and vary in size from 3-30 cm. The surface may be smooth or may have papillary excrescences called papilloma either in the inner or outer surface. Unlike the malignant counterparts, they do not show surface or peritoneal invasion. Benign serous cystadenomas do not show KRAS and BRAF mutations unlike borderline and low-grade serous tumours.[32]

Mucinous cystadenoma: Mucinous cystadenomas account for 41% of benign ovarian tumours and are also common in the 3rd – 5th decade.[33] They are usually unilateral (only 5% bilateral), loculated, have a smooth surface and often grow to large sizes asymptomatically, sometimes filling the abdomen and pelvis.[34] Complex mucinous tumours may have multiple loculi and daughter cysts, giving appearances of pseudoinvasion. They also may contain florid and complex glandular and papillary arrangement, making invasion very difficult to rule out and posing histological challenges.

Endometrioid tumours: Ovarian cysts can also contain cysts that have walls lined by glandular cells similar to the endometrium and are called endometriomas or chocolate cysts. It is a benign oestrogen-dependent tumour that is caused by endometriosis, which is a condition characterised by growth of endometrial tissue outside the uterine cavity. Endometriomas contain old degenerated blood that oozes dark brown coloured fluid when the cyst is opened or bursts. Endometriosis has a strong correlation with history of infertility and nulliparity. Cyst walls are often thick and fibrotic and densely adherent to other structures. They are multiloculated and anechoic, having solid components on imaging due to fluid blood debris or old clot, making them difficult to distinguish from a malignant tumour. Patients also often have raised CA125 due to the associated inflammation. Despite classification as a benign disease, there is mounting evidence that endometriomas have a potential to become malignant,[35] with associations with clear cell and endometrioid cancers.[36]

Clear cell cystadenomas and adenofibromas: These are rare epithelial cell tumours that may contain several cell types; clear cells and hobnail cells being the most common.[37] Initially, they have been thought to arise from the mesonephric rest cells of the female reproductive system, and later others have suggested a Mullerian origin.[38] The benign tumours tend to be relatively more solid than their malignant counterparts.

Serous cell tends to be the most common epithelial component in cystadenomas and adenofibromas; less commonly they may have endometrioid, mucinous, clear call or mixed cell type[39] and careful histological examination is necessary to enable correct identification.

Brenner tumour: Brenner tumours are a rare, solid ovarian tumours mostly occurring in postmenopausal women. The majority of Brenner tumours are benign and less than 5% of are malignant.[40] They arise from ovarian surface epithelium or the pelvic mesothelium by transitional cell metaplasia.[41] They are well circumscribed tumours with a grey or slightly

yellow surface and hard or fibromatous appearance. They comprise approximately 3% of neoplasms and are asymptomatic and often discovered incidentally.[34]

Germ cell tumours

Germ cell tumours (GCT) are derived from the primordial cells of the ovary and, as the name suggests, contain all 3 embryonic germ cell layers ectoderm, endoderm, and mesoderm. They commonly occur around 10-30 years of age, and comprise 10 - 25% of ovarian tumours, but only 5% are malignant.[42] Malignancy is more common in the childhood and adolescent age group and GCTs contribute 70% of ovarian malignancy in this group.[43] They include dysgerminomas, teratomas, embryonal tumours and endodermal sinus tumour.

All the germ cell tumours are malignant except for teratomas, which have a benign counterpart called dermoid cysts or mature cystic teratomas. Teratomas usually form unilocular cysts lined by squamous epithelium and containing hair follicles and sebaceous material. They may also contain teeth, bone cartilage and thyroid tissues. Mature ectoderm layer (skin and hair) and mesoderm layer (fat and muscle) is the commonest composition.[44] They have a characteristic feature on ultrasound and are often detected incidentally. They can present acutely with abdominal pain, as they are prone to torsion (16%) due to the fat content.[45] Fertility preserving cystectomy is often adequate treatment.

Unlike epithelial tumours, often germ cell tumours are not associated with raised CA125. Different types of biomarkers, such as alpha feto protein (AFP), human chorionic gonadotropin (HCG) and lactate dehydrogenase (LDH) may be raised depending on the sub type of GCT. GCTs are also often symptomatic and identified at an early stage, respond to chemotherapy and have good prognosis when treated appropriately.

Stromal cell tumours

Stromal cell are heterogenous group of tumours in the ovary that arise from the primitive sex cord or stromal cells. They are uncommon and comprise only 7% of all ovarian tumour.[46] The sex cord cells give rise to Granulosa and Sertoli cells tumours, whilst the stromal cells give rise to Thecomas, Fibromas and Leydig cell tumours, and may present on their own or as mixed types. Granulosa cell and Theca cell tumours are oestrogenic and may present as postmenopausal bleeding, whilst Sertoli and Leydig cell tumours are androgenic and may

present with virilising symptoms. They occur in a wide range of women but are more common in younger and older women. They were reclassified in the new WHO classification and are grouped under 3 main categories of pure stromal tumours, pure sex cord stromal tumours and mixed stromal-sex cord tumours.

Sclerosing stromal tumour and Fibromas are hormonally inert. Thecomas and Luteomas often produce oestrogen. Unlike sclerosing ovarian tumours that arise in the first three decades of life, Thecomas are common in the 5th decade and in postmenopausal women. Even though Thecomas are themselves benign they can be associated with oestrogenic sequelae, such as endometrial adenocarcinoma or rarely endometrial stromal cell sarcoma in 20% of thecomas.[47] Less commonly, Luteomas can also produce androgenic hormones. Leydig cell tumours are also benign androgen producing tumours seen more commonly in postmenopausal women. They all appear as solid ovarian tumour, grey-white or red-brown in colour. They can be incidental findings on imaging or present with hormonal symptoms and need to be distinguished from the rare malignancies that can arise in stromal-sex cord tumours.[47] Granulosa tumours are associated with increased levels of Inhibin, which is used as a biomarker to aid diagnosis and surveillance; the levels of inhibin are influenced by the volume of the tumour.[48]

1.3.2 Borderline ovarian tumours

Borderline ovarian tumours (BOT) are tumours of epithelial origin with upregulated cellular proliferation and mild nuclear atypia but without destructive stromal invasion.[49] They were first recognised as an entity in 1929 and incorporated into the WHO classification in 1973.[50] The current recognised WHO classification (2014) for BOTs, and malignant tumours are as given in Table 1.3. According to the 2014 classification the term “tumours of low malignant potential” is no longer recommended and is replaced by atypical proliferation.[51] There are 6 subtypes of BOTs similar to the malignant counterparts serous (SBOTs - 50%) and mucinous (MBOTs - 45%) being most common. The others less commonly occurring types are endometrioid, clear cell, seromucinous and Brenner.[52] Histologically they have mild to moderate atypia, hyperchromasia, epithelial multilayering and detachment of cells into the lumen.

SBOTS and MBOTS generally occur between 33-44 years, unlike the Endometrioid borderline ovarian tumours (EBOT) where the mean age at diagnosis is 57 years. The clinical or ultrasound features of BOTS are no different from those of benign lesions or early malignancy. BOTs may act as a precursor lesion, and continuum of progression to malignancy has been demonstrated. Serous BOTS often give rise to LGSOC rather than HGSOC and share similar genetic profiling with LGSOC, demonstrating BRAS and KRAF mutations. Areas of BOT and LGSOC may coexist.[53, 54] Mucinous tumours are markedly heterogeneous with the coexistence of benign, borderline and invasive components within the same cyst and extensive sampling is required at histology to rule out invasive malignancy. Benign metaplastic endometriosis outside the uterus can undergo stepwise transformation to EBOTs and then develop malignancy (low grade serous or clear cell cancer), although the overall risk of this is low. Clear Cell (CCBOTs) and Brenner borderline ovarian tumours are rare. CCBOTs represent <1% of BOTs and have a high likelihood of association with clear cell carcinoma.

Apart from the potential for malignant transformation, another addition to the recent classification is the use of the term 'micro invasion' to replace non-invasive implants. These are isolated groups of cells no larger than 5 mm in the largest dimension, associated with BOTs. Unlike the malignant epithelial cell tumours, the majority of BOTs are diagnosed in stage 1 (75%) and have a 5 year survival rate of 97%.[55] Although in the minority, serous borderline ovarian tumours are associated with microinvasion, peritoneal implants, lymphadenopathy and progression to invasive cancers. Hannibal et al. showed the micropapillary variant to be associated with more advanced stage at presentation and a worse prognosis.[56] However, overall the clinical significance of all these variants and its impact on survival is debatable with studies varying in its conclusion. The criteria for diagnosis are largely based on the 2003 NIH consensus criteria.[57] They commonly occur in the reproductive age group, and fertility sparing surgery can be advocated. If fertility is not desired the treatment is removal of the reproductive organs (hysterectomy and removal of both tubes and ovaries) with the omentum. However, there remains the preoperative challenge of distinguishing it from the benign and malignant counterparts as blood tests, and imaging can be similarly abnormal, and it is challenging to make a clear distinction in order to tailor the location or extent of surgery. This is of particular relevance when considering issues such as fertility preservation in premenopausal women.

Table 1. 2: WHO histological classification of ovarian tumours (2013).

Epithelial cell tumours	Germ Cell tumours	Sex cord stromal tumours	Metastatic*
Serous carcinoma	Endodermal sinus tumour	Granulosa cell	Breast
Mucinous carcinoma	Embryonal carcinoma	Sertoli-Leydig cell tumour	Colon
Endometrioid carcinoma	Choriocarcinoma	Theca cell tumour Thecoma	Endometrium
Clear cell carcinoma	Dysgerminoma	Fibroma	Stomach
Malignant Brenner	Immature Teratoma	Sex cord tumour with annular tubules	Cervix
Transitional cell carcinoma	Yolk sac tumour	Lipid cell tumour	
Small cell carcinoma	Polyembryoma	Gynandroblastoma	
Mixed mesodermal	Mixed germ cell tumour		
Undifferentiated Carcinoma			

*common sites of origin

1.3.3 Malignant ovarian tumours

Epithelial tumours are the commonest of all ovarian tumours and contribute to 75% of all tumours and 90% of malignancies. Germ cell tumours and stromal cell tumours contribute to 15 - 20% of neoplasms (commoner in adolescents) and 5 - 10% of malignant tumours respectively. Metastatic tumours including Krukenberg tumours, account for 5%.

Table 1. 3: New WHO histological classification of epithelial ovarian tumours.[1]

Previous	NEW (2014)
Serous tumours	
Benign type	
Cystadenoma	
Papillary cystadenoma	Cystadenoma
Surface papilloma	Adenofibroma
Adenofibroma and cystadenofibroma	Surface papilloma
Borderline (SBOT)	
Papillary cystic BOT	Serous BOT/Atypical proliferating tumour
Papillary surface BOT	SBOT micropapillary type/Non-invasive serous, low grade carcinoma
Adenofibrormatous and cystadenofibromatous BOT	
Malignant type	Serous low- grade carcinoma
Adenocarcinoma	Serous high-grade carcinoma
Papillary surface carcinoma	
Adenocarcinofibroma	

Mucinous tumours	
Benign type	
Cystadenoma	
Adenofibroma and cystadenofibroma	
Mucinous type with mural nodules	
Mucinous cystic tumour with pseudomyxoma peritonei	Cystadenoma
Borderline (MBOT)	Adenofibroma
Intestinal type	
Endocervical type	
Malignant type	
Adenocarcinoma	
Adenocarcinofibroma (malignant adenofibroma)	Mucinous BOT/atypical proliferating mucinous tumours
	Mucinous carcinoma
Endometrioid tumours	
Benign type	
Cystadenoma	Endometriosis cyst
Adenofibroma and cystadenofibroma	Endometrioid cystadenoma
Borderline (EBOT)	Endometrioid cystadenofibroma
Cystic tumour	
Adenofibroma and cystadenofibroma	Endometrioid EBOT/Atypical proliferative endometrioid tumour
Malignant type	
Adenocarcinoma NOS	Endometrioid carcinoma
Adenocarcinofibroma (malignant adenofibroma)	
Malignant Mullerian mixed tumour (carcinosarcoma)	
Adenosarcoma	
Endometrioid stromal sarcoma (low- grade)	
Undifferentiated ovarian sarcoma	
Clear cell tumours	
Benign type	
Cystadenoma	
Adenofibroma and cystadenofibroma	
Borderline (CBOT)	Cystadenoma
Cystic tumour	
Adenofibroma and cystadenofibroma	CBOT/atypical proliferating clear cell tumour
Malignant type	
Adenocarcinoma	
Adenocarcinofibroma (malignant adenofibroma)	Clear cell carcinoma
Transitional cell tumours	**Brenner tumours**
Benign type	
Brenner tumour	Brenner tumour
Metaplastic type	
Borderline	Borderline Brenner tumour/Atypical proliferating Brenner tumour
Borderline Brenner tumour	
Proliferative type	
Malignant type	Malignant Brenner tumour
Transitional cell carcinoma	
Malignant Brenner tumour	**Seromucinous tumour**
	Benign type
	Seromucinous cystadenoma
	Seromucinous adenofibroma
	Borderline
	Seromucinous borderline tumour/Atypical proliferating seromucinous tumour
	Malignant type
	Seromucinous carcinoma
Squamous epithelial tumours	
Mixed epithelial tumours	
Undifferentiated and unclassifiable tumours	**Undifferentiated carcinoma**

1.4 Factors influencing risk of ovarian cancer

The cause of OC is not currently known, but several factors have been noted that may increase or reduce the risk of developing OC. Age, inherited gene mutations, reproductive history and family history are some of the factors that have been linked to risk of OC. Understanding and awareness of these are important as the risk profile alludes to the pre test probability of OC.

1.4.1 Age

The risk of developing OC increases with age. According to the American Cancer Society (ACS), more than half of the women who develop OC are older than 63 years, and it is commonest in women aged 75-79 years.[58] The average number of new cases per year and age specific incidence rates per 100.000 females is shown in Figure 1.1. More than 80% of the epithelial ovarian cancer occurs in the perimenopausal and postmenopausal age group (> 40 years of age). Germ cell tumours account for 15-20% of all ovarian tumours, but unlike Epithelial Ovarian Cancers, 2/3rd of them appear in younger women and children, and only 3-5% are malignant.[59, 60] Familial OC occurs at younger ages compared to sporadic OC.[14] With successive generations a birth cohort effect for familial (mainly BRCA related) cancers is seen. i.e. increasing OC risks with successive generations.[61]

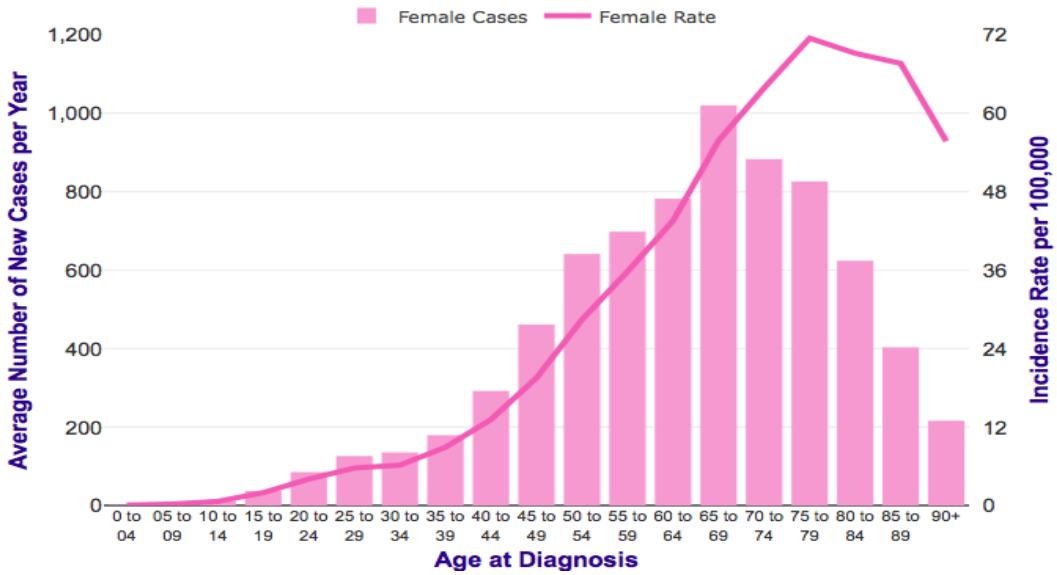


Figure **1. 1: Age-Specific Incidence Rates per 100,000 population, females, UK, 2013-2015.**[10]

1.4.2 Inherited genetic risk

The majority of the ovarian cancers are sporadic and occur due to somatic mutations that are acquired in a person's lifetime. However, a family history of ovarian cancer or breast cancer is considered to be one of the most significant risk factors for development of ovarian cancer. Therefore, family history is one of the referral criteria in the ACOG guidelines. About 15-20% of non-mucinous EOCs are inherited as autosomal dominant mutations in the germ line cells; they are carried in families and passed by either parent. The majority of non-mucinous EOCs occur due to an inherited mutation in breast cancer genes (e.g. BRCA 1 and 2),[14, 62] that increase susceptibility to develop ovarian cancer in comparison to the general population. Serous and Endometrioid are two of the commonest histology types seen in BRCA carriers, whereas BOTs and non-mucinous are less common.[63, 64, 65] Mutations in the Breast cancer genes (BRCA 1 and 2) are the most common heritable mutation and have high penetrance. BRCA 1 and BRCA 2 are tumour suppressor genes and women with first degree relatives (i.e. a mother or sister or daughters) with a history of OC, have an increased risk of developing ovarian cancer especially if the disease has developed at a younger age. The risk reduces with increasing age.[66] The lifetime risk of developing by age 70 years can be up to 63% and 27% in BRCA1 and BRCA2 carriers, respectively.[15, 67] 39% of women with BRCA 1 and 11-17% of women with BRCA2 develop ovarian cancer by 70 years.[68, 69] The risk increases significantly if more than one relative is affected, ranging from 3.0 - 23.5 depending on the number of affected members in the family.[66, 70] Genetic risk for carrying BRCA genes may be identified, and lifetime risk can be calculated using tools such as BOADICEA and Manchester scoring system. Surgical prevention in the form of risk reducing bilateral Salpingo-oophorectomy (RRBSO) in these high-risk women can reduce the risk of ovarian cancer in BRCA carriers by 80-96%, but a 1-6% residual risk of primary peritoneal cancer (PPC) remains.[71]

Another, more commonly known heritable risk factor is families with Lynch syndrome II, also called hereditary nonpolyposis colorectal cancer syndrome (HNPCC). Due to defects in the mismatch repair genes MLH1, MSH2, MSH6 PMS1 and PMS2.[72, 73] The other less known but heritable causes of OC are PTEN tumour hamartoma syndrome (due to mutation of the PTEN gene), Peutz-Jeghers syndrome (mutations in STK11) and MUTYH-associated polyposis. These mutations are associated with increased risk of breast cancer, colon cancer and endometrial cancer, and hence family history of these cancers also indicate high risk of developing OC.

BRIP1 and RAD51c and RAD51D are moderate penetrance mutations in DNA repair genes and are known to be associated with increased inherited risk of OC of 5.8, 5.2 and 12% respectively.[74, 75] A consensus recommendations from the UK genetics group for ovarian cancer panel testing recommend that women with non-mucinous EOC should undergo testing for BRCA1, BRCA2 BRIP1, MLH1, MSH2 MSH6 and RAD51C. The consensus committee does not recommend testing for EPCAM deletions, TP53, or PMS2 based on current evidence, despite the initial majority recommendation in the survey, as the risk was not considered significant [76] in a wider population.

Mainstream testing (i.e. the routine testing of BRCA 1 and 2 genes status in women with HGSOC and CCC) is now standard of care in the UK.

Although epidemiology on reproductive factors, Environmental, dietary and host factors are of interest as risk factors, none have been incorporated into models of diagnostic test accuracy and are not discussed in this thesis**.**

1.5 Diagnosis of ovarian cancer

Survival for OC in England is lower in comparison to Scandinavian and many European countries. This difference is unlikely due to availability of different tests but may be due to different factors such as attitudes or access to health care, better general health or different methods to estimate relative survival. Five thousand or more deaths within the 5 years of diagnosis can be avoided in England in comparison to the European average.[77] This has been attributed to delays in diagnosis, and the delays can be in patient presentation to primary care practitioners, delays in primary care referrals, delays in secondary care referrals or delays in treatment. The first year survival in England is much worse in comparison to their European counterparts with comparable health care systems and seen as reflection of delayed diagnosis.[78] The document "*on route to diagnosis*" suggests that 30% of OC cases in England are diagnosed via the emergency route. These cases show far worse 1 year survival rates (45%) in comparison to the 84% survival at 1 year for women diagnosed by the urgent cancer referral pathway (2 week referral) and the non urgent GP referral route.[79] Of the women who died in their first year after diagnosis, 58% did not receive any form of treatment, reflecting a need to improve symptom awareness, early diagnostic pathways and access to treatment.[80]

The majority of OCs (75%) are diagnosed at an advanced stage and responsible for the highest mortality among gynaecological cancers. The lack of awareness and recognition of symptoms by patients and physicians is considered one of the main factors in delayed diagnosis and poor outcomes. Diagnosis of OC is challenging because of variable presentation, the non-specific nature of symptoms,[81] and low prevalence in primary care (0.09 - 0.18%).[82] The heterogeneity of OC does not lend itself to diagnosis by a single test as they have different clinical presentations. The known tests, such as ultrasound and biomarkers, have varying sensitivity and specificity for different sub groups.

1.5.1 Diagnostic challenges and dilemmas in the diagnosis of ovarian cancer

Historically OC was called a "silent killer". However, we know now that most women with OC have symptoms many months prior to their initial diagnosis.[83, 84] The common presenting symptoms known to be associated with OC are pelvic or abdominal pain, abdominal distension or bloating, loss of appetite or feeling full quickly and urinary frequency and urgency.[85-87]

These symptoms are not only non-specific but mimic symptoms of common benign conditions including gynaecological, non-gynaecological and functional disorders (e.g. menstruation, irritable bowel syndrome, endometriosis, etc.). A symptom index has been designed by Goff et al. and is defined by the presence of any of the following 6 symptoms: pelvic/abdominal pain, abdominal bloating/increased size, reduced appetite or a feeling of early fullness experienced more than 12 times a month and present for less than one year;[84] this is the most validated questionnaire used.[88] The median duration of symptoms from presentation to diagnosis has been noted to be between 3 - 5 months and its value in prompting early detection and effect on improving outcomes have been questioned by some experts.[89-91]

Gilbert et al. performed a pilot study evaluating the role of symptoms which triggered testing for early diagnosis of ovarian cancer in women over 50 years (DOvE).[92] This study found that despite not altering the stage at diagnosis, burden of disease was lower in the cohort of women that were identified by symptom triggered testing. Also, a higher proportion of women in the symptom triggered group achieved optimal cytoreduction. The authors concluded that earlier detection using symptom triggered testing resulted in earlier diagnosis and lower volume disease in women with HGSOC, and further research is awaited. As maximal cytoreduction

and lower residual disease is a the most significant prognostic factor[93-95] in improving ovarian cancer outcomes, the role of symptom triggered testing in the future may play a promising role. Furthermore, nearly 22% of women in the UK diagnosed with OC will not receive any cancer treatment at all; likely due to poor performance status at presentation. Thus, symptom triggered testing may also enable diagnosis at a better functional status, consequently improving survival.

The National Institute of Clinical Excellence (NICE) recommends that the persistent presence of pelvic/abdominal pain, abdominal bloating/increased size, reduced appetite or early fullness, or urinary frequency, especially in women over 50, for more than 2 weeks and less than 12 months duration, should trigger primary care doctors to investigate women with CA125 followed by ultrasound if CA125 is abnormal (>= 35U/ml), and referral to secondary care if both tests are abnormal (see Figure 1.1).

However, CA125 is only elevated in EOCs, not elevated in approximately half of the early stage OCs, and may be normal in mucinous OCs. Hence it may be less useful for early diagnosis.[96] On the other hand, CA125 is an inflammatory marker and not specific to ovarian cancer; therefore CA125 may be elevated in other benign conditions including endometriosis, menstrual disturbances and non gynaecological conditions.[97] A majority of women with these symptoms do not have OC, and a raised CA125 may lead to many unnecessary referrals and overwhelm the services as well as causing patient anxiety. A multicentre audit looking at adherence to NICE recommended pathway in the UK showed that 90% of the GP referrals did not follow the sequential testing and referral guidance prescribed by NICE; however, the symptom triggered referral led to more diagnosis by the urgent referrals dedicated for cancer pathway rather than emergency admissions.[98] A study by Lim et al. identified that 1 in 2 women between the ages of 45 and 70 would present to their GP with an OC symptom once a year. Accurate diagnostic strategies, both in primary care and in secondary care, are vital for the early diagnosis of OC, whilst not overburdening the health service and women with false positive referrals.

Germ cells tumours and sex-cord stromal tumours can also present with severe abdominal pain, distension and fever because of rapid growth of the tumours causing capsular distension, haemorrhage and necrosis which mimics ovarian torsion. The symptoms and signs overlap and hence women may present to primary care, secondary care in gynaecology or other specialities, or acutely to emergency care.

As diagnostic tests, women with symptoms undergo imaging (pelvic or abdominal ultrasound, Computed tomography or Magnetic resonance imaging) or blood test, commonly CA125 (HE4 and OVA1 approved for clinical use in USA). The diagnosis is more difficult in premenopausal women as the prevalence of benign conditions is higher and may have associated raised CA125, confounding the reliability of the test. Non-epithelial ovarian cancers (i.e. germs cell tumours and sex-cord stromal tumours) may not have raised CA125, and instead, biomarkers may be raised (AFP, HCG, LDH, Inhibin) which are not specific to OC. Another problem critically affecting early diagnosis is the absence of raised CA125 in nearly 50% of early stage EOC.[96] Differentiating the more common benign tumours from malignant ovarian tumours on imaging is based on a combination of various morphological parameters, blood flow or pattern recognition, which can be influenced by the operator's experience and skill.

OC is managed by a combination of surgery to remove all visible disease and chemotherapy. Patients with benign tumours can have conservative management, i.e. surveillance or limited surgery. The treatment for OC is radical surgery but benign tumours, if requiring surgery, can have conservative surgery. The outcome of OC is better when operated by a gynaecological oncologist.[99, 100]

An accurate diagnosis is critical to ensure the right women get the right treatment in the right place to have optimal outcome. However, all these influencing factors influence care, making the diagnosis of OC difficult and affect the clinical pathway.

1.5.2 Effect of diagnostic practice on clinical treatment pathway

Current data do not support screening for OC. In the absence of any screening studies the recommendation from UK, Europe and USA is to generally increase awareness of symptoms related to OC in the population and use symptoms as triggers for further testing, with appropriate referral to secondary care (general gynaecologists/ cancer units) or tertiary care (gynaecological oncologists/doctors trained to perform surgery for advanced cancer).

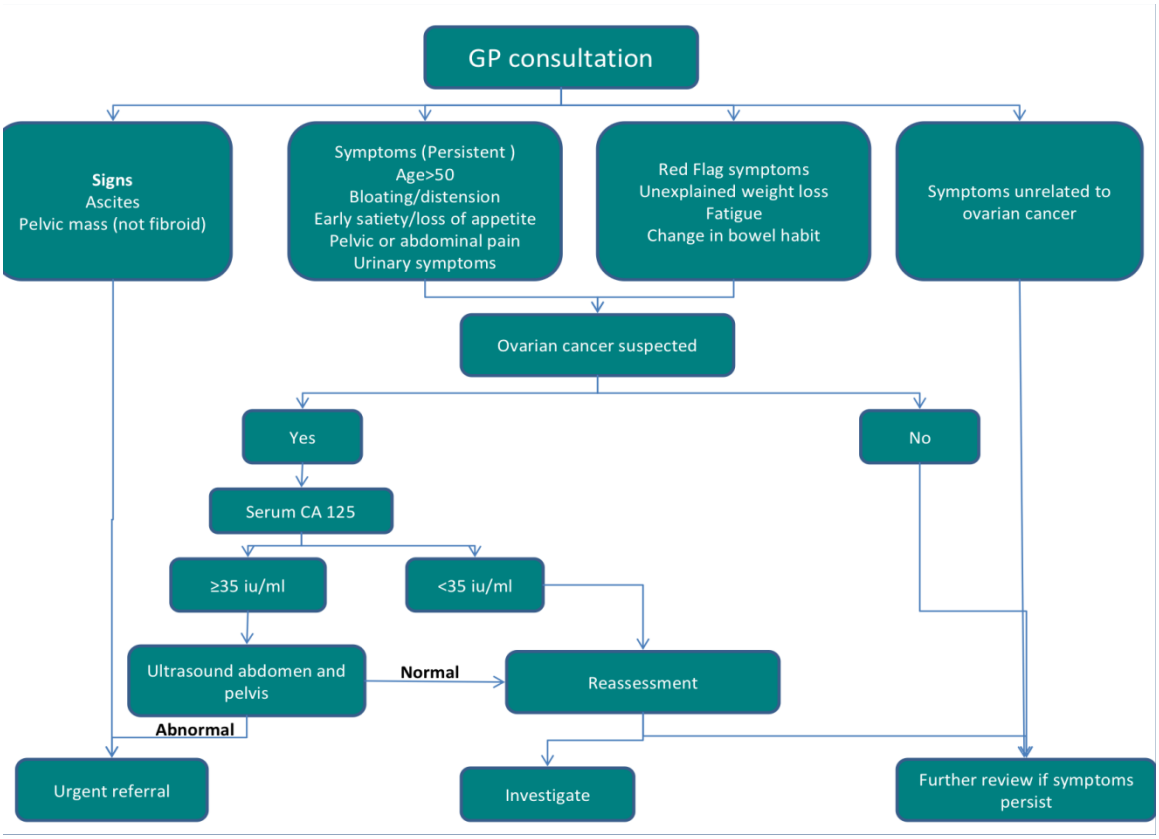


Figure 1. 2: NICE pathway for ovarian cancer referral in primary care.

Unfortunately, as mentioned in previous sections, the symptoms are non specific and there is lack of awareness regarding these symptoms amongst both the population and medical practitioners, leading to unstreamlined diagnostic pathways. This leads to women presenting at different points of health care leading to investigations and referrals, causing diagnostic delays and delays in access to right treatment. In reality, unlike the streamlined pathway prescribed by NICE, the pathways seen in clinical practice are more likely to resemble that shown below (Figure 1.3).

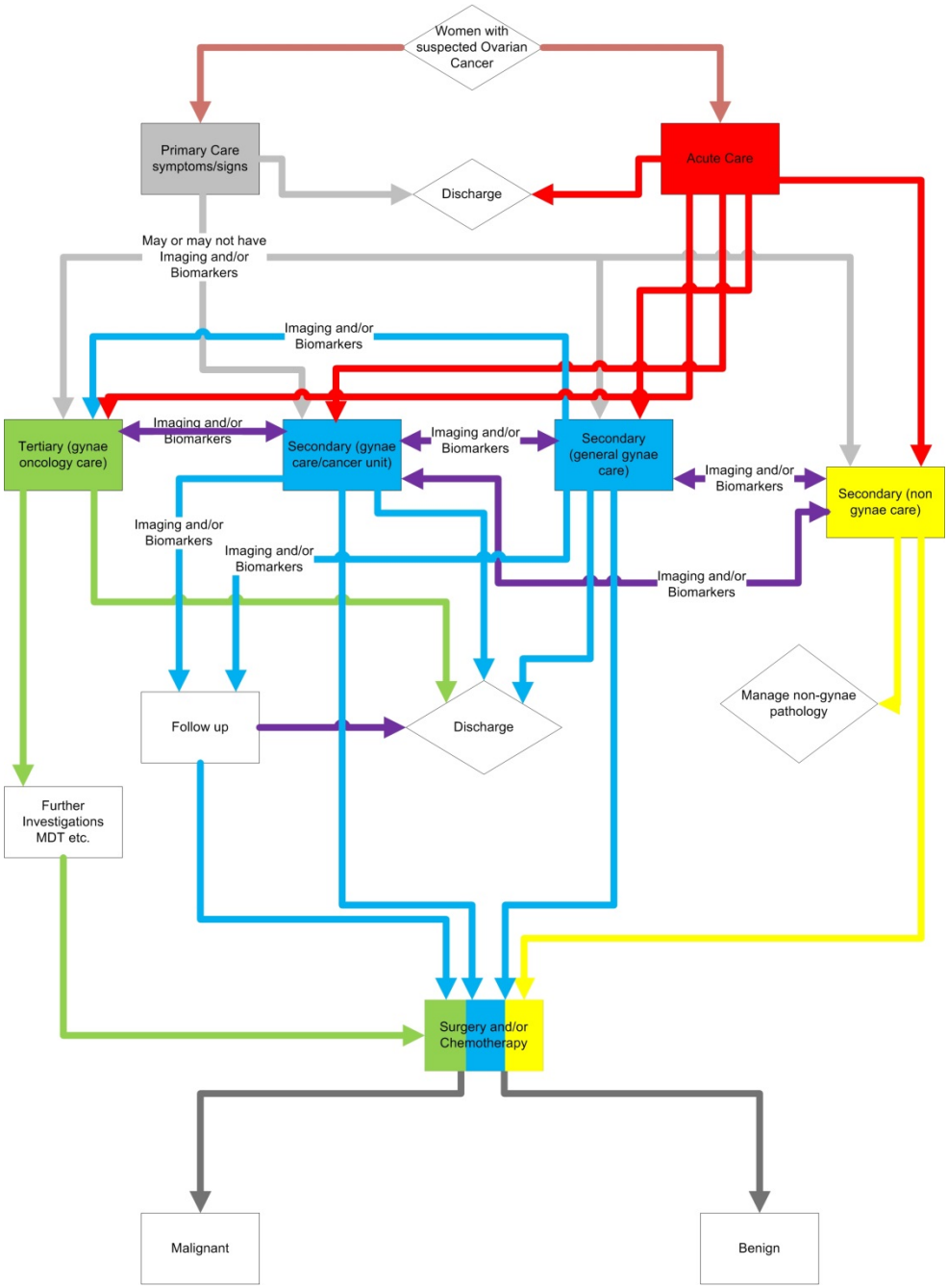


Figure 1. 3: Diagnostic pathway of ovarian cancer in current clinical practice.

As mentioned in previous sections, currently available tests include blood tests, biomarkers (such as CA125, HE4 and OVA1 etc.), imaging (transvaginal ultrasound, Computed tomography, Magnetic resonance imaging etc.) and combination testing using any permutation of symptoms / demographics / family history / blood test / imaging (RMI, ROMA, LR2, ADNEX etc.). However, there is no consensus regarding the test of choice, the point of pathway they are used in or accepted thresholds for test positivity for many of these test(s).

1.6 Aims and objectives

The aims and objectives of my research were to systematically identify the accuracy and applicability of existing tests for OC from existing literature. My research formed part of the broader ROCkeTS project; www.birmingham.ac.uk/rockets which aim to identify best tests for OC in primary care and in secondary care, in pre and postmenopausal women. My research comprised Phase 1 of ROCkeTS. I started my project with a critical evaluation of published systematic reviews and progressed to conduct fresh systematic reviews that addressed the limitations of published reviews. My project also enabled improved design and analysis of ROCkeTS primary study.

1.6.1 Primary objective

i) To critically evaluate published systematic reviews and identify evidence gaps. ii) To establish the accuracy of biomarkers alone or in iii) combination for the diagnosis of OC in pre and postmenopausal women.

Index tests were analysed separately in the following groups:

- Medicines and Healthcare Products Regulatory Agency (MHRA), National Horizon Scanning Centre (NHSC), National Institute for Health and Care Excellence (NICE), European Medicines Agency (EMA), FDA, Royal College of Obstetricians and Gynaecologists (RCOG) and American College of Obstetricians and Gynecologists (ACOG) approved or recommended Biomarkers alone or in combination (CA125, CEA, HE4, OVA1, LDH, HCG, AFP)
- Combinations of tests of more than one modality (biomarkers, imaging) across categories above and or combined with family history / demographics (either as rules or multivariate models
- To compare the accuracy of different tests or test combinations.

1.6.2 Secondary objectives

The following sources of heterogeneity were investigated in my thesis as secondary objectives:

Population:

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

Index tests:

- Test positivity threshold.

Target condition:

- Epithelial ovarian cancer versus Non-Epithelial ovarian cancer.
- Early versus late stage.
- Type I versus Type II.

Study quality:

- Case-control versus other study designs.
- Study quality: for study participants not receiving surgery initially following a negative index test result: 12 months follow-up versus less than 12 months follow-up.

CHAPTER 2: METHODS

Diagnostic Test Accuracy (DTA) studies evaluate the accuracy of a test of interest (index test) with the best available method (reference standard) for determining the true presence or absence of the condition of interest (target condition). Systematic reviews of diagnostic test accuracy (DTA) studies play an important role in providing estimates of test accuracy based on all available evidence. The review of studies provides an opportunity to improve the precision of test accuracy and investigate reasons for variation in test results (heterogeneity) by undertaking a meta-analysis. This thesis is an investigation of the comparative accuracy of biomarkers and tests used in combination for the diagnosis of OC.

Development of protocol

In the methods chapter, I describe first the methods planned at the start of the research as per the published generic protocol[101] detailing methods for both the systematic review of biomarker tests and the review of test combinations for the diagnosis of ovarian cancer in symptomatic women. Changes to the protocol during the course of the research and the rationale for these changes are also presented. The methods and results for the systematic review of reviews are presented together in the next chapter.

The methods for this review are based on Cochrane DTA review methods for diagnostic test accuracy systematic reviews. (https://training.cochrane.org/resource/cochrane-handbook-systematic-reviews-diagnostic-test-accuracy) Review methods were tailored to reflect the topic area. In particular, it was anticipated that test accuracy would vary with variation in test positivity threshold (the definition of a positive test result in included studies) and according to histological subtypes of ovarian cancer. For the latter, menopausal status was hypothesised to reflect differences in the prevalence of different histotypes of ovarian cancer (population spectrum). The aim was, therefore, to stratify results based on menopausal status and test positivity threshold. Other factors hypothesised to cause variation in test accuracy were included in planned investigations of heterogeneity or sensitivity analysis and are detailed below.

2.1 Published generic protocol

2.1.1 Question formulation, Inclusion and exclusion criteria

Population (presentation and prior tests)

Adult women aged 18 years or older, irrespective of menopausal status were included. Studies restricted exclusively to populations under 18 were excluded. Studies in pregnant women or women with previous history of ovarian cancer were also excluded.

Prior tests

The reviews are concerned with women in whom a diagnosis of ovarian cancer was suspected based on symptoms, risk factors or signs. As a minimum, women should have self-referred to a healthcare professional on the basis of the presence of symptoms and or risk factors for ovarian cancer. At the point of inclusion, women receive a biomarker or combination test. In practice they may have had prior testing with one or more of history and examination by a healthcare professional, testing with one or more biomarkers or imaging with ultrasound prior to the index test of interest for inclusion. Studies explicitly describing included participants as asymptomatic were excluded, for example where the index test was being applied as a screening test, or where studies included asymptomatic participants and these could not be disaggregated from participants who were symptomatic, had physical signs or were considered at risk of a diagnosis of ovarian cancer. Where the symptom or at risk status of participants was unclear, this was reflected by downgrading a study as part of topic tailored quality assessment of included studies (QUADAS-2) in the patient applicability domain.

Index tests

By definition tests included in the biomarker and combination reviews had to be applicable to both generalist and specialist settings.

Biomarkers

Any of the following recommended or approved biomarkers as detailed in the objectives of the introduction chapter were considered: CA125, CEA, HE4, OVA1, LDH, HCG, AFP. Although

HCG and AFP are not approved for use as tests in ovarian cancer, they are used clinically and recommended by the RCOG and ACOG for women under 40 as additional markers for germ cell ovarian tumours:

- CA125
- CEA
- HE4
- OVA1
- LDH
- HCG
- AFP

Combinations of tests

Any combination of two or more of the following test types 1) symptoms 2) imaging with ultrasound, biochemical markers, 3) risk factors, e.g. family history for ovarian cancer, age. A combination of multiple biomarkers would be a combination of only one test type and therefore considered part of the review of biomarkers; tests were included if used at any threshold and combined in any order. The combination tests included in the review are listed below in Table 2.1.

Reference standards

Histology was used as the reference standard for OC in women who had undergone surgery for an adnexal mass. As it would be unethical to undertake major surgery in women with a low index of suspicion for OC, follow up for a minimum of 6 months of women with a low index of suspicion and not undergoing surgery or women declining surgery was considered acceptable as an alternative reference standard. For studies using clinical follow-up, a year of follow-up was considered of adequate quality. Studies employing follow up for >6 but < 12 months were downgraded as part of quality assessment.

Table 2. 1: Definition of included test combination.

Index Test Combination	Detail	Test positivity thresholds included (reference)
RMI1 ***U* × *M* × CA125**	**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 **Serum CA125:** CA125U/ml applied directly to the calculation	200, 250 (Jacobs 1993)
RMI2 ***U* × *M* × CA125**	**Ultrasound (U):** (1 point for each of multilocular cysts, solid areas, metastases, ascites and bilateral lesions) where a total ultrasound point score of 0 or 1=1, and a point score of ≥2 =4 **Menopausal status (M):** premenopausal=1 and postmenopausal =3 **Serum CA125:** CA125U/ml applied directly to the calculation	200, 250 (Mol 2001)
RMI3 ***U* × *M* × CA125**	**Ultrasound (U):** (1 point for each of multilocular cysts, solid areas, metastases, ascites and bilateral lesions) where a total ultrasound point score of 0 or 1=1, and a point score of ≥2 =3 **Menopausal status (M):** premenopausal=1 and postmenopausal =3 **Serum CA125:** CA125U/ml applied directly to the calculation	200, 250 (Yamomoto 2009)
RMI4 ***U* × *M* × *S* (size in centimetres) × CA125**	**Ultrasound (U):** (1 point for each of multilocular cysts, solid areas, metastases, ascites and bilateral lesions) where a total ultrasound point score of 0 or 1=1, and a point score of ≥2 =4 **Menopausal status (M):** premenopausal=1 and postmenopausal =4 **Size (S):** tumour size (single greatest diameter) of <7 cm =1, and ≥7 cm =2. **Serum CA125:** CA125U/ml applied directly to the calculation	400, 450 (Yamomoto 2009)
ROMA **Algorithm using CA125, HE4 and menopausal status**	**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	Pre 7.4% & Post 25.3% (Bandiera 2011) Pre 12.5% & Post 14.4% (Van Gorp 2011)

Table 2.1 Continued...

ACOG version 1 **ACOG v1**	**-Premenopausal**: 1 of CA125 >200; family history positive for either breast ca or OC in first degree relatives; Ascites; abdo or distant mets/ pleural effusion **-Postmenopausal**: 1 of CA125 >35; family history positive for either breast ca or OC in first degree relatives; Ascites; abdo or distant mets/ pleural effusion; nodular or fixed pelvic mass	Presence of any one positive test component (ACOG 2002)
ACOG version 2 **ACOG v2**	**-Premenopausal:** 1 of CA125 >67; family history positive for either breast ca or OC in first degree relatives; Ascites; abdo or distant mets/ pleural effusion **-Postmenopausal:** 1 of CA125 >35; family history positive for either breast ca or OC in first degree relatives; Ascites; abdo or distant mets/ pleural effusion; nodular or fixed pelvic mass	Presence of any one positive test component (Dearking 2007)
ACOG version 3 **ACOG v3 (CA125)**	**-Premenopausal:** 1 of CA125 >67; Ascites; abdo or distant mets/ pleural effusion **-Postmenopausal**: 1 of CA125 >35; Ascites; abdo or distant mets/ pleural effusion; nodular or fixed pelvic mass	Presence of any one positive test component (Dearking 2007)
ACOG version 3 **ACOG v3 (OVA1)**	**-Premenopausal**: 1 of OVA1; family history positive for either breast ca or OC in first degree relatives Ascites; abdo or distant mets/ pleural effusion **-Postmenopausal**: 1 of OVA1; family history positive for either breast ca or OC in first degree relatives Ascites; abdo or distant mets/ pleural effusion; nodular or fixed pelvic mass	Presence of any one positive test component (Miller 2011)
LR2	-age of the woman (in years) -presence of ascites (yes, 1; no, 0) -presence of blood flow within a solid papillary projection (yes, 1; no, 0) -maximum diameter of the solid component of the adnexal mass (expressed in millimetres, but with no increase 950 mm) -irregular internal cyst walls (yes, 1; no, 0) -presence of acoustic shadows (yes, 1; no, 0), The probability of malignancy is	Test positivity threshold to achieve a post test probability of malignancy 10% (Timmerman 2010)
ADNEX	-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 -more than 10 cyst locules -acoustic shadows	Test positivity threshold to achieve a post test probability of malignancy of 3%, 5%, 10% and 15% (Van Calster 2014)

Target condition

OC, of all stages and histological types, was included. Studies restricted to specific ovarian pathologies were excluded with the exception of studies restricted to only epithelial ovarian cancer (EOC) as this is the most common of the ovarian cancers (> 90% of tumours in postmenopausal women) and is associated with the highest mortality. Studies restricted to metastatic or recurrent ovarian cancer or where data on recurrent tumours could not be disaggregated were excluded. Studies, where it was not possible to disaggregate metastatic tumours from primary ovarian cancer, were downgraded as part of quality assessment. Metastatic tumours are defined as tumours originating from non-ovarian/ adnexal areas, e.g. colon, breast, endometrium that metastasise to the ovary (also see Table 1.2). Metastatic tumours to the ovary are outside the remit of this review as they will represent the primary tumour pathologically and may not secrete the same biomarkers or exhibit the same ultrasound characteristics.

The quality of included studies was downgraded if borderline ovarian tumours (BOTs) or non-epithelial ovarian cancer (non-EOCs) were excluded from or not reported for the purposes of data extraction and analysis in this review. Non-EOCs are germ cell, and stromal cell tumours originating from the ovary and excluding them contribute as a source of bias as they do not secrete the same biomarkers, e.g. CA125, HE4.

Types of studies

Case control (provided the control arm comprised women with benign ovarian pathology and these could be disaggregated from any healthy controls), cross-sectional (retrospective and prospective data collection) and comparative diagnostic test accuracy studies of any design (within person or between person) were included. In view of the low prevalence of OC, it was anticipated that the majority of 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. Studies were included if there was sufficient data to extract 2 x 2 tables on diagnostic test performance. Inclusion of studies not providing verification of index test negatives was included where 2 x 2 tables could be constructed by imputation using setting-specific prevalence estimates. However, there were no included studies where index test negatives were not verified.

2.1.2 Study identification and selection

Electronic searches

Sensitive search strategies combining terms for the target condition (ovarian cancer) and each of 3 index tests that might be used alone or in combination: biochemical markers, symptom indices and ultrasound imaging were used. In addition, terms to describe testing models and algorithms were added, including the IOTA (International Ovarian Tumour Analysis) variables. The 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).

A date restriction was applied (1991 onwards) to ensure applicability to current technology. Searches were completed in April 2015. No language restrictions were applied. The search strategy used for MEDLINE (Ovid) biomarkers is shown as an example in Appendix 1.

Searching other resources

To identify ongoing and unpublished studies the following trials registers and conference abstracts and proceedings were searched without date restrictions: ClinicalTrials.gov, UK Clinical Research Network Study Portfolio Database (UKCRN) and WHO International Clinical Trials Registry Platform (ICTRP). Conference proceedings were searched from the European Society of Gynaecological Oncology (ESGO), International Gynaecologic 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). The reference lists of existing systematic reviews and guidelines were 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 study selection at the title and abstract and subsequently full text stage was carried out independently and in duplicate by 2 reviewers (NR, RC, PS) with disagreements resolved by a third reviewer (CD, SS). Data

extraction and quality assessment were performed by one reviewer (NR or PS), and 30% of studies were randomly checked independently by a second reviewer (RC or CD) with disagreements resolved by discussion.

Selection of studies

Unique titles and abstracts were reviewed against predefined 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.

2.1.3 Study quality assessment and data extraction

Data extraction and management

A pre-defined data collection form (Appendix 2) was used to extract the following data (if available) into an excel database prior to entry into Review Manager version 5.3: study design; country; setting, including single or multi-centre; method of recruitment; reasons for exclusion; number of participants; number of women with a diagnosis of ovarian cancer and borderline ovarian tumours; age; menopausal status (directly or using age over 50 years or history of previous hysterectomy in the absence of information about postmenopausal status); prior tests; index tests and index test threshold(s); index test operator experience (for symptoms and ultrasound imaging); reference standard (including where relevant duration of follow-up); stage and histological subtype of ovarian cancer. Data to derive a 2 x 2 table for each study was extracted in duplicate by either a clinician (NR) or reviewer (PS) and independently by either a methodologist or statistician (CD, JD, SM, KS).

Assessment of methodological quality

Quality assessment was undertaken using the QUADAS-2 checklist tailored according to the topic and detailed in Appendix 3. Key aspects of tailoring for this review included:

Addition of a question concerning specification of the order and rule for combining individual tests in test combinations (index test domain)

Addition of 2 questions in the index test domain relevant to multivariable model / composite index test validation studies drawing on the PROBAST (prediction model risk of bias assessment) tool for diagnostic and prediction models[102]: pre-specification of thresholds and comparable assessment of all model /test components)

Addition of one question in the reference standard domain concerning the ability to disaggregate index test and reference standard results for primary ovarian cancer, metastatic disease and borderline tumours

Addition of a comparative domain for within and between person comparative studies. concerning participant selection (were they likely to be different from non comparative studies) and index test (were index tests interpreted blind to each other and if the interval between index tests < 3 months)

STARD is a reporting tool for primary test accuracy studies. We did not use the tool to assess reporting quality in our reviews as Cochrane methods follow the STARD guidelines implicitly as part of data extraction and quality assessment; lack of clarity of reporting of included studies is reflected by downgrading in quality assessment

2.1.4 Statistical analysis and data synthesis

Index test subgroups

For each index test, where data allowed separate meta-analyses were conducted for:

Patient characteristics: Mixed populations (pre and postmenopausal combined) versus premenopausal versus postmenopausal status.

Index test characteristics: Index test version, (for example data on 4 different versions of RMI were identified for this review). Test threshold where studies reported a common threshold.

Target condition: Appropriate grouping of borderline test results for construction of the 2 x 2 table. For these reviews, based on current clinical practice, it was considered appropriate if included studies classified histologically borderline ovarian tumours as positive (grouped with

histologically malignant ovarian tumours for estimation of test accuracy). It was considered inappropriate, if included studies excluded borderline ovarian tumours from estimation of test accuracy or if the classification of borderline ovarian tumours in included studies was unclear (borderline tumours may not have been identified in study participants, or classification of borderline tumours as positive (malignant) or negative (benign) for the estimation of test accuracy was unclear.

Methods for meta-analysis

Exploratory analyses included forest plots of study estimates of sensitivity and specificity grouped by test threshold, and plotting sensitivity and specificity in ROC space with test thresholds indicated. Analyses were conducted in Stata version 14.[103] Where adequate data were available, and it was considered reasonable to pool results, meta-analyses were performed using hierarchical models to estimate summary effects. These models take into account the correlation between sensitivity and specificity, threshold effects and the fact that factors causing heterogeneity may affect sensitivity differently to specificity. Hierarchical models deal with statistical distributions at two levels. At the lower level, they model the cell counts in the 2×2 tables extracted from each study using binomial distributions and logistic (log-odds) transformations of proportions. This level deals with sampling variability in individual studies and the correlation that exists between sensitivity and specificity. At the higher level, random study effects analysis is used to account for heterogeneity in test accuracy between studies. The weighted means of a test sensitivity (weighted by prevalence) or specificity (weighted by complement of prevalence) provides overall accuracy; this can provide distorted impression of a validity of the test and have potentially negative consequences especially as the difference between sensitivity and specificity increases and making it inferior to the hierarchical models balanced consideration of sensitivity and specificity. Univariate analysis of sensitivity separate to specificity is not recommended because it does not take into account this correlation. Forrest plots are displayed for a visual representation of test accuracy estimates across studies.[104] Random-effects univariate analyses (which ignore any correlation between sensitivity and specificity) were used where pooling was considered an appropriate approach, but bivariate models failed to converge. Summary estimates of sensitivity and specificity were translated into summary estimates of the absolute numbers of true positives, false negatives, false positives and true negatives using a hypothetical population of 1000 women. Three different estimates of disease prevalence (pre test probability) were used reflecting primary care (0.23%)

and specialist settings (secondary care 10% and tertiary care 30%) taken from the published literature and hospital audits.

The meta-analyses are represented by curves that are fitted when the HSROC hierarchical model is used. This model is appropriate where there is not a common threshold used across studies. The curve illustrates the average estimate of test accuracy across the range of thresholds included. Where included studies can be grouped according to the threshold used (as in the biomarker review) the bivariate hierarchical model is used and a pooled point estimate of test accuracy at a particular threshold can be estimated. This point estimate should be indicated on the SROC plot. It is not possible to produce a point estimate of sensitivity and specificity without reference to a single threshold. In the combination review, to overcome this problem, we estimate sensitivity from the curve at a fixed specificity (80% or 90%) based on clinical consensus about an acceptable FP rate.

Narrative synthesis was used where meta-analysis was not considered appropriate due to clinical or methodological heterogeneity or where only a single study was identified.

Estimation of the accuracy of individual index tests

Since the characteristics measured by index tests could be extracted as 2 x 2 tables reported at common index test thresholds, bivariate model including random effects[105, 106] was used. To estimate average sensitivity and specificity at fixed thresholds, the analysis of each index test version was performed by first restricting to studies that reported thresholds recommended in guidelines and or used in clinical practice and secondly to those thresholds most commonly reported across included studies. In addition, for ROMA, studies using thresholds +/- 2 around the most commonly reported thresholds were included. For CA125 the positivity threshold cut off CA125 35U/ml (RCOG recommended) 65U/ml (modified ACOG) and the 200U/ ml as recommended by RCOG for premenopausal women were investigated along with results within a 5U range of the test thresholds (30-40U/ml and 60-70U/ml). For HE4 as there is no consensus on the clinically agreed threshold cut off for test positivity, thresholds cut off as recommended by manufacture and the most commonly reported clinical thresholds of 70pMol/L and 140pMol/L and also two analyses including results within a 10U range of these

test thresholds (60-80pMol/L, 130-150pMol/L). Studies were excluded if no threshold was given for the reported values of sensitivity and specificity.

Comparison of test accuracy

For the review of test combinations an indirect comparison of each version of each index test was undertaken by fitting HSROC models for each of the 3 patient groups: mixed populations (pre and postmenopausal combined), premenopausal only and postmenopausal only. A summary ROC curve for each version of each index test across all thresholds was estimated. Each included study contributed one threshold to the summary ROC curve. Where an individual study reported more than one threshold, the most commonly reported threshold for that index test across all included studies was selected for the meta-analyses. This selection of threshold was only necessary for ROMA studies where the threshold pair 31.1 (+/-2) and 27.2 (+/2) was the most commonly reported across studies. Summary ROC curves which have a common shape were fitted to the data. Differences in accuracy were reported using the ratio of Diagnostic odds ratios with 95% CI using RMI version 1 as the baseline category, RMI I being the combination test currently in routine clinical use in the UK. The difference in specificity fixed sensitivities of 80% and 90% for each index test version compared to RMI version 1 with 95% CI was also reported.

For the review of biomarkers, test accuracy between different tests was compared first by restricting to studies that make head-to-head (direct) comparisons between tests within the same population as this provides the most reliable evidence.[107] Secondly, tests were also compared by including all relevant studies (indirect comparison), particularly when there are few studies comparing tests within the same population and thirdly comparisons were undertaken based only on studies using appropriate grouping of borderline tumours for analysis (grouped with malignant tumours). Test accuracy were compared by adding a covariate for test type and threshold in the bivariate model and calculating differences and 95% confidence intervals using non-linear estimating methods.

Analyses of differences in accuracy between index tests were performed using the NLMIXED procedure in Statistical Analysis System.[108]

2.1.5 Investigations of heterogeneity

Heterogeneity is investigated in test accuracy meta-analysis by fitting covariates to the hierarchical model used. Likelihood ratio tests are used to test for the significance of covariates on estimates of accuracy.

For each index test group, the number of studies with study characteristics specified in the secondary objectives, and their effect by visual inspection of forest plots and summary ROC plots were explored. Investigations of heterogeneity were undertaken, where an adequate number of studies reporting as including or not including these characteristics were present. Potential sources of heterogeneity were added as separate covariates in the bivariate model to assess association with test performance: menopausal status and grouping of borderline results in analysis (with malignant results or excluded/unclear).

Premenopausal versus postmenopausal women

Accuracy for pre and postmenopausal women for commonly reported thresholds of each index test version were compared.

Comparison of test accuracy for appropriate and inappropriate grouping of borderline test results in 2x2 tables

The effect of appropriate and inappropriate grouping of borderline test results on estimates of test accuracy was investigated by comparing summary ROC curves for studies using inappropriate groupings (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) as the baseline category.

Estimation of differences in accuracy was performed using the NLMIXED procedure in Statistical Analysis System[108] by including menopausal status and borderline grouping as

covariates in the bivariate model. Differences in accuracy were reported using the ratio of Diagnostic odds ratios with 95% CI and associated p values were computed using Wald tests.

2.1.6 Sensitivity analyses

Sensitivity analyses was considered where sufficient studies to investigate the impact on the summary estimates of (i) including only studies with low concern about applicability in the patient selection domain of QUADAS-2, (ii) leaving out potentially highly influential studies and (iii) classification of borderline tumours as malignant or benign.

2.1.7 Assessment of reporting bias

Formal assessment of reporting bias in were not undertake in this review due to current uncertainty about how to assess reporting bias in diagnostic test accuracy reviews, especially in the presence of heterogeneity.[109]

2.2 Differences between protocol and final review methods

2.2.1 Selection of studies

Language

Searches were not restricted to English Language publications but non-English publications were not considered 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.

2.2.2 Study quality assessment and data extraction

Data extraction and quality assessment

Study characteristic data and quality assessment was extracted by a single reviewer (NR or PS) and 30% of studies were randomly checked independently by a second reviewer (RC). Any differences were resolved by discussion.

Methodological quality data was extracted by a single reviewer (NR or PS) and 30% of studies independently checked by a second reviewer (CD). Any differences were resolved by discussion.

Quality assessment

A separate domain for multivariable models was not considered necessary, particularly as no studies including only reporting development of multivariate models were included. Instead 2 questions were added to the participant domain of QUADAS-2 drawing on the PROBAST (prediction model risk of bias assessment) tool for diagnostic and prediction models 1) Prespecification of thresholds and 2) Comparable assessment of all model /test components.

2.2.3 Methods for meta-analysis

Statistical analysis

Test accuracy in pre and postmenopausal women were compared 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.

2.2.4 Heterogeneity and sensitivity analyses

The following planned heterogeneity analyses were not carried out 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 ROB (high/unclear vs low. Case-control study versus other study designs, 12 months follow-up versus > 6 months < 12 months follow-up for study participants not receiving surgery initially following a negative index test result.

Sensitivity analyses were not undertaken for (i) leaving out highly influential studies as this was not considered necessary (ii) including only studies with low concern about applicability in the patient selection domain of QUADAS-2 as there were insufficient studies (iii) 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, a comparison of 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) was undertaken using the HSROC model.

CHAPTER 3: REVIEW OF EXISTING REVIEW

Early detection of ovarian cancer and the aim of reducing mortality has been the goal of research efforts aimed at improving diagnostics in ovarian cancer. Improved understanding of the pathophysiology of ovarian cancer combined with advances in technology in imaging and biochemistry (i.e. proteomics, genomics) has also led to the development of new tests and new studies. Whilst there are many existing guidelines for the detection of ovarian cancer both in Europe (NICE, RCOG and ESGO) and North America (ACOG and SGO) consensus is lacking mainly due to the lack of strong evidence favouring any particular test.

The accuracy of a diagnostic test is not fixed and the methodological design and conduct of test accuracy studies can introduce variation in estimation of accuracy. For example, characteristics of the population being tested including their presentation and prior tests received, how the index test was done and the reference standard used to verify the diagnosis can all influence estimated accuracy and the generalisability of the results.[110] Systematic reviews (SR) and meta-analyses of test accuracy studies therefore can be affected by heterogeneity as a result of variation in these factors in included studies. Willis et al. reported that published SRs of DTAS could be uninformative as they are prone to bias and heterogeneous results.[111]

Given recent developments in meta-analytic and quality assessment methods and the potential for heterogeneity in existing SR of test accuracy it was necessary to evaluate existing systematic reviews addressing the accuracy of tests in OC to assess their currency, validity and applicability to my research question. This will not only inform the need, if any, for a new review but also inform the new review objectives in order to address deficiencies of the existing evidence base.

3.1 Aims and objectives

3.1.1 Primary objectives

1) To systematically identify existing Systematic reviews (SR)of test accuracy for the detection of ovarian cancer

2) To assess the currency, applicability and validity of existing SR of test accuracy for the detection of ovarian cancer reviews to the research questions: “symptoms, ultrasound imaging and biochemical markers alone or in combination for the diagnosis of ovarian cancer in women with symptoms suspicious of ovarian cancer”.

3.1.2 Secondary objectives

1) To assess the reporting quality of the existing SRs using the PRISMA-DTA reporting checklist.
2) To evaluate the usability of the PRISMA-DTA tool.

3.2 Materials and methods

3.2.1 Search strategy and data sources

The search strategy and data sources used for this review are detailed in chapter 2. SRs on the index tests of interest were identified as part of the scoping search for the new test accuracy reviews using a systematic reviews filter.

3.2.2 Inclusion criteria

Objective

Diagnostic test accuracy; studies that evaluated screening, prognosis or recurrence was excluded.

Study design

Systematic reviews- Reviews were included (judged systematic) if they had conducted a systematic literature search including at least 1) two electronic databases, 2) had made an attempt at question formulation: Population, (prior tests and presentation), Index test(s), Reference standard and Target condition and had 3) undertaken quality assessment of individual included studies. Studies were excluded if more than one of these criteria was not present.

Index test(s)

Reviews that investigated the accuracy of one or more of symptoms, FDA approved biomarkers, ultrasound or a combination of these were included.

Target condition

Primary ovarian cancer: all stages and histological types

Non-English publications were not translated. Authors were not contacted.

3.2.3 Study selection

Screening of abstracts and titles, followed by selection of studies for inclusion after reviewing full texts was completed by (NR). Any ambiguity was discussed with a 2nd review (CD) and consensus reached after discussion.

3.2.4 Assessment of applicability, quality assessment and assessment of reporting quality

Data extraction

Data extraction for assessment of applicability, quality of review methods and quality of review reporting was undertaken by NR. The published article and all supplementary figures and tables were reviewed to extract information. Any ambiguity was discussed with CD and consensus reached after discussion. Data extraction was undertaken using EXCEL following piloting on 4 reviews by NR and CD independently.

General information about the review regarding year of publication, title, name of first author and country as identified in the corresponding address and the rationale and objectives of the review was extracted.

There is currently no tool for assessment of the quality (validity and applicability) of SRs of test accuracy. Judgements about the applicability and quality of reviews to the new review

question were therefore made based on existing tools for primary test accuracy studies the QUADAS-2 tool[112] and knowledge of the review topic. This is detailed below.

Judgement of applicability

Population (presentation and prior tests)

The aim of this review was to evaluate the accuracy of tests in symptomatic pre and postmenopausal women suspected of ovarian cancer presenting in non-specialist settings to 'triage women for management in either secondary or tertiary care'.

Index test conduct (operator and threshold)

Many of the tests had several possible thresholds for test positivity. As the accuracy of ultrasound is operator dependent, information on the expertise of US operators by accepted grades of classification EFSUMB or study specific definitions was also sought.

Target condition as defined by the reference standard

Tumours of the ovary are heterogeneous with test accuracy affected by sub types and stages of malignancy as well as characteristics of benign and borderline ovarian tumours. Information about histotypes and stage of malignancy included by reviews and the way borderline tumours were classified for analysis (with either benign or malignant tumours) was therefore noted. Histology in index test positives or follow up for a minimum of 6 months for index test negatives or for women declining surgery was judged to be a suitable reference standard for verification of diagnosis.

Assessment of review quality and currency

Review quality was evaluated using the number of databases searched, attempt at question formulation [population, presentation, prior test, Index test, target condition and reference standard (PppITR)] and any form of quality assessment of included studies. If meta-analysis was performed, statistical methods were noted, in particular use of hierarchical models.[113] Currency of reviews was judged according to the date of completion of search.

Reporting quality

The recently published PRISMA-DTA checklist was used to evaluate the reporting quality of the included reviews. The checklist comprises 27 items covering 6 domains: introduction, methods, results, discussion, conclusion and funding. The PRISMA-DTA tool was piloted by 2 reviewers (NR and CD) before a single reviewer (NR) proceeded to evaluate all. During piloting, it emerged that to accurately capture details pertinent to this review question, it was necessary to tailor items in the checklist. Minimally acceptable information in the study that should be reported for certain items to be scored 'yes' were therefore predefined. For example for the item D1-'intended use of test' It was required authors should state where in the pathway it is to be used, e.g. triage for further testing or management/add on to another test to inform decisions about further testing / management. For item 11- 'definitions for data extraction' presentation and prior tests were required as additional information regarding 'clinical setting'. For item 12 -'risk of bias and applicability' authors were required to report components regarding both -'risk of bias and applicability. If some but not all components of an item were not reported fully, then the item was scored 'limited' and if none of the components were reported, the item was scored 'no'. Tailoring of the PRISMA-DTA checklist is based on the data extraction form and the QUADAS tailoring done to address the questions asked in this review as illustrated in appendix 2 and 3, respectively.

3.2.5 Data synthesis

Data extraction for all elements: assessment of applicability, quality assessment and reporting quality was completed for each review in EXCEL. Results are presented narratively and in tables using descriptive statistics.

3.3 Results

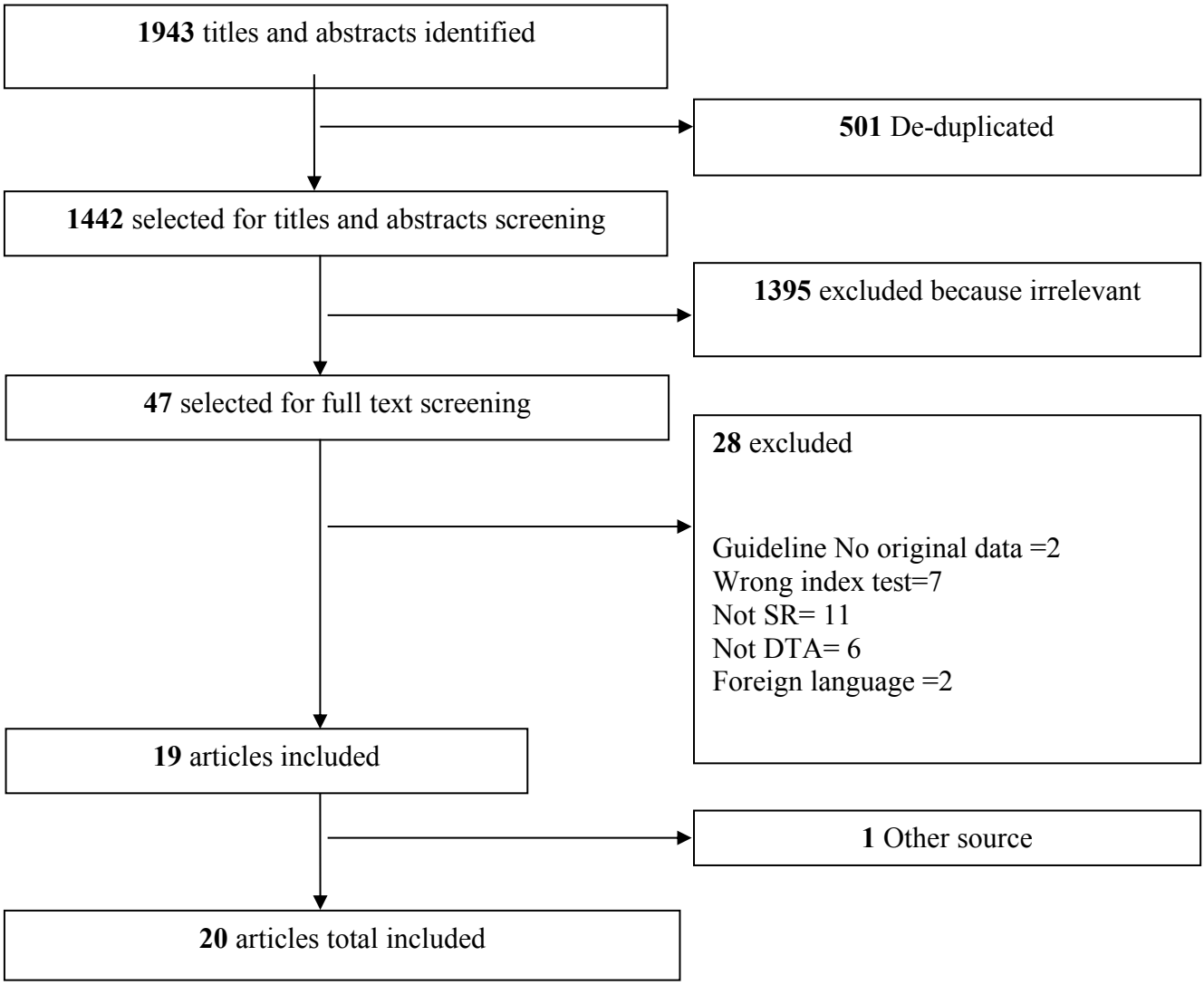


Figure 3. 1: PRISMA of included studies for review of reviews.

3.3.1 Included reviews

After deleting duplicates, 1943 articles underwent title and abstract screening resulting in 47 articles for full text screening (see PRISMA Figure 3.1). 20 publications reporting 23 separate systematic reviews were included. Two reviews investigated the accuracy of symptoms (both individual and in combination), 27 studies included in these reviews investigated the accuracy of a range of biomarkers (CA125, HE4, OVA1) and 35 studies included in these reviews investigated the accuracy of test combinations (RMI1-4, ROMA, ACOG, LR1 and LR2). Where reviews included more than one test assessments of applicability, quality and quality of reporting were undertaken for each test. Date of publication ranged between 2006 [82] and 2016[114]. The reviews were spearheaded from USA, Europe and China. A total of 64 independent syntheses of tests were undertaken across included reviews; These covered single and multiple symptoms (s) (2 syntheses); biomarkers: a total of 27 individual test cohorts that include CA125, HE4 and OVA1; and test combinations: 35 syntheses of individual test combinations including RMI1, RMI2, RMI3, RMI4, ROMA, LR1, LR2 and ACOG. The synthesis of these reviews and assessment of applicability, review quality and currency of the existing reviews are detailed below (also see Table 3.1 and Table 3.2).

3.3.2 Applicability

A total of 64 independent syntheses of tests were undertaken across included reviews; These covered single and multiple symptoms (s) (2 syntheses); biomarkers: CA125 (13 syntheses), HE4 (12 syntheses) and OVA1 (2 syntheses) and test combinations: RMI1 (7 syntheses), RMI2 (7 syntheses), RMI3 (5 syntheses), RMI4 (1 synthesis), ROMA (5 syntheses), LR1 (3 syntheses), LR2(5 syntheses) ACOG (2 syntheses). (see Table 3.1)

Table 3. 1: Review characteristics – applicability to review question(s).

Author and publication date	Population				Index test(s) included		Results can be /have been stratified by:			
	Presentation (A)Symptomatic	Prior tests	Health care setting (Primary/Secondary/ Tertiary)	Biomarkers	Test combination	Symptoms	Menopausal status / Age	BOT are or can be grouped with malignant	Threshold	Histological subtypes
E. R. Myers / Feb 2006	Reported as unable to assess due to lack of information in primary studies	Reported as unable to assess due to lack of information in primary studies	Reported as unable to assess due to lack of information in primary studies	CA125			Yes	Yes	Yes	NR
					RMI1,2 and 3, LR1 and 2		Yes	NR	NR	NR
P. Geomini / Feb 2009	NR	NR	NR		RMI1 and 2		NR	NR	for RMI1	NR
L. R. Medeiros / Feb 2009	NR	NR	NR	CA125			NR	Yes	NA Single threshold included	NR
NICE April /2011 (a) sym	Symptomatic	NR Assume 'no' as presenting in primary care	Primary care			Single and combination symptoms	NR	NR	NR	NR
NICE April /2011 (b) Bio	Noted in summary that results do not directly apply to symptomatic women	'Women with a confirmed adnexal mass or ascites.'	Secondary care	HE4, CA125			NR	NR	NR	NR

Table 3.1 Continued...

NICE April /2011 (c) combi	NR	NR	NR	Yes	RMI1 & 2 and physical examination+ CA125+USS studies from different reviews assuming conditional independence		NR	NR	NR	NR
S. Yu/ Feb 2012	NR	NR	NR	CA125, HE4	No	No	NR	N	Yes for HE4 only	NR
J. E. Dodge/ Apr 2012	NR	NR	NR	CA125	RMI1 and 2 RMI3, LR1, LR2, ACOG	No	NR	NR	NR	NR
F. Li / June 2012	NR	NR	NR		ROMA		Yes*	Yes	NR	EOC Vs OC
				HE4, CA125			Not pooled analysis but represented in Forest plot (as lee than 3 studies)- same done for early vs late	No	NR	
L. Wu/ Sept 2012	NR	NR	NR	HE4	No	No	NR	NR	NR threshold effect found on investigation	NR

Table 3.1 Continued...

J. Lin / Dec 2012	NR	NR	NR	CA125, HE4	ROMA	No	NR	NR	Yes (ideal Vs suggested) for HE4	NR
S.Ferraro/ Feb 2013	NR	NR	Yes (44% in gyn onc setting)	CA125, HE4	No	No	NR (4 studies but not enough to perform pooled analysis)	NR	NR	NR
Z. Yang/ Jul 2013	NR	NR	NR	HE4	No	No	NR	NR	NR	NR
J. Lin / Dec 2013	NR	NR	NR	CA125, HE4	ROMA	No	NR	NR	Yes (ideal Vs suggested) for HE4	NR
J. Kaijser/ Dec 2013	NR	NR	NR	OVA1	Many models including ROMA, RMI (1-4), LR 2	No	Where reported in primary studies	NR	Yes	No (but studies excluded if restricted to specific histological
J.Wang/ Mar 2014	NR	NR	NR	CA125, HE4	ROMA	No	Yes	Yes (EOC VS EOC+BOT)	No (best test threshold reported by study was chosen)	NR§§
S. Zhen /Apr 2014	NR	NR	Yes (60% of the studies- Gyn onc setting)	CA125, HE4	No	No	NR	NR	NR	NR
X. Y. Liao / Nov 2014	NR	NR	NR	CA125	No	No	NA Only postmenopausal women included	NR	NR	NR
A.C.L. Macedo / Sept 2014	NR	NR	NR	HE4	No	No	Yes	Yes	Yes	NR§
M. Stukan / Feb 2015 (no meta analysis	NR	NR	NR	CA125, HE4, OVA1	ROMA, RMI1,2 and 3, LR1, LR2, ACOG	No	Yes	Yes (only studies including borderline with malignant included)	Yes	Yes
M.Ebell / Mar 2016	Symptomatic	NR	NR	No	No	Individual symptoms, combi symptoms	Yes	NR	No/NA	NR

Table 3.1 Continued...

E.M.J. Meys/ May 2016	NR	NR	Yes (LR2 studies 3 studies with 1 tertiary and 2 mixed, RMI -18 studies 9 in secondary or mixed setting); metaregression analysis for heterogeneity with less or more than 25% prevalence	No	LR2, RMI1,2 and 3	No	Yes	Borderline grouped with benign excluded	NR	Studies not reporting all histological subgroups excluded

Note: In the absence of menopausal status age >< 50 used to stratify results

*Additional analyses by stage, methods (low vs high quality);

§early vs late

§§ (also sub group for early vs late and diff testing platforms

**AAN- Artificial Neural network.

Population (presentation and prior tests)

With the exception of the 2 reviews investigating the accuracy of symptoms themselves only 2 reviews concerned with combination and biomarker tests [82, 115] commented on the absence of information on symptom status in included studies.

Only one review discussed the relevance of information on prior tests (conducted before the index test).[82] A further 3 reviews Ferraro[116] (44% study participants in tertiary care setting), Zhen[117] (60% participants in tertiary care setting) and Meys[114] (3 LR2 studies-1 tertiary care and 2 mixed; 18 RMI studies 9/18 in secondary or mixed setting and unknown in 9/18)[114] included information on the clinical setting of included primary studies and provided stratified results according to ovarian cancer prevalence (< or > 25%).

Menopausal status, index test positivity threshold and target condition

Nine reviews stratified test accuracy estimates by menopausal status.[82, 88, 114, 118-122] Seven reviews provided stratified results by test positivity threshold.[82, 119, 121-124] Two reviews presented stratified results according to histological subtype (2)[65, 122] and 6 reviews explicitly reported how borderline ovarian tumours had been categorised for analysis.[82, 118, 120-122, 125, 126]

Important potential sources of heterogeneity for the accuracy of tests for ovarian cancer include population characteristics (presentation and tests received prior to the index test under evaluation), menopausal status, the test positivity threshold used for biomarkers and inclusion of all histological subtypes of ovarian cancer including borderline tumours. A total of 9 reviews did not report study characteristics or undertake investigations of heterogeneity for any of these variables.

3.3.3 Review quality and currency

Currency

The search dates in the included reviews covered the period from database inception[115] to 2016.[114] The search dates of the included reviews had a wide range depending on the data source; the search end dates for different reviews ranged from September 2004[82] to July 2015.[114] (see Table 3.2)

Database coverage

With the exception of 3 reviews,[88, 127, 128] 2 or more bibliographic databases were searched; Medline (15 reviews) and Embase (12 reviews) were the most commonly searched databases followed by the Cochrane library (11). Two reviews searched Pubmed only[88, 128] and 1 review[127] covered Medline alone. Thirteen of 23 reviews did not specify if unpublished literature was searched.

Table 3. 2: Review characteristics – quality.

Author and publication date	Index test(s) included			Search dates	Electronic databases	Study selection (PppIRT)	Quality assessment included studies (tool)	Synthesis employed hierar chical models unless not appropriate to review question or evidence base*
	Biomarkers	**Test combinations**	**Symptoms**					
E.R. Myers / Feb 2006	Yes	Yes	No	1969 - Sept 2004	Cochrane and MEDLINE	No prior tests	Yes (review specific tool)	No
P. Geomini / Feb 2009	No	Yes	No	Inception to Mar 2008	MEDLINE and EMBASE	PppR not reported	Yes (review specific tool)	Yes
L. R. Medeiros / Feb 2009	Yes	No	No	JAN 1985 - DEC 2007	MEDLINE, CANCERLIT, LILACS, EMBASE, Cochrane	Ppp not reported	Oxford evidence level Studies with level 4&5 excluded	No
NICE April /2011 (a) sym	Individual and combination			1806 – 2010	MEDLINE, PreMEDLINE, EMBASE, Cochrane, CINAHL, BNI, PsycINFO, Web of Science, BioMed Central	PppR Not reported	Yes QUADAS	No meta analyses. Forest plots of sensitivity and specificity
NICE April /2011 (b)Bio	CA125, HE4			1806 – 2010	MEDLINE, PreMEDLINE, EMBASE, Cochrane, CINAHL, BNI, PsycINFO, Web of Science, BioMed Central	Ppp not mentioned in methods but results summary notes that results do not directly apply to symptomatic women and timing of tests not reported (page 64)	Yes QUADAS	No
NICE April /2011 combi	Yes	Yes	Yes	Same	Same	PpR(presentation and prior test) After search SR for PE+CA125+ USS combined and Geomini for RMI and 1 additional updated study for RMI and combined tests results from different studies assuming conditional independence	Yes QUADAS	Not Done
S. Yu/ Feb 2012	Yes	No	No	1974-2011	PubMed, MEDLINE, EMBASE, EBSCO, Science Direct, Cochrane, CNKI	Ppp not reported	Yes QUADAS	No
J. E. Dodge/ Apr 2012	Yes	Yes	No	SRs and guidelines 1999-2009 and Jan 2004 - March 2009 ALL studies Myer's SR from 2004 updated	MEDLINE	Details not reported except Core methodology same as Meyer's	Yes AMSTAR for SR Review specific tool for primary studies	Yes

Table 3.2 Continued...

F. Li / June 2012	Yes	Yes	No	NO START DATE - UP TO 22/12/2011	MEDLINE, EMBASE, Web of Science, Google Scholar, Cochrane, Clinical Trials.gov	presentation and p prior test not reported	Yes QUADAS	Limited (univariate analysis was performed to determine if multivariate analysis appropriate based on heterogeneity (I2))
L. Wu/ Sept 2012	Yes	No	No	JAN 1990 - AUG 2011	PubMed	Ppp not reported	Yes QUADAS	No
J. Lin/ Nov 2012	Yes	No	No	SEPT 1995 - NOV 2011	MEDLINE, EMBASE, Cochrane	PppR not reported	Yes (Selected from QUADAS)	No
J.Lin/ Dec 2012	Yes	Yes	No	2009 - June 2012 (Studies before 2009 excluded due to progressive development in assay according to author)	EMBASE, MEDLINE, Cochrane	PPPRT not reported	Yes (Selected from QUADAS)	
S.Ferraro/ Feb 2013	Yes	No	No	Medline (since 1966) and Embase (since 1993) upto Jan 2012	MEDLINE, EMBASE,	Ppp not reported (Chinese women)	Yes QUADAS	2 studies for CA125 and 3 for HE4 were excluded as they were outliers
Z. Yang/ Jul 2013	Yes	No	No	JAN 2000 - MAY 2013	PubMed and EMBASE	PppR not reported	Yes QUADAS	Yes
J.Lin/ Dec 2013	Yes	Yes	No	Until June 2012	EMBASE, MEDLINE, Cochrane	PppR not reported	Yes (Selected from QUADAS)	No
J. Kaijser / Dec 2013	Yes	Yes	No	2008-2013 + added Models from Geomini	MEDLINE and EMBASE	Ppp reported	Yes QUADAS	Yes
J.Wang/ Mar 2014	Yes	Yes	No	Period NR; Only last date of search reported-15-Nov-13	MEDLINE, Science Direct, Cochrane	presentation and prior test not reported	Yes QUADAS	Yes
S. Zhen /Apr 2014	Yes	No	NO	2008-2012		PppT not reported	Yes QUADAS	No
X. Y. Liao / Nov 2014	Yes	No	No	NR	PubMed, WANFANG, CNKI	PppR not reported	NR	
A.C.L. Macedo / Sept 2014	Yes	No	No	1990-APRIL 2013	MEDLINE, EMBASE, Cochrane, Central Register of Controlled Trials, IBEX, BIOSIS, Web of Science, SCOPUS, Congress, Abstracts, Grey Literature	presentation and prior test not reported	Yes QUADAS	No

Table 3.2 Continued...

M. Stukan / Feb 2015	Yes	Yes	No	1984-April 2004	MEDLINE, EMBASE, CINAHL	Ppp not reported	Yes QUADAS	NA
M.Ebell / Mar 2016	No	No	Yes	2001-2014 (only provided in abstract not methods)	PubMed	Ppp not reported	Yes QUADAS	Yes
E.M.J. Meys/ May 2016	No	Yes	No	1990 - 31 JULY 2015	MEDLINE, EMBASE, Cochrane, Central Register of Controlled Trials	Prior test and presentation not reported	Yes QUADAS	Yes

Notes: PppIRT Population, prior tests, presentation, Index test & threshold, Reference Standard and Target condition defined (specified) in objectives or methods

*Meta-analyses in diagnostic test accuracy reviews require multivariate models (bivariate and hierarchical summary receiver operating characteristic), which allow for the trade off between sensitivity and specificity with changes in test positivity threshold and also for variability across studies by incorporating random effects. Univariate methods may be used if the review question is concerned with estimates of sensitivity at a fixed specificity or vice versa. Meta-analysis may be precluded by low volume or heterogeneity of the evidence base

Question formulation

Inclusion criteria were not clear or comprehensive in any of the (23/23) of reviews. Details regarding one of more of population presentation, prior testing, menopausal status, index test threshold and conduct, details of the target condition (histological subtypes including borderline tumours) and details on the methods used to verify of diagnosis (reference standard) were missing from all reviews. Only nine of the reviews [114, 121, 120, 82, 115, 118, 127] explicitly specified inclusion of both pre and postmenopausal women. One review restricted inclusion to only postmenopausal women.[129] No reviews provided information on presentation or prior testing leading up to inclusion in the study cohort.

Index test information provided in reviews was limited to the type of index test and with the exception of one study did not specify thresholds for inclusion.[126] However, 17/23 reviews reported test positivity information in their results.

Seven of 23 reviews did not define the reference standard in the inclusion criteria or provide details of the reference standard used in included studies (histology or duration of clinical follow up). Histological conformation of diagnosis for both index test positive and index test negative was the prerequisite inclusion criteria in the remaining 16 reviews. Histological subtypes or stages of ovarian cancer to be included were not defined in any review although 13/23 of the reviews reported some histological information in their results.

Quality assessment of studies included in reviews

Quality assessment of included studies was undertaken in 22 reviews with 15 of them using the QUADAS tool in its entirety and 2 using only a selection of items rather than the full tool.[124, 130] 6 reviews used a quality assessment tool not specific for test accuracy studies; 3 reviews used a review specific quality assessment tool[82, 125] and one AMSTAR[127] Liao et al was the only review[129] that did not undertake quality assessment of included studies.

Statistical synthesis

Meta-analyses in diagnostic test accuracy reviews require multivariate models (bivariate and hierarchical summary receiver operating characteristic), which not only allows for the trade off between sensitivity and specificity with changes in test positivity threshold but can also take

into consideration variability across studies by incorporating random effects. Univariate methods may be used if the review question is concerned with estimates of sensitivity at a fixed specificity or vice versa. Meta-analysis may be precluded by a low volume of evidence or heterogeneity. Twelve of the reviews did not use hierarchical models for meta-analysis without explanation. One review used hierarchical models in a limited way.[118] Three reviews in 2 publications[115, 122] provided reasons for not undertaking meta-analysis either because included reviews were confined to SRs or because of a low volume of included primary studies.

3.3.4 Quality of review reporting

Quality of reporting using the PRISMA DTA checklist is detailed below. The tables summarising the quality of reporting on the methods used and results reported for the included reviews are presented in appendix 4 and 5 respectively.

Review background

Eight reviews were published by European groups (1 from UK) and 8 from China (all biomarkers) and 4 from America (1 from USA, 1 from Canada and 2 from Brazil).

Twelve of 20 reviews identified themselves as a SR whilst 5 identified themselves as a meta-analysis. Two of the reviews not identifying themselves as a SR were guidelines.[82, 115]

The clinical background was poorly described by the majority of reviews including the role of the test, the potential effect of test results on management and the minimal acceptability level of test accuracy. The majority of reviews in their introduction and rationale for review are limited to only stating "a review of diagnostic tests to diagnose ovarian malignancy in women with adnexal mass". In summary 14 reviews were scored as 'no' and 4 as 'limited' for this section of the reporting tool. With the exception of one study[114] none of the protocols for any of the reviews were registered.

With the exception of one review[115] reporting of the review objectives was scored as "limited" or "no" due to the paucity of information about the population limited to "women".

Methods

The study eligibility criteria that should be reported by a test accuracy review include study characteristics: population, index test, target condition, reference standard and study design. Information on index test (23 scoring 'yes') and target condition (23 scoring 'yes') were consistently reported whilst information on population (14 scoring 'No'), reference standard (8 scoring 'No') and setting (20 scoring 'No') were poorly reported. Study design and language were not reported in 9 reviews. Only one review (NICE "biomarker reviews") reported all eligibility criteria whilst 2 reviews omitted to report one of the above items (NICE "symptoms review" (reference standard) and Myers review (setting). All other reviews had more than 1 eligibility item-missing ranging from 2 items (7 reviews) to 5 items (3 reviews).

Information sources and search

Sub items included for reporting checklist for this item included details of databases included, search dates including date of last search, search strategy, details regarding contact with authors and other sources and any search limits (such as date or study design) applied.

Some information regarding data sources, search period and key words were reported by all reviews. However this reporting item was scored as 'limited' in 19 reviews because the last search date (8 reviews), contact with authors (16 reviews) or both (4 reviews [116, 125, 124, 129] were missing. Search limits were not reported in 5 reviews. Only 8 reviews reported their search strategy whilst 1 other review made it available on request;[121] All sub items were reported in 2 reviews.[82, 115] One review[120] did not report any information and was scored 'no' for this item.

Study selection

All sub items of study selection were reported in 19/23 reviews. Three reviews were scored 'limited' as eligibility for meta-analysis was not reported and 1 review[122] did not report any of the sub items and was scored 'no'.

Data extraction

20/23 reviews were scored as 'Limited' on this item based on the omission of one of the sub items "piloting of data extraction form". Three reviews did not report any sub items and were

scored 'no'. [115, 122, 127] The NICE "combination review" was screened by a single reviewer and included only the SR by Geomini et al along with an other RMI study by Raza et al was regarding definition of terms for data extraction only 2 reviews scored 'yes'.[119, 114] The remaining reviews were scored as "limited" as one or more of the variables index test, target condition, reference standard and study design were undefined for data extraction; target condition was defined by the least number of reviews (17/23) and the reference standard by the most reviews (16/23).

Quality assessment

The quality assessment tool used was reported in 22/23 reviews.

Planning statistical analysis

Regarding definition of the unit of assessment for meta-analysis this was not explicitly stated in any of the reviews. Only one review[119] reported methods on how authors intend to group index tests for comparison. Only 2 reviews did not report methods for meta-analysis and accuracy measures to be derived [122] and the NICE "symptoms review".[115]

Fourteen reviews reported a-priori plans for additional analyses; 2/14 of the reviews reported plans for subgroup and sensitivity analyses and 12 reviews reported additional analyses for investigating heterogeneity, threshold effects and publication bias.

Results

All sub items of study selection were reported in 12 reviews and scored 'yes'. Stukan and Dodge et al[122, 127] did not report any sub items and scored "No". 11 of the reviews did not report a PRISMA diagram. In 3 reviews[114, 122, 117] the reason for excluding studies from the meta analysis was not provided.

Study characteristics

With respect to reporting study characteristics of included studies one review[120] scored "No" as none of the individual study characteristics (patient symptom status, prior test, study design, index test and threshold, target condition, reference standard, sample size and funding) were

reported. The remaining 22 reviews were scored as 'limited' mostly (18 reviews) due to the nonreporting of the sub items regarding patient symptomatology (18), prior testing (21) and funding (22). 13 reviews had 4 or more sub items unreported.

Quality of the primary studies included in the reviews

Six of the reviews reported both results of risk of bias and applicability assessment of individual studies and included a summary table or figure and were scored "Yes"; Three reviews reported a narrative summary of the quality of included studies without detail for individual studies in either narrative or figure form and were scored "No". 2 reviews reported the intention to do Quality assessment in their methods but did not report the quality assessment results. 9 reviews were scored "Yes" and 8 as 'Limited' either due to risk of bias or applicability judgements not being reported.

Numerical results

The PRISMA DTA numerical results items for quantitative analysis include reporting results of individual studies (TN, TP, FN, FP, diagnostic measure e.g. Sensitivity and specificity with confidence intervals, threshold for test positivity, ROC and/ or Forest plot), synthesis across studies (narrative summary or pooled estimates of meta analysis and confidence intervals) and any additional analyses performed or not performed with reasons.

17 reviews lacked 1 or more sub items relating to reporting of results for included studies. Sensitivity and Specificity, threshold for test positivity and ROC curve/Forest plot were not reported in 10, 12 and 3 reviews respectively.

With the exception of one study[122] the results of statistical synthesis were well reported (summary estimates of sensitivity and specificity with confidence intervals).

Additional analyses in reviews are undertaken for investigation of heterogeneity (sub group analyses or meta-regression) or to test the robustness of results (sensitivity analyses). It is often not possible to undertake planned additional analyses due to limitations of quality and / or volume of data in a review. In the case that reviews indicate a plan to undertake additional analyses in their methods the results of these should be reported or reasons provided for why they were not performed. Across included reviews reporting of the results of additional

analyses was scored as “Limited" (15 reviews) and “No” (8 reviews). Additional analyses that were reported included subgroup analyses (17 reviews), sensitivity analyses (6 reviews) and meta-regression (4 reviews). 9 reviews planned and reported additional analyses and 1 review[119] reported planned subgroup analysis by menopausal status but inability to conduct this due to paucity of studies and provided individual data of 2 studies, which were available. 1 review provided subgroup results of menopausal status for biomarkers without an a priori plan in the methods[82] and 3 reviews reported a plan to perform heterogeneity results but did not provide the results or the reasons for not performing it.[125, 127, 88]

None of the reviews reported the funding of the included studies and therefore lacking transparency in industry funded studies or conflict of interest. The funding source was declared in 11 of the reviews.

3.4 Discussions

Systematic reviews of diagnostic test accuracy undertaken outside of the Cochrane collaboration are increasing in number as a resource for evidence based diagnosis and are an important resource of guideline development. Poor quality reporting of DTA reviews limits assessment of their quality and applicability, which in turn can mislead the guidelines and their recommendations drawn from these reviews and hinder policy development and assessing the need for further research. McInnes et al. published the Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies-The PRISMA-DTA statement to address this deficiency in 2018 [131]. To date, there is no tool for the quality assessment of systematic reviews of test accuracy.

3.4.1 What is the applicability of existing SRs for OC diagnosis?

Judgment of the applicability of existing reviews for OC diagnosis was difficult to assess in a large number of included reviews. For the majority of reviews where pertinent information was not available (symptomology, prior tests received and healthcare setting), no reason for its omission was provided. Only a minority (2 reviews) cited difficulty as a result of lack of information in included studies. Based on information reported the studies included in the reviews largely included patient populations from specialist settings presenting with adnexal

masses, with no details provided on presentation (presence of absence of symptoms) and the point in the clinical pathway.

Ovarian Cancer often presents with symptoms or may be detected incidentally during investigation for other conditions. The presenting symptoms are variable, non-specific, often mimic bowel symptoms and range in intensity, duration and frequency. The nature of symptoms will influence the healthcare setting in which a woman first presents (primary care, emergency setting, and the specialist team they are referred to), tests that will have been done prior to using an index test and the prevalence and severity of disease at the time of index testing (the further along a referral pathway – the higher prevalence and greater proportion of more advanced disease is encountered). The performance of the same test is, therefore expected to vary according to the healthcare setting in which it is used. Absence of reporting of this information in systematic reviews precludes judgements about applicability to a particular healthcare setting and contributes to unknown sources of heterogeneity when undertaking meta-analyses. The importance of information about patient characteristics can be illustrated using included reviews that did report relevant information. Meys[114] presented stratified for prevalence less than and greater than 25% based on knowledge of the range of prevalence in included populations: Of 24 studies included in the analyses the lowest prevalence was 9.5%, 3 studies reported a prevalence less than 15% and 6 studies reported a prevalence over 40%. The prevalence of OC varies on the setting and menopausal status and there is no published literature alluding to it. In generalist settings the prevalence of ovarian cancer is in the region of 0.03%. The implications for making judgements about the applicability of the existing systematic review evidence presented here is clearly crucial.

The ratio of the prevalence of ovarian cancer to benign conditions in women presenting with symptoms suspicious for an adnexal mass increases with age and after attaining menopause. Ovarian cancer is a heterogeneous disease and tests are known to perform differently according to histological subtypes. Similarly, borderline ovarian tumours (of low malignant potential) can display features of benign or malignant tumours with either US testing or testing for tumour markers. Tests would therefore be expected to perform differently in populations where the distribution of tumour types is different, for example, pre and post menopause. A lack of information on different histotypes identified in included along with the absence of information on participant age and/ or menopausal status also impacts on the applicability judgements of an evidence base.

3.4.2 Quality: How reliable are the results of existing SRs for OC diagnosis

In the absence of DTA specific quality tool consensus criteria for systematic review conduct was taken from the critical appraisal tool for systematic reviews, AMSTAR, combined with best practice guidance from the Cochrane Diagnostic Test Accuracy Review Handbook.

Search strategy

Regarding the search strategy of a review it is important to know the dates and range of information searches interrogated and the use of search limits such as date, language or study design. The search period was provided in all the reviews allowing assessment of currency. With the exception of 3 reviews a minimum of 2 electronic databases and citation checking was provided by the other included reviews. The concept of publication bias is not well understood for DTA studies but current guidance advocates searching for unpublished literature in DTA reviews; however more than 50% of included reviews did not report searching for unpublished literature in their methods.

Question formulation (review this section)

Question formulation is a critical step for the review that follows as it provides clarity on the objectivity of the review by predefining what forms the basis for the inclusion criteria. It provides the framework and objectivity for including and excluding studies and provides transparency and quality control of included studies in the review. The question formulation in a DTA review typically includes PppITR so that it reflects and defines the inclusion criteria that allow one to determine applicability of the review and contextualises the results and conclusion drawn.

The relevance regarding applicability of the reviews to the relevant population in clinical practice could not be drawn due to the lack of information and clarity missing for details for inclusion criteria regarding population, presentation and prior testing. The review by Meyers et al. (AHRQ)[82] also commented on this. However, the review that followed do not provide further clarity or address this weakness.

Even though the type of index test was reported in all reviews, 39% of review did not prespecify test positivity thresholds to be included. As test accuracy varies with test positivity, if index test thresholds are not predefined there is a risk that thresholds will be chosen that optimise test performance in a particular population, therefore potentially introducing bias into the review. This also potentially introduces heterogeneity to results of the review when the meta analyses that follow do not take this into account and prespecify the plan to deal with multiple different thresholds from included studies. The reviews also in their results/ conclusion did not make any recommendation of the clinically useful threshold for testing; the results of accuracy are meaningless when the threshold at which it can be applied, and this precludes decision on what is an acceptable false positive or false negative for the patient, clinician or stakeholder.

Similarly, information on target condition was not defined and limited to broadly just stating ovarian cancer in majority of the reviews thus showing an inability to understand the heterologous nature of ovarian tumours and its attendant effect on test accuracy. The lack of predefined criteria on target condition and the following lack of predefined plan on how to deal with borderline tumours, different histotypes or stages also makes the results less useful as again the patients, clinicians or stakeholders cannot infer its applicability. With the exception of one review[82], the lack of clarity on the spectrum of ovarian tumours included (benign, borderline and malignant and different histotypes) considered (inclusion and investigation of heterogeneity) was unclear. Without information about the range of histological subtypes included in a review and how these have been dealt with for meta-analyses (for example have borderline tumours been excluded) or has review authors selectively included studies with histological subtypes in which the index test is known to perform better, for example, HE4 is suggested to perform better for EOC,[132] applicability cannot be determined and recommendations made.

The quality of the reference standard used in test accuracy studies is key to the validity of estimates of test accuracy generated. Poor quality reference standards may underestimate or overestimate test accuracy. Reference standards that share characteristics of the index test may be chosen to inflate the observed test accuracy of an index test. In addition, inclusion of more than one type of reference standard in reviews is a potential source of heterogeneity for estimates of test accuracy.

Quality assessment in reviews

Assessment of the quality of included studies in a review is necessary so that appropriate inferences can be made about the validity of review results and the impact of study quality on heterogeneity can be investigated as appropriate. With the exception of one review[129] review, authors undertook quality assessment of included studies. The original DTA-specific quality assessment tool, QUADAS (QUality Assessment of Diagnostic Accuracy Studies) tool was published in 2003[112] and was superseded by QUADAS-2 in 2011;[133] All reviews published after 2011 included QUADAS as the quality assessment tool. Six of the reviews undertaking quality assessment did not provide details and presented these findings only with a figure.

Statistical analysis

Because of the correlation between sensitivity and specificity induced by threshold effects, meta-analysis of test accuracy outcomes requires hierarchical models. Hierarchical models deal with statistical distributions at two levels. At the lower level, they model the cell counts in the 2×2 tables extracted from each study using binomial distributions. This level deals with sampling variability in individual studies and the correlation that exists between sensitivity and specificity. At the higher level, random study effects analysis is used to account for heterogeneity in test accuracy between studies. The Bivariate and HSROC hierarchical models were developed between 2001 and 2005 and have been shown to be equivalent under many circumstances.[134] Hierarchical models are now advocated over separate (univariate) analysis of sensitivity and specificity.[135] Univariate analysis fails to take into account the correlation between sensitivity and specificity and is known to underestimate both metrics. Despite all included reviews being published after 2006 (earliest Myers Feb 2006) Only 8 of the (insert number of reviews undertaking meta-analysis as denominator here) reviews undertaking meta-analysis used hierarchical models or justified why they were considered inappropriate meaning that summary estimates of test accuracy may not be valid.

3.4.3 Reporting quality

There was consistent poor reporting regarding the methodology employed for most of the sub items such as rationale for review, protocol registration, search strategy, pilot data forms, defining index test thresholds, minimally acceptable test accuracy, definition of DTA elements,

plan for results synthesis on handling differences in primary studies during meta-analysis and funding of individual studies. Information on database searched, period included, eligibility criteria and personnel involved in data extraction were reported with key details missing thus lacking clarity on population (menopausal status), Index test (assay method, threshold for test positivity) spectrum of target condition. Lack of apriori plan or poor reporting of a apriori plan on how to how handle differences in different thresholds, different definitions of the target condition, indeterminate results or other potential variation across studies contribute to potential source of bias in review conduct and impacts on applicability and accuracy of the results.

The poor quality of reporting was as also reflected in the reporting of results with lack of information on data extraction, data on presentation, prior testing, patient characteristics, target condition and reference standard. The results of the meta-analyses for SN, SP, CI and ROC were well reported but subgroup / sensitivity analyses and stratification of results by menopausal status, threshold cut off or borderline or staging are poorly reported.

Methodologically, this lack of information not only introduces potential sources of heterogeneity but the lack of transparency prevents comparing or reproducing these results, which basically infers to poor quality scientific evidence.

Stukan et al. reported study specific results of the accuracy of index tests that included combination tests and biomarkers in epithelial ovarian cancer compared to other histotypes when available in included studies. Fake Li et al[118] is the only review that undertook meta-analyses and reported a stratified result for epithelial ovarian cancer compared to other histotypes for the accuracy of HE4 (biomarker). Poor reporting on accuracy of results without reporting stratification and adequate assessment of quality and or heterogeneity affects accuracy, transparency and applicability.

Information is lacking regarding presentation and clinical setting and limited to women in surgical setting; this affects the prevalence and spectrum of disease, information crucial to assessment of the applicability of review results to clinical practice apart from affecting accuracy of the index tests

These generalist results also cannot be applied to the population in a meaningful way, thus contributing to the continuing debate and lack of consensus on the clinical utility of these results.

While piloting the PRISMA–DTA tool, it was also noted that methodological and topic specific clinical expertise is required for meaningful interpretation of questions which may limit its usability to stakeholders. Secondly, the word count in most journals could be a stumbling block and inclusion of this information in the appendix apart from providing transparency is also likely to help potential reviewers, stakeholders and future researchers to identify relevant information to avoid reduplication of work and concentrate resources where the gaps are identified.

A requirement by journals to adhere to the reporting items as minimal standard as is already the case for many journals for RCTs (CONSORT) and systematic reviews of interventions (PRISMA). Adherence to PRISMA-DTA as well as a greater use of web-based material accompanying publications would improve the reporting of DTA reviews and therefore their usability.

3.5 Conclusion

The many identified SR were variable in quality. The poor quality of reporting precludes, making judgement on generalisability and applicability and therefore underlining the need for a review, which is systematic and compare tests across the board.

The poor reporting and where this could be assessed, deficiencies in quality and lack of applicability to the planned reviews (biomarkers and combination) led to the conclusion that updating an existing review was not an option. The lack of information in the reviews could have been either due to lack of the concerned review or the primary studies, and hence an attempt was made to not only look for all relevant information in the primary studies but also provide transparency to the review process.

The gaps in current knowledge discovered during the review of reviews were used to refine the review question and define the PppITR inclusion criteria. The information distilled from review of reviews was used to inform the methodology and tailor the quality assessment questions of the planned review to identify gaps in review and report test accuracy results giving

consideration to menopausal status, test positivity thresholds, status of borderline tumours and improve the reporting applicability of the review to make it more clinically relevant.

The QUADAS 2 tool was tailored to evaluate the primary review to evaluate heterogeneity, transparency and applicability of the studies to enable assessment and also contextualise results to aid patient and stakeholders alike.

The newly published DTA reporting PRISMA guide will be used to help complete reporting.

CHAPTER 4: RESULTS OF COMBINATION REVIEW

4.1 Screening and selection of studies

The original primary searches and subsequent updates retrieved a total of 36,005 records after de-duplication. After reviewing titles and abstracts, 34,968 records were subsequently excluded. Full-text copies of 1,037 potentially relevant reports were obtained and screened for inclusion, of which 72 were deemed eligible for inclusion for combination test. Figure 4.1 shows the PRISMA diagram of the study selection process. Studies were excluded if they failed to meet one or more of the specified inclusion criteria with regard to study design, participants, target conditions or ability to derive a 2x2 table. In this systematic review index tests or test combinations that did not report test positivity threshold or, had been superseded (for example LR1) or that had not been used at some point in clinical practise were excluded.

The test combinations included in the review were RMI (versions 1-4), ROMA, LR2, ACOG and modifications of ACOG and ADNEX (Table 2.1). A total number of 72 studies were identified for inclusion in this review (see section quantity of evidence). 48 studies including 15542 women were identified investigating the accuracy of RMI; 31 studies including 12335 women for RMI I, 19 studies including 4989 women for RMI II, 19 studies including 5945 women for RMI III and 6 studies including 1971 women for RMI IV. 22 studies, including 6497 women, were identified investigating the accuracy of ROMA, 4 studies including 4888 women, were identified investigating the accuracy of LR2. Four studies, including 2787 women, were identified investigating the accuracy of 4 different versions of ACOG. Three studies, including 1301 women for the ACOG version 1, 3 studies including 837 women for ACOG version 2, 3 studies including 2123 women for ACOG version 3 incorporating CA125 and 1 study including 516 women for ACOG version 3 incorporating OVA1. One study, including 2124 women, was identified investigating the accuracy of ADNEX. Reason for full text study exclusions are detailed in Appendix 6.

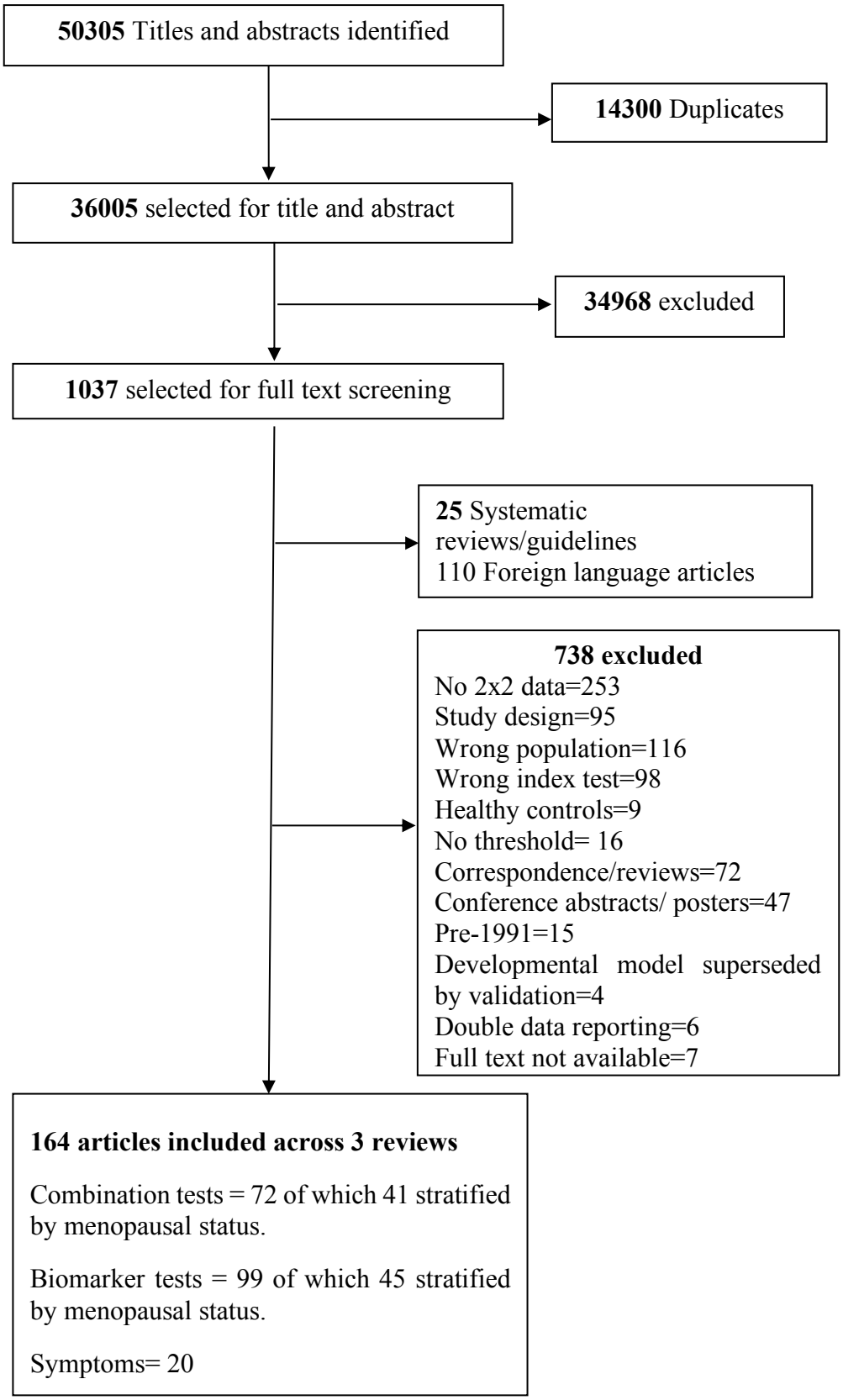


Figure 4. 1: PRISMA of included studies for the combination and biomarker review. See Appendix 6 for reasons of exclusion.

Forty-one of 72 included studies allowed separation of results by menopausal status and are included in the analysis. Details about included test combinations are detailed in Table 2.1, 4.1, 4.2, 4.3, 4.4 and 4.5. Where multiple publications were available for the same cohort of patients, publications reporting accuracy by menopausal status were chosen over non-differentiated results to reflect well-documented differences in spectrum in pre and postmenopausal women as detailed in the introduction chapter. Similarly, publications reporting accuracy on the basis of clinical assessment in primary care were preferred over those reporting accuracy on the basis of clinical assessment by gynaecologists in secondary care or gynaecological oncologists in tertiary care and specialist ultrasonographers. This was to reflect the setting in which index tests were to be used.

Seven of the included studies compared two index tests. Five patient cohorts reported in 6 studies compared ROMA with RMI [136, 137, 138, 139-141]. Two studies compared RMI with LR2 for both pre and postmenopausal women.[142, 143]

4.2 Study characteristics and quality of included studies

4.2.1 Study characteristics of included studies

The clinical pathways of patients from presentation to the decision/referral for surgical intervention were not detailed in any included studies. Testing prior to surgical intervention in this patient group will have included clinical history and examination, biomarker measurement and imaging. These tests can be carried out in primary care, by Gynaecologists/other specialities, by Gynaecological oncologists or a mixture of both. In clinical practice in the UK, clinical assessment and biomarker testing are most commonly initially performed in primary care. Ultrasound assessment may be initiated by a primary care practitioner but conducted in a secondary care setting. If the primary care practitioner decides to refer a woman for further management because of a suspicion of ovarian cancer, clinical assessment and ultrasound assessment or CT/MRI may be used as adjuncts to aid in decision-making and management plan in secondary or tertiary care.

The characteristics of all included studies, baseline characteristics of participants and the nature of the index test combination are described below under the headings RMI, ROMA, LR2, ACOG and ADNEX.

RMI and RMI variation studies

(Menopausal status, US findings and serum CA125). See also Table 2.1

The study characteristics of 48 studies, including 15542 women evaluating RMI for the diagnosis of ovarian cancer in pre and or postmenopausal women, are described in Table 4.1. Of 48 studies, 7 studies were multicentre, conducted within Europe and participating women were recruited mostly from mixed clinical settings (secondary or tertiary) [144, 142, 143, 145, 146, 147, 148] or from secondary care.[149] The remaining 41 studies were single centre, mostly conducted in secondary care (24 studies), 14 in tertiary car and in 3 studies the setting was not clear.[150, 151, 152] The majority of single centre studies (23/41) were conducted in Europe [153, 154, 150, 155, 156, 157, 158, 159, 160, 161, 162, 138, 163, 151, 164, 165, 166, 167, 137, 168, 169, 170, 140] 15 were conducted in Asia[171, 172, 173, 174, 175, 176, 177 177, 152,178, 179, 180, 181, 141, 182, 183] and one each were conducted in Canada[184,] Australia[185] and Brazil.[136]

Studies used various versions of RMI (1-4) and at different test positivity thresholds. RMI version 1 at threshold of 200 was the most commonly reported threshold (31/48 studies) followed by RMI version 2 at a threshold of 200 (19/48 studies) and RMI version 3 at a threshold of 200 (19/48 studies). RMI versions 1, 2 or 3 at thresholds of 250 were reported in 17 studies. RMI version 4 at a threshold of 400 or 450 was reported in 6 studies.[153, 178, 147, 182, 183, 149]

Table 4. 1: Study characteristics: RMI I.

Author Year Country	Setting	Participants characteristics	Index test threshold
Akturk 2011 **Turkey**	**Study criteria:** Women with pelvic mass undergoing laparotomy or laparoscopy **Clinical setting:** Secondary care **Prior tests:** Unclear **Exclusions:** Borderline tumours **Centre:** Single centre	**N:** 100 **Postmen n (%):** 16 (16) **Ovarian cancer n (%):** 18 (18) **Age:** Not reported **Separated by menopausal status:** No	**Threshold:** 200, 250 **Prespecified:** No
Anton 2012 **Brazil**	**Study criteria:** Women referred with pelvic mass diagnosed by US, 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 **Postmen n (%):** 73 (60) **Ovarian cancer n (%):** 30 (25) **Borderline n (%):** 17 (14) **Age mean:** -Malignant: 54.7 -Borderline: 56.4 -Benign: 50.7 **Separated by menopausal status:** Yes	**Threshold:** 200 **Prespecified:** Yes
Arun-Muthuvel 2014 **India**	**Study criteria:** women with pelvic tumours scheduled for surgery **Clinical setting:** Tertiary care **Prior test:** Unclear **Exclusions:** Nil **Centre:** Single centre	**N:** 467 **Postmen n (%):** 105 (22.4) **Ovarian cancer n (%):** 82 (18) **Borderline n (%):** 16 (3) **Age mean:** 39 **Separated by menopausal status:** No	**Thresholds:** 200, 250 **Prespecified:** No

Table 4.1 Continued...

Asif 2004 **India**	**Study criteria:** Women who underwent elective surgical exploration for ovarian mass **Clinical setting:** Secondary care **Prior test:** Unclear **Exclusions:** Unclear **Centre:** Single centre	**N:** 100 **Postmen n (%):** 30 (30) **Ovarian cancer n (%):** 55 (55) **Borderline n (%):** Not reported **Age mean:** - Malignant: 45 - Benign: 37 **Separated by menopausal status:** No	**Thresholds:** 200, 250 **Prespecified:** No
Aslam 2000 a **UK**	**Study criteria:** Women undergoing surgery for adnexal mass was examined preoperatively by US **Clinical setting:** Tertiary care **Prior test:** Unclear **Exclusions:** Nil **Centre:** Single centre	**N:** 61 **Postmen n (%):** 25 (41) **Ovarian cancer n (%):** 19 (31) **Borderline n (%):** 4 (7) **Age mean:** - Malignant: 53 - Benign: 43 **Separated by menopausal status:** No	**Thresholds:** 200 **Prespecified:** Yes
Bouzari 2011 **Iran**	**Study criteria:** Women with pelvic masses admitted for laparotomy **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** Not reported **Centre:** Single centre	**N:** 182 **Postmen n (%):** 21 (12) **Ovarian cancer n (%):** 23 (13) **Borderline n (%):** 8 (4) **Age mean:** 39.9 **Separated by menopausal status:** No	**Thresholds:** 200 **Prespecified:** No

Table 4.1 Continued...

Clarke 2009 **Canada**	**Study criteria:** Women with adnexal masses who underwent surgery **Clinical setting:** Tertiary **Prior test:** Unclear **Exclusions:** Nil **Centre:** Single centre	**N**: 163 **Postmen n (%):** 102 (63) **Ovarian cancer n (%):** 58 (36) **Borderline n (%):** 1 (1) **Age mean:** - Malignant: 57 - Benign: 54 **Separated by menopausal status:** No	**Thresholds:** 200 **Prespecified:** No
Davies 1993 **UK**	**Study criteria:** Women undergoing surgery for an adnexal mass **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** unclear **Centre:** Single centre	**N:** 124 **Postmen n (%):** 58 (47) **Ovarian cancer n (%):** 28 (23) **Borderline n (%):** 7 (6) **Age mean:** Not reported **Separated by menopausal status:** No	**Thresholds:** 200, 250 **Prespecified:** No
Di Legge 2012 **European countries**	**Study criteria:** Women who underwent surgery within 120 days after ultrasound examination for adnexal mass **Clinical setting:** Mixed **Prior test:** Unclear **Exclusions:** Nil **Centre:** Multicentre	**N:** 2445 **Postmen n (%):** 945 (39) **Ovarian cancer n (%):** 398 (16) **Borderline n (%):** 131 (5) **Age mean:** Not reported **Separated by menopausal status:** No	**Thresholds:** 200 **Prespecified:** Yes

Table 4.1 Continued...

Dotlic 2011 **Serbia**	**Study criteria:** adnexal masses hospitalised for surgery **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** Nil **Centre:** Single centre	**N:** 81 **Postmen n (%):** 30 (37) **Ovarian cancer n (%):** 30 (37) **Borderline n (%):** Not reported **Age mean:** 44.13 **Separated by menopausal status:** No	**Thresholds:** 200 **Prespecified:** Yes
Enakpene 2009 **Germany**	**Study criteria**: Women with suspicious adnexal mass undergoing surgery **Clinical setting:** Tertiary **Prior test:** Unclear **Exclusions:** Women with obvious clinical symptoms of malignant disease and ultrasound features of metastases were excluded Centre: Single centre	**N:** 302 **Postmen n (%):** 137 (45) **Ovarian cancer n (%):** 127 (42) **Borderline n (%):** 31 (10) **Age mean:** Not reported **Separated by menopausal status:** No	**Thresholds:** 250 **Prespecified:** Yes
Huchon 2012 **France**	**Study criteria:** Women who presented with an adnexal mass requiring surgery **Clinical setting:** Tertiary **Prior test:** Unclear **Exclusions:** Nil **Centre:** Single centre	**N:** 99 **Postmen n (%):** Not reported **Ovarian cancer n (%):** 6 (6) **Borderline n (%):** 5 (5) **Age mean:** 45.8 - Benign: 45.1 - Borderline: 52.3 - Malignant: 49.9 **Separated by menopausal status:** No	**Thresholds:** 200 **Prespecified:** No
Irshad 2013 **Pakistan**	**Study criteria:** Unclear (ovarian masses) **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** Unclear **Centre:** Single centre	**N:** 36 **Postmen n (%):** 36 (100) **Ovarian cancer n (%):** 24 (37) **Borderline n (%):** Not reported **Age mean:** 58 **Separated by menopausal status:** Yes	**Thresholds:** 250 **Prespecified:** Yes

Table 4.1 Continued...

Jabeen 2015 **Pakistan**	**Study criteria:** Women with adnexal mass undergoing surgery **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** only women over 30 were included **Centre:** Single centre	**N:** 60 **Postmen n (%):** Not reported **Ovarian cancer n (%):** 17 (28) **Borderline n (%):** 1 (2) **Age mean:** 50.8 **Separated by menopausal status:** No	**Thresholds:** 200 **Prespecified:** Yes
Jacobs 1993 **UK**	**Study criteria:** Women undergoing laparotomy for suspected adnexal mass Clinical setting: Secondary **Prior test:** Unclear **Exclusions:** Nil **Centre:** Single centre	**N:** 143 **Postmen n (%):** Not reported **Ovarian cancer n (%):** 37 (26) **Borderline n (%):** 4 (3) **Age mean:** Not reported **Separated by menopausal status:** No	**Thresholds:** 200 **Prespecified:** No
Kader Ali Mohan 2010 **Australia**	**Study criteria:** Women undergoing surgery for abdominopelvic mass **Clinical setting:** Tertiary **Prior test:** Unclear **Exclusions:** Only women over 30 included **Centre:** Single centre	**N:** 224 **Postmen n (%):** 124(55.3) **Ovarian cancer n (%):** 40 (18) **Borderline n (%):** 22 (10) **Age mean:** Not reported **Separated by menopausal status:** No	**Thresholds:** 200 **Prespecified:** Yes
Manjunath 2001 **India**	**Study criteria:** Women with a pelvic mass admitted for Laparotomy/ Laparoscopy **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** BOT excluded from analysis **Centre:** Single centre	**N:** 152 **Postmen n (%):** 43(65) **Ovarian cancer n (%):** 91 (60) **Borderline n (%):** (excluded from analysis) 4 (3) **Age :** Not reported **Separated by menopausal status:** No	**Thresholds:** 200, 250 **Prespecified thresholds:** No

Table 4.1 Continued...

Mol 2001 **Netherlands**	**Study criteria:** women who had surgery for adnexal mass **Clinical setting:** Unclear **Prior test:** Unclear **Exclusions:** Nil **Centre:** Single centre	**N:** 170 **Postmen n (%):** 61 (36) **Ovarian cancer n (%):** 30 (18) **Borderline n (%):** Not reported **Age mean:** Not reported **Separated by menopausal status:** No	**Thresholds:** 200 **Prespecified:** No
Moolthiya 2009 **Thailand**	**Study criteria:** women with pelvic mass admitted for laparotomy **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** only women >30 years included **Centre:** Single centre	**N:** 209 **Postmen n (%):** 88 (42) **Ovarian cancer n (%):** 60 (29) **Borderline n (%):** 11 (5) **Age mean:** 50 **Separated by menopausal status:** No	**Thresholds:** 200 **Prespecified:** Yes
Morgante 1999 **Italy**	**Study criteria:** Women over 30 years undergoing surgery for ovarian mass **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** Only women over 30 years included **Centre:** Single centre	**N:** 124 **Postmen n (%):** 55 (44) **Ovarian cancer n (%):** 27 (22) **Borderline n (%):** 2 (2) **Age mean:** Not reported **Separated by menopausal status:** No	**Thresholds:** 200 **Prespecified:** No
Obeidat 2004 **Jordan**	**Study criteria:** Women with a pelvic mass admitted for laparotomy **Clinical setting:** Tertiary **Prior test:** Unclear **Exclusions:** Unclear **Centre:** Single centre	**N:** 100 **Postmen n (%):** 73 (73) **Ovarian cancer n (%):** 56 (56) **Borderline n (%):** 16 (16) **Age mean:** Not reported **Separated by menopausal status:** No	**Thresholds:** 200, 250 **Prespecified:** No

Table 4.1 Continued...

Ong 2013 **Singapore**	**Study criteria:** Women admitted to the hospital for surgery due to ovarian mass **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** Nil **Centre:** Single centre	**N:** 228 **Postmen n (%):** 66(2.9) **Ovarian cancer n (%):** 6 (3) **Borderline n (%):** 11 (5) **Age mean:** Not reported **Separated by menopausal status:** No	**Thresholds:** 200 **Prespecified:** Yes
Radosa 2011 **Germany**	**Study criteria:** women with adnexal mass who subsequently underwent surgery were selected **Clinical setting:** Tertiary **Prior test:** Unclear **Exclusions:** Nil **Centre:** Single centre	**N:** 442 **Postmen n (%):** 141 (32) **Ovarian cancer n (%):** 79 **Borderline n (%):** 19 **Age mean:** 43.3 **Separated by menopausal status:** Yes	**Thresholds:** 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:** Nil **Centre:** Multi centre	**N:** 255 **Postmen n (%):** 117 (46) **Ovarian cancer n (%):** 48 (19) **Borderline n (%):** 18 (7) **Age mean:** 46 **Separated by menopausal status:** Yes	**Thresholds:** 200 **Prespecified:** Yes
Sayasneh 2013 (secondary publication)		**N:** 301 **Postmen n (%):** 116 (39) **Ovarian cancer n (%):** 65 (22) **Borderline n (%):** 18 (6) **Age mean:** 47 **Separated by menopausal status:** No	

Table 4.1 Continued...

Simsek 2014 **Turkey**	**Study criteria:** Women referred with adnexal mass and managed surgically Clinical setting: Tertiary **Prior test**: Unclear **Exclusions:** Nil **Centre:** Single centre	**N:** 569 **Postmen n (%):** 164 (29) **Ovarian cancer n (%):** 234 (41) **Borderline n (%):** 8 (1) **Age mean:** - Benign: 35.23 - Malignant: 50.78 **Separated by menopausal status:** No	**Thresholds:** 200 **Prespecified:** Yes
Terzic 2013 **Serbia**	**Study criteria:** Women treated for adnexal tumours **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** Nil **Centre:** Single centre	**N:** 689 **Postmen n (%):** 138 (20) **Ovarian cancer n (%):** 112 (16) **Borderline n (%):** 33 (5) **Age mean:** - Benign: 42.8 - Borderline: 53.6 - Malignant: 57.25 **Separated by menopausal status:** Yes	**Thresholds:** 250 **Prespecified:** Yes
Testa 2014 **European countries**	**Study criteria:** Women presenting with adnexal mass and undergoing TVS by one of the principal investigators and surgery within 120 days after examination **Clinical setting:** Mixed **Prior test:** Unclear **Exclusions:** Nil **Centre:** Single centre	**N:** 2403 **Postmen n (%):** 1049 (44) **Ovarian cancer n (%):** 701 (29) **Borderline n (%):** 153 (6) **Age mean:** Not reported **Separated by menopausal status:** Yes	**Thresholds:** 200 **Prespecified:** Yes

Table 4.1 Continued...

Tingulstad 1996 **Norway**	**Study criteria:** Women with pelvic mass undergoing laparotomy **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** women under 30 years were excluded **Centre:** Single centre	**N:** 173 **Postmen n (%):** 39 (27) **Ovarian cancer n (%):** 45 (26) **Borderline n (%):** 6 (3) **Age mean:** Not reported **Separated by menopausal status:** No	**Thresholds:** 200, 250 **Prespecified:** No
Ulusoy 2007 **Turkey**	**Study criteria:** Women undergoing surgery for an adnexal mass Clinical setting: Tertiary **Prior test:** Unclear **Exclusions:** Not reported **Centre**: Single centre	**N:** 296 **Postmen n (%):** 39.2 (49) Ovarian cancer n (%): 84 (28) **Borderline n (%):** 15 (5) **Age mean:** 42 **Separated by menopausal status:** No	**Thresholds:** 200, 250 **Prespecified:** Yes
Van Calster 2012 **European countries**	**Study criteria:** Women presenting with adnexal mass and undergoing surgery within 120 days **Clinical setting:** Mixed **Prior test:** Unclear **Exclusions:** Nil **Centre:** Multicentre	**N:** 1938 **Postmen n (%):** 742 (38) **Ovarian cancer n (%):** 472 (24) **Borderline n (%):** 111 (6) **Age mean:** Not reported **Separated by menopausal status:** No	**Thresholds:** 250 **Prespecified:** Yes

Table 4.1 Continued...

Van Gorp (IOTA) 2012 **Belgium**	**Study criteria:** Women with a pelvic mass, scheduled for surgery **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** Nil **Centre:** Single centre	**N:** 374 **Postmen n (%):**196 (52) **Ovarian cancer n (%):** 94 (25) **Borderline n (%):** 31 (8) **Age mean:** - Benign: 46.2 - Malignant: 57.7 **Separated by menopausal status:** Yes	**Thresholds:** 200 **Prespecified:** Yes
Van Holsbeke 2007 **Belgium, Sweden, Italy, UK**	**Study criteria:** Women with persistent adnexal mass undergoing surgery within 120 days **Clinical setting:** Mixed **Prior test:** Unclear **Exclusions:** Unclear **Centre:** Multicentre	**N:** 1066 **Postmen n (%):** 431 (40) **Ovarian cancer n (%):** 266 (25) **Borderline n (%):** Not reported **Age mean:** Not reported **Separated by menopausal status:** No	**Thresholds:** 200 **Prespecified:** Yes
Yamamoto 2009 **Japan**	**Study criteria:** Women with pelvic mass scheduled for laparotomy or laparoscopy **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** Nil **Centre:** Single centre	**N:** 253 **Postmen n (%):** 39 (15) **Ovarian cancer n (%):** 32 (13) **Borderline n (%):** 8 (3) **Age mean:** - Benign: 39.8 - Malignant: 54 **Separated by menopausal status:** No	**Thresholds:** 200, 250 **Prespecified:** No

Table 4.1 Continued...

Yamamoto 2015 **Japan**	**Study criteria:** Women with pelvic mass scheduled for laparotomy or laparoscopy **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** Nil **Centre:** Single centre	**N:** 296 **Postmen n (%):** 103 (35) **Ovarian cancer n (%):** 47 (16) **Borderline n (%):** 19 (6) **Age mean:** - Benign: 43.6 - Malignant: 56.3 **Separated by menopausal status:** No	**Thresholds:** 200 **Prespecified:** No
Yavuzcan 2013 **Turkey**	**Study criteria:** women with adnexal mass who underwent surgery Clinical setting: Unclear Prior test: Unclear **Exclusions:** Unclear **Centre:** Multicentre	**N:** 153 **Postmen n (%):** 54 (35) **Ovarian cancer n (%):** 24 (16) **Borderline n (%):** 8 (5) Age mean: 46 **Separated by menopausal status:** Yes	**Thresholds:** 200, 250 **Prespecified:** No

Thresholds extracted for RMI I: 200 and 250

Spectrum: presentation

With the exception of one study, all included studies recruited patients undergoing surgical investigation of an adnexal mass. In one study[138] women underwent surgery because of a 'high index of suspicion' for ovarian cancer, which may or may not have included the presence of an adnexal mass. This was the only RMI study where authors explicitly acknowledged that included women might have been symptomatic. However, the proportion of symptomatic women was not reported. Participants were recruited from a secondary care setting (general gynaecology unit), a tertiary care setting (specialist gynaecology oncology unit) or a mixed setting that included both general gynaecology and specialist gynaecology oncology units. None of the included studies recruited participants from a primary care setting. These highly selected populations are not representative of unselected, symptomatic women presenting in primary care where testing is initiated because of a suspicion of ovarian cancer.

Spectrum: target conditions included

The types of ovarian malignancy reported in included studies varied. Two studies reported including only epithelial ovarian cancer (EOC) and other benign ovarian tumours.[137, 141] Excluding certain tumour types changes the spectrum of disease in the study population, and this is known to affect estimates of test accuracy. For example, CA125 and HE4 are known to have a higher sensitivity in EOC compared to other types of ovarian tumour such as stromal and germ cell tumours. Similarly, 9 of the included studies either explicitly excluded those with borderline tumours [153, 174, 137] or did not identify borderline tumours in the study cohort.[150, 172, 158, 151, 166, 148]

Spectrum: age and menopausal status

1. Information on participant age was reported in 36/48 RMI studies. Twenty eight studies reported mean age which varied between 26.0 years[149] and 55.0 years.[184, 160, 166] Age range was reported in 16 studies and ranged from 10 years[178] to 94 years.[147, 144] Four of 48 RMI studies included women less than 16 years old among their participants.[171, 144, 178, 147] Nine studies restricted inclusion to women over 30 years of age.[154, 177, 185, 152, 180, 164, 169, 143, 146] Twelve studies did not report any information about the age of included participants.

Menopausal status is a risk factor for ovarian cancer. In addition, the spectrum of disease (the severity of ovarian cancer 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. The accuracy of tests is expected to vary between pre and post menopause and therefore I consider that stratifying test performance by menopausal status an important feature of studies. Only eight of 48 RMI studies stratified results by menopausal status.[136, 160, 165, 142, 137, 168, 143, 140]

ROMA studies

(Menopausal status, serum HE4 and CA125). See also Table 2.1

The study characteristics of 22 studies including 6489 women using ROMA for the diagnosis of ovarian cancer in pre and postmenopausal women are described in Table 4.2

Table 4. 2: Study characteristics: ROMA.

Author Year	Setting	Participant characteristics	Index test threshold*
Anton 2012 **Brazil**	**Study criteria:** Women presenting 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 centre	**N:** 120 **Postmen n (%):** 73 (60.8%) **Ovarian cancer n (%):** 30 (25%) **Borderline n (%):** 17 (14%) **Age mean:** -Malignant: 54.7 -Borderline: 56.4 -Benign: 50.7 **Separated by menopausal status:** Yes	**Thresholds:** -Premen 13.1; -Postmen 27.7 **Pre-specified:** Yes
Bandiera 2011 **USA**	**Study criteria:** Not reported **Clinical setting:** Tertiary care **Prior tests:** Not reported **Exclusions:** Non Epithelial Ovarian Cancer excluded **Centre:** Single	**N:** 278 **Postmen n (%):** 183 (65.8) **Ovarian cancer n (%):** 113 (41) **Borderline n (%):** Not reported **Mean age:** Premenopausal: -Malignant: 44.7 -Benign: 41.5 Postmenopausal: -Malignant: 66.3 -Benign: 64.0 **Separated by menopausal status:** Yes	**Thresholds:** -Premen 7.4; -Postmen 25.3 **Pre-specified:** Yes

Table 4.2 Continued...

Chan 2013 **Countries in Asia-Pacific region**	**Study criteria:** Women over 18 years diagnosed with adnexal mass diagnosed by any Imaging method (US, CT or MRI) **Clinical setting:** Unclear **Prior test:** Unclear **Exclusions:** Nil **Centre:** Multicentre	**N:** 414 **Postmen n (%):** 26 (108) **Ovarian cancer n (%):** 74 (18) **Borderline n (%):** 16 (4) **Age mean:** Not reported **Separated by menopausal status:** Yes	**Thresholds:** -Premen 7.4; -Postmen 25.3 **Pre-specified:** Yes
Chen 2015 **China**	**Study criteria:** Women with pelvic masses scheduled for surgery **Clinical setting:** Unclear **Prior test:** Unclear **Exclusions:** Nil **Centre:** Single	**N:** 232 **Postmen n (%):** Not reported **Ovarian cancer n (%):** 60 (26) **Borderline n (%):** Not reported **Age mean:** - Benign: 33 - Malignant: 53 **Separated by menopausal status:** Yes	**Thresholds:** - Premen 11.4 **Pre-specified:** Yes
Chen 2014 **China**	**Study criteria:** Women with EOC and Benign lesions **Clinical setting:** Tertiary **Prior test:** Unclear **Exclusions:** Women with non EOC excluded **Centre:** Single	**N:** 192 **Postmen n (%):** 84 (44) **Ovarian cancer n (%):** 123 (64) **Borderline n (%):** Not reported **Age mean:** not reported **Separated by menopausal status:** Yes	**Thresholds:** - Premen 12.2; - Postmen 25.8 **Pre-specified:** 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:** Excluded non EOC **Centre:** Single	**N:** 99 **Postmen n (%):** 31 (31) **Ovarian cancer n (%):** 43 (43) **Borderline n (%):** Not reported **Age mean:** - Benign: 39 -Malignant (EOC): 51 **Separated by menopausal status:** Yes	**Thresholds:** - Premen 11.5; - Postmen 25.5 **Pre-specified:** Yes

Table 4.2 Continued...

Fujiwara 2015 **Japan**	**Study criteria:** Women with adnexal mass diagnosed on Ultrasound/CT/MRI/PET and scheduled for surgery **Clinical setting:** Unclear **Prior test:** Unclear **Exclusions:** Women with previous history of any gynaecological disease and NON EOC excluded **Centre:** Multicentre	**N:** 221 **Postmen n (%):** 90 (41) **Ovarian cancer n (%):** 71 (32) **Borderline n (%):** 19 (9) **Age mean:** - Benign: 43 - Borderline: 47 - Malignant: 54 **Separated by menopausal status:** Yes	**Thresholds:** 8.8 (premen and postmen) **Pre-specified:** Yes
Grenache 2015 **USA**	**Study criteria:** Women with abnormal adnexal mass detected on physical examination and Imaging Included Ultrasound, CT or MRI) followed by surgery **Clinical setting:** Unclear **Prior test:** Unclear **Exclusions:** Unclear **Centre:** Multicentre	**N:** 146 **Postmen n (%):** 76 (52) **Ovarian cancer n (%):** 19 (13) **Borderline n (%):** 7 (5) **Age mean:** 52 **Separated by menopausal status:** Yes	**Thresholds:** - Premen 8.6, 13.1; - Postmen 27.7 **Pre-specified:** 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 **Exclusions:** Nil **Centre:** Single	**N:** 1218 **Postmen n (%):** 621 (51) **Ovarian cancer n (%):** 261 (21) **Borderline n (%):** 79 (6) **Age mean:** Not reported **Separated by menopausal status:** Yes	**Thresholds:** - Premen 7.4; - Postmen 25.3 **Pre-specified:** Yes
Kadija 2012 **Serbia**	**Study criteria:** Women diagnosed with adnexal mass scheduled to undergo surgery **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** Nil **Centre:** Single	**N:** 108 **Postmen n (%):** 41 (38) **Ovarian cancer n (%):** 24 (22) **Borderline n (%):** 5 (5) **Age mean:** Not reported **Separated by menopausal status:** Yes	**Thresholds:** - Premen 12.5; - Postmen 14.4 **Pre-specified:** No

Table 4.2 Continued...

Kim 2011 **S. Korea**	**Study criteria:** Women diagnosed with adnexal mass on the first visit to the gyn oncology clinic and underwent surgery **Clinical setting:** Tertiary **Prior test:** Unclear **Exclusions:** Only EOC included **Centre:** Single	**N:** 159 **Postmen n (%):** 108 (68) **Ovarian cancer n (%):** 68 (43) **Borderline n (%):** 10 (6) **Age mean:** - Benign: 35.7 - Malignant: 51.7 **Separated by menopausal status:** **Yes	**Threshold:** - Premen: 7.6 **Pre-specified:** Yes
Molina 2011 **Spain**	**Study criteria:** Not reported **Clinical setting:** Unclear **Prior test:** Unclear **Exclusions:** Nil **Centre:** Single	**N:** 396 **Postmen n (%):** 143 (36) **Ovarian cancer n (%):** 111 (28) **Borderline n (%):** Not reported **Age mean:** Not reported **Separated by menopausal status:** Yes	**Thresholds:** - Premen 13.1; - Postmen 27.7 **Pre-specified:** 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 **Centre:** Single	**N:** 104 **Postmen n (%):** 53 (51) **Ovarian cancer n (%):** 55 (53) **Borderline n (%):** Excluded **Age mean:** **-** Malignant: 56.9 - Benign: 42 **Separated by menopausal status:** Yes	**Thresholds:** - Premen 12.5; - Postmen 14.4 **Pre-specified:** Yes
Moore 2009 **USA**	**Study criteria:** Women with ovarian cyst scheduled to undergo surgery **Clinical setting:** Unclear **Prior test:** Unclear **Exclusions:** Nil **Centre:** Multicentre	**N:** 513 **Postmen n (%):** 150 (29) **Ovarian cancer n (%):** 143 (28) **Borderline n (%):** 22 (4) **Age mean:** 54 **Separated by menopausal status:** Yes	**Thresholds:** - Premen 13.1; - Postmen 27.7 **Pre-specified:** Yes

Table 4.2 Continued...

Moore 2011 **USA**	**Study criteria:** Women with ovarian cyst scheduled to undergo surgery **Clinical setting:** Mixed **Prior test:** Unclear **Exclusions:** Nil **Centre:** Multicentre	**N:** 472 **Postmen n (%):** 217 (46) **Ovarian cancer n (%):** 68 (14) **Borderline n (%):** 19 (4) **Age mean:** 50.3 **Separated by menopausal status:** Yes	**Thresholds:** - Premen 13.1; - Postmen 27.7 **Pre-specified:** Yes
Novotny 2012 **Czech Republic**	**Study criteria:** Women with pelvic abnormalities **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** premenopausal women excluded **Centre:** Single	**N:** 256 **Postmen n (%):** 256 (100) **Ovarian cancer n (%):** 21 (8) **Borderline n (%):** Not reported **Age mean:** - Benign: 65.28 - Malignant: 64.37 **Separated by menopausal status:** Yes	**Thresholds:** - Postmen 26.3 **Pre-specified:** No
Ortiz-Munoz 2014 **Spain**	**Study criteria:** Women with gynaecological symptoms, diagnosed with primary ovarian cancer **Clinical setting:** Tertiary **Prior test:** Symptoms **Exclusions:** Nil **Centre:** Single	**N:** 148 **Postmen n (%):** 104 (70) **Ovarian cancer n (%):** 29 (20) **Borderline n (%):** Not reported **Age mean:** not reported **Separated by menopausal status:** **Yes	**Thresholds:** - Premen 11.4; - Postmen 29.9 **Pre-specified:** 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 were excluded **Centre:** Single	**N:** 374 **Postmen n (%):** 276 (74) **Ovarian cancer n (%):** 108 (29) **Borderline n (%):** 45 (12) **Age mean:** Not reported **Separated by menopausal status:** **Yes	**Thresholds:** - Premen 17.3 - Postmen 26.0 **Pre-specified:** Yes

Table 4.2 Continued...

Sandri 2013 **Italy**	**Study criteria:** Women aged 18 years or older presenting to a gynaecologist undergoing surgery at a gyn oncology centre **Clinical setting:** Unclear **Prior test:** Unclear **Exclusions:** Nil **Centre:** Multicentre	**N:** 349 **Postmen n (%):** 191 (55) **Ovarian cancer n (%):** 160 (46) **Borderline n (%):** 25 (7) **Age mean:** 51.6 **Separated by menopausal status:** No	**Thresholds:** - Premen 7.4; - Postmen 25.3 **Pre-specified:** Yes
Van Gorp 2011 **(Van Gorp 2012- secondary publication- smaller cohort)** **Belgium**	**Study criteria:** All patients diagnosed with pelvic mass undergoing surgery **Clinical setting:** Unclear **Prior test:** Unclear **Exclusions:** Ni **Centre:** Single	**N**: 389 **Postmen n (%):** 161 (41) **Ovarian cancer n (%):** 161 (41) **Borderline n (%):** Not reported **Age mean:** - Benign: 46.3 - Malignant: 57.8 **Separated by menopausal status:** Yes	**Thresholds:** - Premen 12.5; - Postmen 14.4 **Pre-specified:** Yes
Winarto 2014 **Indonesia**	**Study criteria:** Women diagnosed with ovarian tumour by physical examination and TVS **Clinical setting:** Secondary care **Prior test:** Unclear **Exclusions:** Women with non EOC and unresectable tumours excluded **Centre:** Single	**N:** 128 **Postmen n (%):** 42 (33) **Ovarian cancer n (%):** 50 (39) **Borderline n (%):** 17 (13) **Age mean:** - Benign: 41 - Malignant: 44 **Separated by menopausal status:** No	**Thresholds:** - Premen 7.4; - Postmen 25.3 **Pre-specified:** No

Notes to table: *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. EOC: Epithelial ovarian cancer

Six of the 22 included ROMA studies were multicentre and conducted across Asia-pacific [186, 187], the USA [188, 189, 190,] and Italy.[191] The remaining 16 studies were single centre; the majority conducted in Europe[138, 137, 192 139, 193-197], 5 in Asia[198,141, 199, 200] and one each in the USA[201] and Brazil.[136]

In 7/22 studies women were recruited from secondary care.[136, 138, 195, 197, 200 141, 196] In 5 studies, women were recruited from tertiary care[194, 198, 201, 192, 137] and in one study, women were recruited from mixed hospital settings (secondary or tertiary).[190] In nine studies the hospital setting for recruitment (secondary or tertiary) was unclear.[186, 202, 199, 187, 188, 193, 189, 191, 139]

Of the 22 ROMA included studies, 14 provided results for premenopausal women and for postmenopausal women separately using different thresholds. Four studies only included accuracy data for premenopausal women [202, 187, 198, 194] and 2 studies only provided accuracy data for postmenopausal women.[196, 192] Two studies did not stratify accuracy by menopausal status.[191, 141]

In premenopausal women a test positivity threshold of 13.1 (+/-2) was most commonly used (13/18 studies)[136, 202, 199, 200, 188, 197, 193, 195, 189, 190, 194, 137, 139] followed by a threshold of 7.4 (+/-2) (5/18 studies).[201, 186, 187, 138, 198] In postmenopausal women, a test positivity threshold of 27.7 was used in 5/16 studies,[136, 188, 193, 189,] thresholds between 25.5 to 27.8 in 5/16 studies,[199, 200, 196, 192, 137] a threshold of 25.3 in 3/16 studies[201, 186, 138] and a threshold of 14.4 in 3/16 studies.[197, 195, 139]

Spectrum: presentations

All the ROMA studies included women with an adnexal mass scheduled for surgery for suspected ovarian cancer. Only three studies specified the presence of symptoms including 'gynaecological symptom's[194], pelvic pain[138] and vaginal bleeding, pain, distension and weight loss.[200] However, it is unclear whether the presence of symptoms prompted women's health seeking behaviour or whether the presence of symptoms was recorded following identification of an adnexal mass. The majority of studies (18/22) did not detail the imaging modality leading to the diagnosis of pelvic mass. Four studies reported that an adnexal mass was identified following investigation with one of either ultrasound, MRI or CT.[186, 136, 188, 187]

Spectrum: target conditions included

The types of ovarian malignancy reported in included studies varied. Eight studies included only EOC.[201, 199, 200, 187, 198, 195, 137, 141] In two other studies it was not clear if germ cell tumours were purposively excluded or not identified in the study cohort.[194, 188] Eight studies either excluded borderline tumours[195, 137] or did not identify borderline tumours in the study cohort.[201, 199, 202, 200, 193, 194, 196, 139]

Spectrum: age and menopausal status

Information on participant age was reported in 18/22 ROMA studies. Mean age varied from 42.5[141] to 65 years.[196] The age range of participants was reported in 12/18 studies and ranged across studies from 14 years[198] to 93 years.[196] Four studies did not report any information about the age of included participants.

A total of 18 studies provided test accuracy estimates for ROMA in premenopausal women only and 16 studies for postmenopausal women only. Only 2 studies did not stratify results by menopausal status.[141, 191]

LR2 studies

(Age, US findings). See also Table 2.1

The study characteristics of four studies including 4888 women[142, 143, 203, 204] in pre and or postmenopausal women are described in Table 4.3. Three of the studies were multicentre studies conducted across Europe[143, 204] and in the UK[142] where women were recruited from mixed clinical settings (secondary or tertiary care). One study was single centre and conducted in the UK in a secondary care setting.[203] All studies used LR2 at a test positivity threshold corresponding to a 10% probability of malignancy. For studies with overlapping data, the study with the most complete data was chosen after checking the studies for duplicate data.

Table 4. 3: Study characteristics: LR2.

Author Year	Study criteria and setting*	Participants characteristics	Index test threshold
Nunes 2013 **UK**	**Study criteria:** Women referred with US evidence of adnexal mass and undergoing surgery within 120 days **Clinical setting:** Secondary care **Prior tests:** Not reported **Exclusions:** None reported **Centre:** Single	**N:** 292 **Postmen n (%):** 137 (46.9) **Malignant n (%):** 90 (30.8) **Borderline n (%):** 17 (5.8) **Mean age (+/-SD):** 51 (not reported) **Median age (range):** Not reported (16-91) **Separated by menopausal status:** No	**Threshold:** 10% post test probability of malignancy **Prespecified:** Yes
Sayasneh 2013 (a) **Secondary study:** **Sayasneh 2013** **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 **Postmen n (%):** 117 (45.9) **Malignant n (%):** 48 (18.8) **Borderline n (%):** 18 (7.1) **Mean age (+/-SD):** 46 (not reported) **Separated by menopausal status:** Yes	**Threshold:** 10% post test probability of malignancy **Prespecified threshold:** Yes
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 **Postmen n (%):** 1049 (43.7) **Malignant n (%):** 701(18.8) **Borderline n (%):** 153 (6.4) **Median age (range):** -Malignant:57(33-66) -Benign: 44 (not reported) **Separated by menopausal status:** Yes	**Threshold:** 10% post test probability of malignancy Prespecified threshold: Yes
Timmerman 2010 **Secondary study:** **Di Legge 2012** **Eurpoe**	**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 **Postmen n (%):** 742 (38.0) **Malignant n (%):** 373 (19.2) **Borderline n (%):** 111 (5.7) **Mean age (+/-SD):** 46 (not reported) **Median age (range):** Not reported Separated by menopausal status: Yes	**Threshold**: 10% post test probability of malignancy Prespecified threshold: Yes

Spectrum: presentation

All LR2 studies included women with an adnexal mass scheduled for surgery.

Spectrum: target conditions included

All 4 studies reported a range of ovarian cancer tumour types following histology including borderline tumours. No studies explicitly excluded any type of ovarian cancer or benign ovarian pathology.

Spectrum: age and menopausal status

Information on participant age was reported for all 4 studies. The age range of participants ranged from 11 to 94 years.[143, 203, 204] One study only included women who were older than 30 years.[143] Where reported the mean age was similar across studies varying between 46 years and 51 years.[142, 203, 204]

Three of the 4 included LR2 studies stratified results by menopausal status[142, 143, 204] whilst one study[203] only reported test accuracy for all participating women regardless of menopausal status.

ACOG and ACOG variation studies

(Serum CA125 or OVA1, US findings and for ACOG v1 family history of breast or ovarian cancer). See also Table 2.1

The study characteristics of four studies (2787 women) using ACOG and variations of ACOG[205-208] for the diagnosis of ovarian cancer in pre and or postmenopausal women are described in Table 4.4.

All four ACOG studies were conducted in the USA. Three studies were multicentre and women were recruited from mixed clinical settings (secondary or tertiary care).[205, 207, 208] The test positivity threshold for ACOG is defined as the presence of any one of the variables being measured (positive family history of breast or ovarian cancer, raised biomarker level, any one ultrasound finding) The 4 included ACOG papers evaluated the accuracy of four different versions of ACOG. The first modification entailed changing the threshold of serum CA125

used in premenopausal women, the second modification entailed removing family history of ovarian or breast cancer and the third modification entailed replacing serum CA125 with OVA1 (see Table 4.4). Three of 4 ACOG studies evaluated more than one version of ACOG. The original version of ACOG was reported in three studies;[206-208] modified ACOG version 1 in three studies;[206, 208, 209] modified ACOG version 2 in three studies;[206-208] and modified ACOG version 3 in one study.[208]

Table 4. 4: Study characteristics: ACOG V3 (CA125) and ACOG V3 (OVA1).

Author Year	Study criteria and setting*	Participants characteristics	Index test version and threshold **
Bristow 2013 Secondary publications Bristow 2013a, Goodrich 2014, Longoria 2014 Ueland 2011 USA	**Study criteria:** Women undergoing surgery within 3 months of a documented pelvic mass on imaging (CT, US or MRI) **Clinical setting**: Mixed secondary and tertiary care **Prior tests:** Not reported **Exclusions:** None reported **(Longoria 2014**: Women with advanced disease, ascites and metastatic disease excluded**)** **Centre:** Multicentre	**N:** 770 **Postmen n (%):** 217 (43.9) **Malignant n (%):** 115 (15) **Borderline n (%):** 29 (3.8) **Mean age (+/-SD):** 49 (13.97) **Median age (range):** 48 (18-90) **Separated by menopausal status:** Yes	ACOG v3 (CA125)
Dearking 2007 USA	**Study criteria:** Women who fulfilled ACOG/SGO guidelines criteria for referral and undergoing surgery **Clinical setting:** Tertiary care **Prior tests:** Not reported **Exclusions:** Women aged < 35 **Centre:** Single	**N:** 837 **Postmen n (%):** 597 (71) **Malignant n (%):** 277 (33) **Borderline n (%):** Not reported **Mean age (+/-SD):** -Malignant: 64.9 (10.5) -Benign: 44.7 (4.9) **Age range:** (35-91) **Separated by menopausal status:** Yes	ACOG v3 (CA125)
Ware Miller 2011 USA	**Study criteria:** Women undergoing surgery within 3 months of a documented pelvic mass on imaging (CT, US or MRI) **Clinical setting:** Mixed secondary and tertiary care **Prior tests:** Not reported **Exclusions:** None reported **Centre:** Multicentre	**N:** 516 **Postmen n (%):** 281 (54) **Malignant n (%):** 115 (22) **Borderline n (%):** 28 (5) **Age range:** (18-92) **Separated by menopausal status:** Yes	ACOG v3 (CA125) ACOG v3 (OVA1)

* Setting: Secondary care: dedicated gynaecologist in a general hospital; Tertiary care – gynaecological oncology centre.

American College of Obstetrics and Gynaecologists (ACOG) v3 (CA125): **Premenopausal: 1 of CA125 >67; Ascites; abdo or distant mets/ pleural effusion. **Postmenopausal**: 1 of CA125 >35; Ascites; abdo or distant mets/ pleural effusion; nodular or fixed pelvic mass.

American College of Obstetrics and Gynaecologists (ACOG) v3 (OVA1): **Premenopausal**: 1 of: OVA1 positive; US: presence of ascites, abdominal or distant mets, pleural effusion. **Postmenopausal**: 1 of OVA1 positive; US: presence of ascites, abdominal or distant metastases, pleural effusion, nodular or fixed pelvic mass.

Threshold: any one positive test result.

Spectrum: presentation

All 4 ACOG studies included women with an adnexal mass scheduled for surgery.

Spectrum: target conditions included

All 4 studies reported a range of ovarian cancer tumour types following histology. One study either excluded borderline tumours or did not identify borderline tumours in the study cohort.[207]

Spectrum: age and menopausal status

Information on participant age was reported in all 4 studies and ranged from 18 years to 92 years. One study[206] restricted inclusion to women above 35 years old only. Mean age varied across studies in these studies between 49 and 52 years old.

All four included studies stratified results by menopausal status.

ADNEX studies

(Age, serum CA125, referral centre (tertiary or secondary), US findings). See also Table 2.1

One study evaluating ADNEX for the diagnosis of ovarian cancer in pre and or postmenopausal women[210] is described in Table 4.5. The ADNEX model was developed on 3506 patients recruited between 1999 and 2007, temporally validated on 2403 patients recruited between 2009 and 2012, and then updated on all 5909 patients. For the purposes of this review results were considered from validated data on 2403 participants. This was a multicentre study conducted across 10 countries in Europe. The study evaluated 4 test positivity thresholds: 3%, 5%, 10% and 15% probabilities of malignancy.

Table 4. 5: Study characteristics: ADNEX.

Author Year	Study criteria and setting*	Participants characteristics	Index test threshold
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 **Postmen 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

Spectrum: presentation

Women presenting with an adnexal mass scheduled for surgery were recruited from mixed settings (secondary and tertiary care).

Spectrum: target conditions included

The study reported a range of ovarian cancer tumour types following histology including borderline tumours. No type of ovarian cancer or benign ovarian pathology was explicitly excluded.

Spectrum: age and menopausal status

The study did not provide information on participant age in the validation data set (n=2403). Results were not stratified by menopausal status.

4.2.2 Methodological quality of included studies

The methodological quality of all studies evaluating one or more of RMI, ROMA, LR2 and ACOG and ADNEX studies is summarised in Figure 4.2. Separate Figures summarise study quality grouped by index tests RMI, ROMA, LR2 and ACOG (Appendix 7, Appendix 8,

Appendix 9, and Appendix 10. Methodological quality of the 41 studies reporting test accuracy data separately for premenopausal and postmenopausal women was not notably different to the 31 studies where accuracy was not reported separately according to menopausal status.

The data was extracted using a modified QUADAS-2 criteria proforma (Appendix 3) focusing on four domains of methodological quality: patient selection; index test; reference standard; and flow and timing. A comparative domain was completed for studies comparing different index tests. Tailoring of QUADAS-2 to the clinical topic required consideration of the following.

The target population was unselected women in whom a diagnosis of ovarian cancer was suspected. Thus, studies were considered to be at high risk of bias if they excluded certain types of malignant or benign pathology that might affect the accuracy of index tests specifically for detecting primary ovarian cancer. Examples include endometriosis (which for example causes raised CA125 levels) and borderline ovarian tumours (which although managed as malignant tumours in current clinical practice may not result in a positive test result). Additionally, restricting populations by age was considered to place studies at high risk of bias as age alters disease spectrum. For example, epithelial malignant tumours are more prevalent in older women and germ cell tumours in younger women and it is known that index test performance differs in different types of malignancy.[211, 212] The target population was unselected, symptomatic women in whom testing was initiated and conducted in a primary care setting and resulted in a suspicion of ovarian cancer. Thus, studies were considered to be at high risk of applicability in the presence of one or more of the following characteristics. If index tests were not performed on the basis of presentation and prior testing in primary care and / or if women were not symptomatic at presentation.

Menopausal status is a risk factor for ovarian cancer. In addition, the spectrum of disease (the severity of ovarian cancer 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. I therefore 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 review included a variety of index tests, themselves all comprised of several tests covering demographic information (such as age and family history of ovarian cancer), biochemical testing and ultrasound examination. Studies were considered at high risk of bias if ultrasound assessment (as a relatively subjective test) was not conducted blind to the results of other index test components (such as biochemical marker results). Similarly, where index tests included ultrasound as a component studies were considered at high risk of bias if the index test was not conducted and interpreted blind to the disease status/reference standard result. Studies that did not prespecify the test positivity threshold were also 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 composite index tests based on multivariable models (ACOG and ADNEX) additional factors resulting in a judgement of high risk of bias were whether all model components and thresholds were not prespecified or if individual test components were not assessed in a similar way (for example in similar healthcare settings / by individuals with similar levels of expertise). Assessment of applicability of index tests comprised consideration of the expertise of clinicians undertaking more subjective tests such as history taking, examination and ultrasonography.

Histological diagnosis or clinical follow up for a minimum of 6 months (in the absence of histological diagnosis) are both likely to correctly classify the target condition (therefore a low risk of bias) and therefore this was considered to be a valid reference standard for the purpose of this review. Most included studies were conducted in women with an adnexal mass scheduled for surgery and so histological diagnosis was available for all participants. In a minority of studies women with a negative index test result (considered to have a benign condition) were followed up. In these studies, risk of bias was considered high if follow up was less than 6 months and unclear if the duration of follow up was not reported. Concerning the applicability of the target condition /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 2x2 table and reflecting current clinical practice borderline tumours classified as malignant was considered as appropriate for the purposes of estimation of a test's accuracy in this review. Thus, studies either grouping borderline tumours with malignant tumours or reporting results in a way which allowed this grouping for the purpose of meta-analysis were considered to be of low applicability concern. Studies excluding borderline tumours from their analysis were considered as a high applicability concern as this supposes that borderline tumours would not be encountered in a typical patient population

being tested for ovarian cancer and is likely to lead to an overoptimistic estimation of a test's accuracy.

Concerning the flow and timing domain studies, risk of bias was increased if more than one reference standard (histology and follow up) was used in an individual study, if the interval between index test and reference standard application was > 3 months or if some patients were not included in the analysis.

Methodological quality of studies including RMI and variations of RMI

48 studies included an evaluation of the accuracy of RMI. The majority (39 of 48) RMI studies did not stratify women according to menopausal status. Appendix 7 summarise quality assessment for all 11 studies stratifying by menopausal status.

Risk of bias

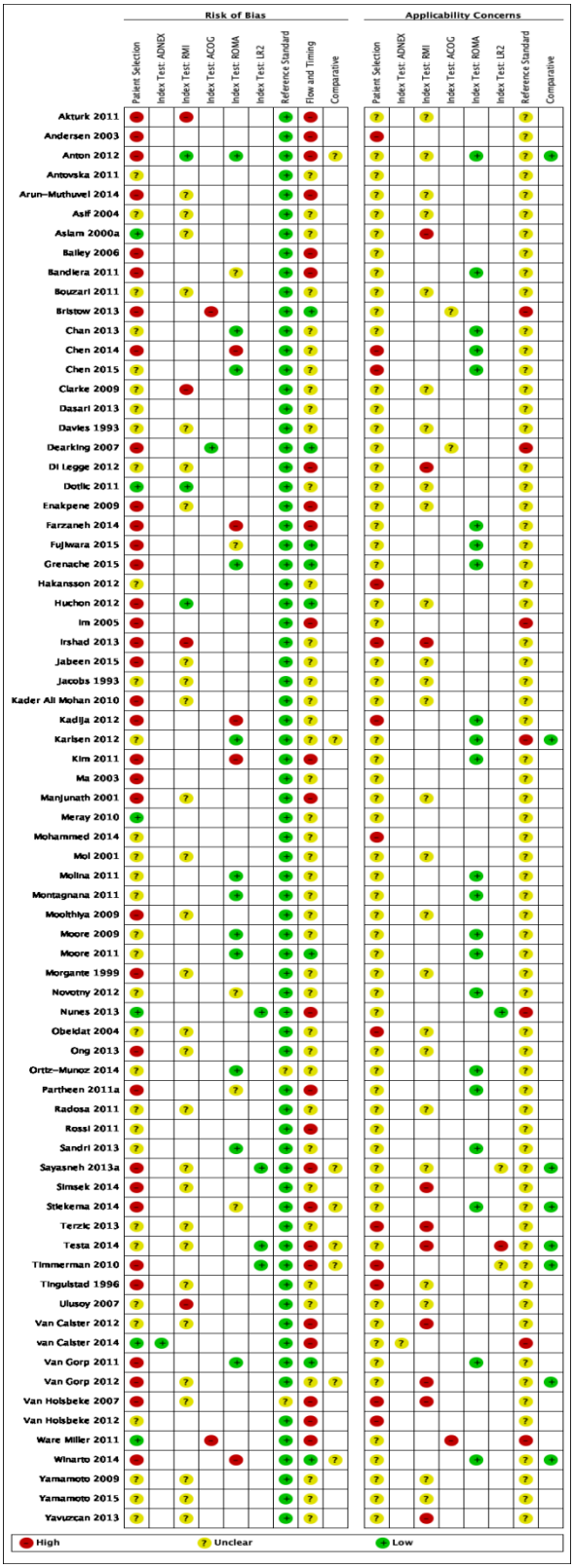
Patient recruitment was prospective in 18 studies,[150, 153, 155, 158, 160, 161, 171, 173, 138, 142, 151, 163, 166, 177, 140, 147, 168, 182] retrospective in 19 studies[144, 152, 157, 159, 175, 184, 159, 174, 178, 180, 146, 179, 181, 137, 167, 141, 143, 148, 183, 149] and was unclear in 11 studies.[156, 172, 176, 213, 162, 164, 165, 169, 170, 185]

For the participant selection domain 22/48 studies were judged at high risk of bias and 23 at unclear risk of bias. Only 3 studies were judged at low risk of bias[155, 158, 163] on the basis that authors explicitly reported consecutive sampling in a representative population (representative on the basis of reporting of patient characteristics (age range, menopausal status) and comprehensive listing of tumour types / pathology identified at histology. For studies rated as at unclear risk of bias this was most often because of a lack of an explicit mention of consecutive sampling in combination with a limited range of tumour types and benign pathology presented. For the index test and flow and timing domains the majority of studies were judged as at unclear risk of bias (40/48 unclear). The most common reason for an unclear risk of bias judgement in the index test domain was lack of clarity about blinding of the reference standard result to the index result. This is important because ultrasound forms a component of RMI, thus there is the potential for knowledge about true disease status to affect index test interpretation. Four of 48 studies were judged at high risk of bias either because of a lack of blinding of the index test to the reference standard or because of lack of prespecification of the index test positivity threshold. For the reference standard domain, the majority of studies

(47/48) were judged as low risk of bias. One study was judged at unclear risk of bias[154] because the length of follow up for 10% of participants was not reported. For the flow and timing domains the majority of studies were judged as at unclear risk of bias (31/48). Fifteen of 48 studies were judged at high risk of bias. The most common reason for an unclear risk of bias judgement in the flow and timing domain was lack of clarity of the interval between the conduct of the index test and the reference standard whereas those studies judged at high risk of bias in this domain did not include all patients in their analysis, for example excluding borderline tumours.

Risk of applicability

No studies were judged as having a low applicability concern for any of the participant selection, index test and reference standard domains. For participant selection this was due to the fact that study participants did not obviously represent symptomatic women presenting in primary care. For the index test domain this was due to fact that ultrasound was conducted by specialist sonographers or their level of specialism was unclear. The accuracy of US estimated from RMI studies could therefore not be assumed to be achievable in non-specialist, primary care settings. For the reference standard domain 47 of 48 RMI studies were judged as having an unclear applicability concern as it was unclear how borderline tumours had been handled for the paper's analysis. One study excluded borderline tumours from analysis and was judged to be of high applicability concern.[138]



Empty cells indicate that an index test was not evaluated by a study.

Figure 4. 2: Risk of bias and applicability concerns for 72 individual included studies for index tests: ADNEX: Assessment of Different NEoplasias in the adneXa model; RMI 1: Risk of Malignancy Index 1; ACOG: American College of Obstetrics and Gynaecology

Guidelines ; ROMA: The Risk of Ovarian Malignancy Algorithm; LR2: Logistic Regression 2 model.

Methodological quality of studies including ROMA

Twenty two studies included an evaluation of the accuracy of ROMA. Fourteen of 22 ROMA studies did not stratified women according to menopausal status. Appendix 8 summarise quality assessment for the 14 studies stratifying by menopausal status.

Risk of bias

Patient recruitment was prospective in 10 studies,[138, 186, 187, 197, 200, 139, 189-192] retrospective in 6 studies[188, 193, 194, 199, 137, 141] and unclear in 6 studies.[136, 195, 196, 198, 201, 202]

For the participant selection domain 12/22 were judged at high risk of bias and 10 studies at unclear risk of bias. The reason for unclear or high risk of bias in the patient selection domain was non-consecutive sampling and evidence of inappropriate exclusions or lack of clarity about these aspects of study conduct. For the index test domain 5 studies were judged as at high risk of bias because of lack of prespecification of the index test positivity threshold and three studies were judged at unclear risk because this was unclear. The majority of studies (21/22) were judged as low risk of bias for the reference standard domain and one study at unclear risk[194] because the duration of follow up was not reported for 15% of participants. For the flow and timing domains the majority of studies were judged as at high (6/22) or unclear (11/22) risk of bias because of exclusion of participants from analysis and a lack of clarity of the interval between the conduct of the index test and the reference standard respectively.

Risk of applicability

All studies were judged as being of unclear applicability concern for the patient selection domain due to the fact that study participants did not obviously represent symptomatic women presenting in primary care. For the index test domain all studies were judged as being of low applicability concern because all components of ROMA are biomarker tests and not likely to be subject to the influence of operator experience and skill. All studies were judged as being of unclear applicability in the reference standard domain because it was how borderline tumours had been handled for the paper's analysis.

Methodological quality of studies including LR2

Four studies included an evaluation of the accuracy of LR2. Only 1 of the 4 LR2 studies[203] did not stratify women according to menopausal status. Appendix 9 summarises quality assessment for all 4 studies.

Risk of bias

Patient recruitment was prospective in 3 studies[203, 142, 204] and retrospective in 1 study.[143]

For the participant selection domain 2 studies were judged at high risk of bias[142, 204] and 1 study was judged as at unclear risk of bias[143] as a result of non consecutive sampling and evidence of inappropriate exclusions or lack of clarity with respect to this aspect of study conduct. One study was judged to be at low risk of bias in the participant selection domain.[203] All studies were judged as low risk of bias for the index test and reference standard domains. All 4 studies were judged at high risk of bias for the flow and timing domain on the basis that not all study participants were included in the analysis and the interval between conduct of the index test and reference standard was not within 3 months.

Risk of applicability

For the patient selection domain 1 study was judged as being of high concer[204] and the remaining studies of unclear concern[142, 143, 203] due to the fact that study participants did not obviously represent symptomatic women presenting in primary care. For the index test domain 1 study was judged as being of high concern because US was performed by specialists,[143] 2 of unclear concern because it was unclear whether ultrasonographers were specialists[142, 204] and therefore the performance of US could not be assumed to be applicable to that achievable in primary care. One study was of low concern because ultrasonographers were not specialists.[203] For the reference standard domain 1 study was judged as being of high concern because borderline tumours were excluded from the study[203] and in the remaining 3 studies it was unclear whether borderline tumours had been excluded and if not excluded how they had been classified in the paper's analysis.[142, 143, 204]

Methodological quality of studies including ACOG and variations of ACOG

Four studies included an evaluation of the accuracy of ACOG. All 4 of the ACOG studies stratified women according to menopausal status. Appendix 10 summarises quality assessment for all 4 studies.

Risk of bias

Patient recruitment was prospective in 1 study,[208] retrospective in 2 studies[207, 209] and unclear in 1 study.[206]

Three of 4 studies were judged at high risk of bias for the participant selection domain[206, 207, 209] with the remaining study judged to be of unclear risk[208] as a result of non-consecutive sampling and evidence of inappropriate exclusions. For the index test domain 2 studies were judged at high risk of bias[208, 209] because individual components of the multivariable model were not prespecified. One study was judged at unclear risk of bias[207] because it was not clear if index test results were interpreted without knowledge of disease status. One study was judged at low risk of bias in the index test domain.[206] All studies were considered at low risk of bias for the reference standard domain. For the flow and timing domain 2 studies were judged at high risk of bias[207, 208] because not all participants were included in the analysis and 2 studies were judged to be at low risk of bias.[206, 209]

Risk of applicability

For the patient selection domain all studies were judged as being of unclear concern due to the fact that study participants did not obviously represent symptomatic women presenting in primary care. For the index test domain 1 study was judged as being of high concern[208] and the remaining 3 studies of unclear concern[206, 207, 209] because US was not performed by non specialised sonographers or this was not clear and therefore the performance of US could not be assumed to be applicable to that achievable in primary care. For the reference standard domain all studies were judged as being of high concern because borderline tumours were excluded from analysis.

Methodological quality of the single ADNEX study

Only one study was identified which used the ADNEX model.[210]

Risk of bias

This prospective study did not stratify women according to menopausal status. Figure 4.2 includes quality assessment for this single study.

The study was judged at unclear risk of bias for patient selection because the method of patient recruitment was not clearly reported. For the index test, reference standard and index test domains the study was judged at low risk of bias. For the index test domain this judgement was importantly based on the fact that all components and thresholds of the nine individual test components of ADNEX were prespecified and assessed in a similar way (for example similar healthcare setting) for all participants.

Risk of applicability

The study was of unclear concern in the patient domain on the basis that study participants did not obviously represent symptomatic women presenting in primary care. The index test domain was judged as of unclear concern on the basis that the skill of the sonographer undertaking US examinations was not reported and therefore the performance of US could not be assumed to be applicable to that achievable in primary care. The study was judged as being at low concern in the reference standard domain because borderline tumours could be grouped with malignant tumours for the purpose of analysis.

4.3 Results of meta-analyses

4.3.1 Test positivity thresholds used in included studies

For RMI I accuracy was estimated at 2 different thresholds (200 and 250), for RMI II at two thresholds (200 and 250), for RMI III at two thresholds (200 and 250) and for RMI IV at two thresholds (400 and 450). For ROMA accuracy was estimated at three different test positivity pairs (7.4 +/- 2 for premenopausal women and 25.3 +/- 2 for postmenopausal women; 12.5 for premenopausal women and 14.4 for postmenopausal women; 13.1 +/- 2 for premenopausal

women and 27.7 +/- 2 for postmenopausal women). For LR2 accuracy was estimated at a threshold to achieve a post test probability of malignancy of 10%. For ACOG, referral of women with an adnexal mass to specialist care is recommended if either CA125 is raised (threshold menopausal status dependent) or any one of a list of US features suggestive of malignancy are present. For ADNEX estimation of accuracy was reported at 4 different thresholds to achieve post test probabilities of malignancy of 3%, 5%, 10% and 15%.

4.3.2 Accuracy in premenopausal versus postmenopausal women

Where data allowed, for individual index test versions at defined test positivity thresholds, accuracy was compared between premenopausal women and postmenopausal women. This comparison was possible for RMI I (thresholds of 200 and 250), RMI III (threshold of 200), LR2, ROMA (all 3 threshold pairs), ACOG version 1, ACOG version 3 (incorporating CA125) and ACOG version 3 (incorporating OVA1) (Appendix 11).

In the populations of women included in this review a consistent difference was observed in sensitivity (higher in postmenopausal women) and specificity (lower in postmenopausal women) across all versions of all index tests at all thresholds analysed. For RMI I at a threshold of 200 (5 studies) and 250 (3 studies) sensitivity was significantly higher and specificity significantly lower for postmenopausal women compared to premenopausal women. RMI I at a threshold of 200: sensitivity post compared to premenopausal ROR 2.71 (95% CI 1.92 to 3.83) p =0.0001 and specificity post compared to premenopausal ROR 0.42 (95% CI 0.19 to 0.94) p=0.03; RMI I at a threshold of 250: sensitivity post compared to premenopausal [ROR 3.88 (95% CI 1.61 to 9.34)] p=0.003 and specificity post compared to premenopausal [ROR 0.52 (95% CI 0.30 to 0.90)] p= 0.02. For LR2 (3 studies) sensitivity was significantly higher and specificity significantly lower for postmenopausal women compared to premenopausal women; sensitivity post compared to premenopausal [ROR 3.02(95% CI 2.04 to 4.49) p=0.0001] and specificity post compared to premenopausal [ROR 0.21(95% CI 0.11 to 0.43) p=0.0001].

Subsequently, where study reporting allowed, sensitivity and specificity for all index test versions and all test positivity thresholds were estimated separately for premenopausal women only and postmenopausal women only. Estimates of test accuracy for pre and postmenopausal

women combined for the 41/ 72 studies where 2x2 tables could not be extracted separately according to menopausal status are also presented in Appendix 12.

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 ovarian cancer, it would be reasonable to expect tests to perform with lower specificity in premenopausal women compared to postmenopausal women, reflecting the greater possibility of false test positives from physiological alterations in premenopausal women that cause raised CA125 and ovarian cysts (e.g. the menstrual cycle) and conditions such as endometriosis that are commoner in premenopausal women.

Test accuracy is described based on estimates for premenopausal separate to postmenopausal women for those index test versions considered relevant to clinical practice and at test positivity thresholds relevant to clinical practice, or, in the absence of clinical consensus those thresholds most commonly reported in included studies are presented in Table 4.6 and Figure 4.6, Figure 4.7, Figure 4.8, Figure 4.9, Figure 4.10. Test versions and thresholds described include RMI I at thresholds of 200 and 250; ROMA at each of 3 thresholds pairs reported in included studies: 7.4 +/- 2 for premenopausal women and 25.3 +/- 2 for postmenopausal women; 12.5 for premenopausal women and 14.4 for postmenopausal women; 13.1 +/- 2 for premenopausal women and 27.7 +/- 2 for postmenopausal women); LR2 at a threshold to achieve a post test probability of malignancy of 10%, ACOG v3 (incorporating CA125), ACOG v3 (incorporating OVA1) and ADNEX at thresholds to achieve a post test probability of malignancy of 5% and 10%.

Table 4. 6: Sensitivity, False Negatives, Specificity and False positives of RMI I, ROMA, LR2, ACOGv3 (CA125), ACOGv3 (OVA1) and ADNEX.

Test	Threshold / Version	Women, Cases (%) (Studies)		Sensitivity (95% CI)		False Negative rate N (%) of cancers missed for every 100 women with ovarian cancer		Specificity (95% CI)		False Positives rate N (%) of women tested/treated unnecessarily for every 100 women without ovarian cancer	
		Premenopause	Postmenopaue	Premenopause	Postmenopause	Premenopause	Post menopause	Pre menopause	Post menopause	Pre menopause	Postmenopause
RMI 1	**200**	2634, 502 (19%) (5)	1879, 862 (46%) (5)	52.2 (45.9, 58.5)	75.0 (69.5, 79.8)	48 (41 to 54)	25 (20 to 30)	95.4 (92.5, 97.3)	90.1 (83.1, 94.4)	5 (3 to 7)	10 (5 to 17)
	250	356, 31 (9%) (1)	220, 97 (44%) (2)	54.8 (36.0, 72.7)	82.5 (73.6, 88.8)	45 (27 to 64)	17 (11 to 26)	88.3 (84.3, 91.6)	79.7 (71.6, 85.9)	12 (8 to 16)	20 (14 to 28)
ROMA	**7.4, 25.3**	1269, 173 (14%) (6)	1396, 544 (40%) (7)	82.6 (71.6, 89.9)	92.2 (87.4, 95.2)	17 (10 to 18)	8 (5 to 13)	80.0 (68.4, 88.1)	81.1 (71.8, 88.8)	20 (12 to 32)	19 (11 to 28)
	12.5, 14.4	302, 68 (23%) (3)	299, 177 (59%) (3)	63.5 (51.0, 74.4)	88.0 (80.6, 92.8)	36 (26 to 49)	12 (7 to 19)	89.3 (80.8, 94.3)	68.3 (57.4, 77.5)	11 (6 to 11)	32 (22 to 43)

Table 4.6 Continued...

	13.1, 27.7	1487, 307 (21%) (13)	1497, 610 (41%) (9)	78.3 (71.2, 84.0)	90.7 (86.4, 93.7)	22 (16 to 29)	9 (6 to 14)	83.3 (78.2, 87.5)	79.1 (74.6, 83.1)	17 (12 to 22)	21 (17 to 25)
LR2	**PTProb OC 10%**	2715, 588 (22%) (3)	1861, 1008 (54%) (3)	83.6 (78.6, 87.6)	93.9 (92.3, 95.3)	16 (12 to 21)	6 (5 to 8)	90.4 (81.3, 95.3)	63.9 (60.6, 67.0)	10 (5 to 19)	36 (33 to 39)
ACOG	**V3 CA125**	899, 150 (17%) (3)	1224, 486 (40%) (3)	78.0 (70.4, 84.1)	84.6 (76.0, 90.5)	22 (16 to 30)	15 (9 to 24)	83.9 (65.1, 92.6)	78.1 (71.0, 83.9)	16 (7 to 35)	22 (16 to 29)
	V3 OVA1	235, 45 (19%) (1)	281, 116 (41%) (1)	91.1 (78.8, 97.5)	94.8 (89.1, 98.1)	9 (2 to 21)	5 (2 to 11)	43.2 (36.0, 50.5)	25.5 (19.0, 32.8)	57 (49 to 64)	74 (67 to 81)
ADNEX	**PTProb OC 5%**	1354, 378 (28%) (1)	1049, 602 (57%) (1)	97.6 (95.5-98.7)	98.8 (97.6-99.4)	2 (1 to 4)	1 (1 to 2)	69.5 (66.5-72.3)	37.4 (33.0-41.9)	30 (28 to 33)	63 (48 to 67)
	PTProb OC 10%	1354, 378 (28%) (1)	1049, 602 (57%) (1)	94.7 (92.0-96.6)	97.7 (96.1-98.6)	5 (3 to 8)	2 (1 to 4)	78.6 (75.9-81.1)	55.5 (50.9-60.0)	21 (19 to 24)	44 (40 to 49)
	PTProb OC 15%	1354, 378 (28%) (1)	1049, 602 (57%) (1)	90.5 (87.1-93.0)	96.5 (94.7-97.7)	11 (7 to 13)	3 (2 to 5)	83.4 (80.9-85.6)	63.5 (59.0-67.9)	16 (14 to 18)	36 (32 to 41)

Threshold – ROMA was reported at threshold pairs of 7.4 (+/-2) pre and 25.3 (+/-2) postmenopausal; 12.5 pre and 14.4 postmenopausal; 13.1 (+/-2) pre and 27.7 (+/-2) postmenopausal. PTProb OC: Test threshold used in studies included in the review to achieve the described post test probability of ovarian cancer.

Accuracy of RMI I

RMI I premenopausal women (threshold 200 and 250). (See Figure 4.3)

RMI1 threshold 200 in premenopausal women (5 studies, 2634 women of whom 502 had a diagnosis of ovarian cancer).

The sensitivity of RMI1 in premenopausal women at a threshold of 200 was 52.2% (95%CI 45.9 to 58.5). For every 100 women with an actual diagnosis of ovarian cancer an estimated 52 (46 to 59) would have their cancer identified whilst an estimated 48 (41 to 54) would have their cancer missed (test negative in error).

The specificity of RMI1 in premenopausal women at a threshold of 200 was 95.4% (95% CI 92.5 to 97.3). For every 100 women without an actual diagnosis of ovarian cancer, 95 (93 to 98) would correctly be identified as negative by the test whilst 5 (2 to 7) would receive a positive diagnosis of ovarian cancer in error (a false positive result).

RMI1 threshold 250 premenopausal women (1 study, 356 women of whom 31 had a diagnosis of ovarian cancer):

Based on a single study the sensitivity of RMI1 in premenopausal women at a threshold of 250 was 54.8% (95% CI 36.0 to 72.7). For every 100 women with an actual diagnosis of ovarian cancer an estimated 55 (36 to 73) would have their cancer identified whilst an estimated 45 (27 to 64) would have their cancer missed (test negative in error).

The specificity of RMI I in premenopausal women at a threshold of 250 was 88.3% (95% CI 84.3 to 91.6). For every 100 women without an actual diagnosis of ovarian cancer 88 (84 to 92) would correctly be identified as negative by the test whilst 12 (8 to 16) would receive a positive diagnosis of ovarian cancer in error (a false positive result).

RMI I postmenopausal women (threshold 200 and 250)

RMI1 threshold 200 in postmenopausal women (5 studies, 1879 women of whom 862 had a diagnosis of ovarian cancer).

The sensitivity of RMI1 in postmenopausal women at a threshold of 200 was 75.0% (95% CI 69.5 to 79.8). For every 100 women with an actual diagnosis of ovarian cancer an estimated 75 (70 to 80) would have their cancer identified whilst an estimated 25 (20 to 30) would have their cancer missed (test negative in error).

The specificity of RMI1 in postmenopausal women at a threshold of 200 was 90.1% (95% CI 83.1 to 94.4). For every 100 women without an actual diagnosis of ovarian cancer, 90 (83 to 95) would correctly be identified as negative by the test whilst 10 (52 to 17) would receive a positive diagnosis of ovarian cancer in error (a false positive result).

RMI1 threshold 250 in postmenopausal women (2 studies, 220 women of whom 97 had a diagnosis of ovarian cancer).

The sensitivity of RMI1 in postmenopausal women at a threshold of 250 was 82.5% (95% CI 73.6 to 88.8). For every 100 women with an actual diagnosis of ovarian cancer an estimated 83 (74 to 89) would have their cancer identified whilst an estimated 17 (11 to 26) would have their cancer missed (test negative in error).

The specificity of RMI1 in postmenopausal women at a threshold of 250 was 79.7% (95% CI 71.6 to 85.9). For every 100 women without an actual diagnosis of ovarian cancer 80 (72 to 86) would correctly be identified as negative by the test whilst 20 (14 to 28) would receive a positive diagnosis of ovarian cancer in error (a false positive result).

The expected trade off between sensitivity and specificity with changes in threshold was observed; with an increase in test positivity threshold, sensitivity increased and specificity decreased. There was no evidence of a significant difference in sensitivity between the 2 test positivity thresholds reported in included RMI studies. Estimated specificity was lower (88.3% (95% CI 84.3 to 91.6) at a threshold of 250 compared to 95.4% (95% CI 92.5 to 97.3) at a threshold of 200 in premenopausal women which appeared to be of borderline significance. However, this finding should be interpreted with caution as it is based on a small number of studies.

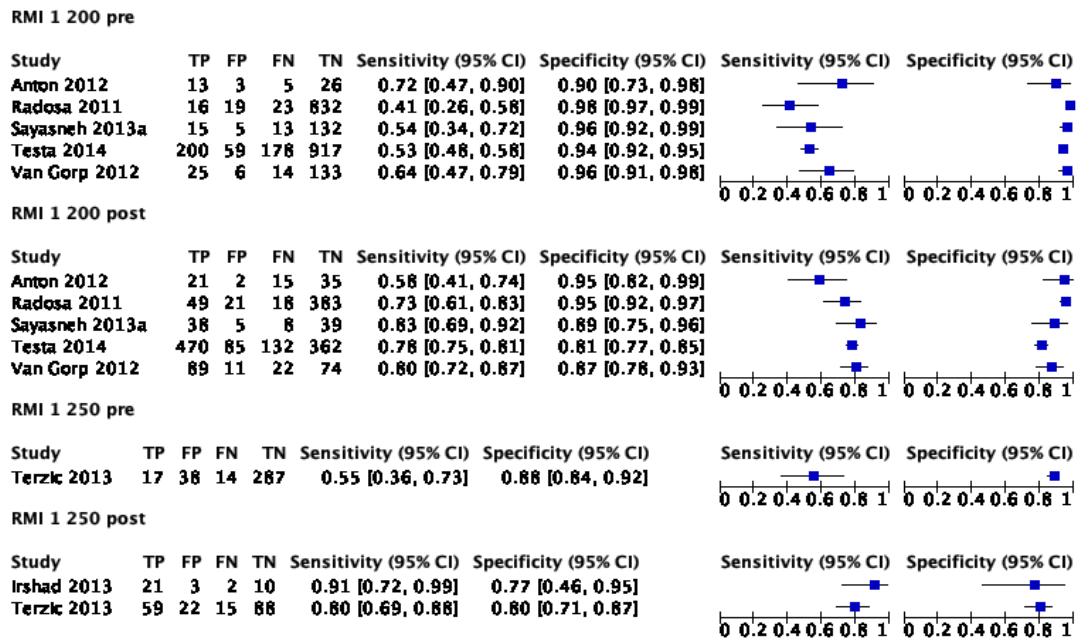


Figure 4. 3: Forest plot of tests: RMI1 at thresholds of 200 and 250, separately for premenopausal and postmenopausal women.

Accuracy of ROMA (See Figure 4.4)

The test accuracy of ROMA was investigated at three test positivity threshold pairs: 12.5 for premenopausal women and 14.4 for postmenopausal women; 13.1 +/- 2 for premenopausal women and 27.7 +/- 2 for postmenopausal women; 7.4 +/- 2 for premenopausal women and 25.3 +/- 2 for postmenopausal women.

ROMA premenopausal women (thresholds 7.4 and 13.1)

Sensitivity in premenopausal women varied from 82.6% (95 % CI 71.6 to 89.9) at a threshold of 7.4 (+/-2) to 78.3% (95% CI 71.2 to 84.0) at a threshold of 13.1 (+/-2). For every 100 women with an actual diagnosis of ovarian cancer an estimated 78 (71 to 84) to 83 (72 to 90) would have their cancer identified whilst an estimated 17 (10 to 28) to 22 (16 to 29 would have their cancer missed (test negative in error).

Specificity varied from 80.0% (95% CI 68.4 to 88.1) at a threshold of 7.4 (+/-2) to 83.3% (95% CI 78.2 to 87.5) at a threshold of 13.1 (+/-2). For every 100 women without an actual diagnosis of ovarian cancer, 83 (78 to 88) to 80 (68 to 88) would correctly be identified as negative by

the test whilst 17 (22 to 12) to 20 (32 to 12) would receive a positive diagnosis of ovarian cancer in error (a false positive result).

ROMA postmenopausal women (thresholds 14.4 and 27.7)

Sensitivity in postmenopausal women varied from 88.0% (95% CI 80.6 to 92.8) at a threshold of 14.4 to 90.7% (95% CI 86.4 to 93.7) at a threshold of 27.7 (+/-2). For every 100 women with an actual diagnosis of ovarian cancer an estimated 88 (81 to 93) to 91 (86 to 94) would have their cancer identified whilst an estimated 11 (6 to 14) to 12 (7 to 19) would have their cancer missed (test negative in error).

Specificity in postmenopausal women varied from 68.3% (95% CI 57.4 to 77.5) at a threshold of 14.4 to 79.1% (95% CI 74.6 to 83.1) at a threshold of 27.7 (+/-2). For every 100 women without an actual diagnosis of ovarian cancer, 68 (57 to 78) to 79 (75 to 83) would correctly be identified as negative by the test whilst 21 (17 to 25) to 32 (28 to 43) would receive a positive diagnosis of ovarian cancer in error (a false positive result).

The expected trade off between sensitivity and specificity with changes in threshold was observed; with an increase in test positivity threshold, sensitivity increased and specificity decreased. There was no evidence of a significant difference in accuracy between thresholds reported in included ROMA studies. It is not possible to recommend a specific ROMA threshold for use in clinical practice from the evidence reviewed.

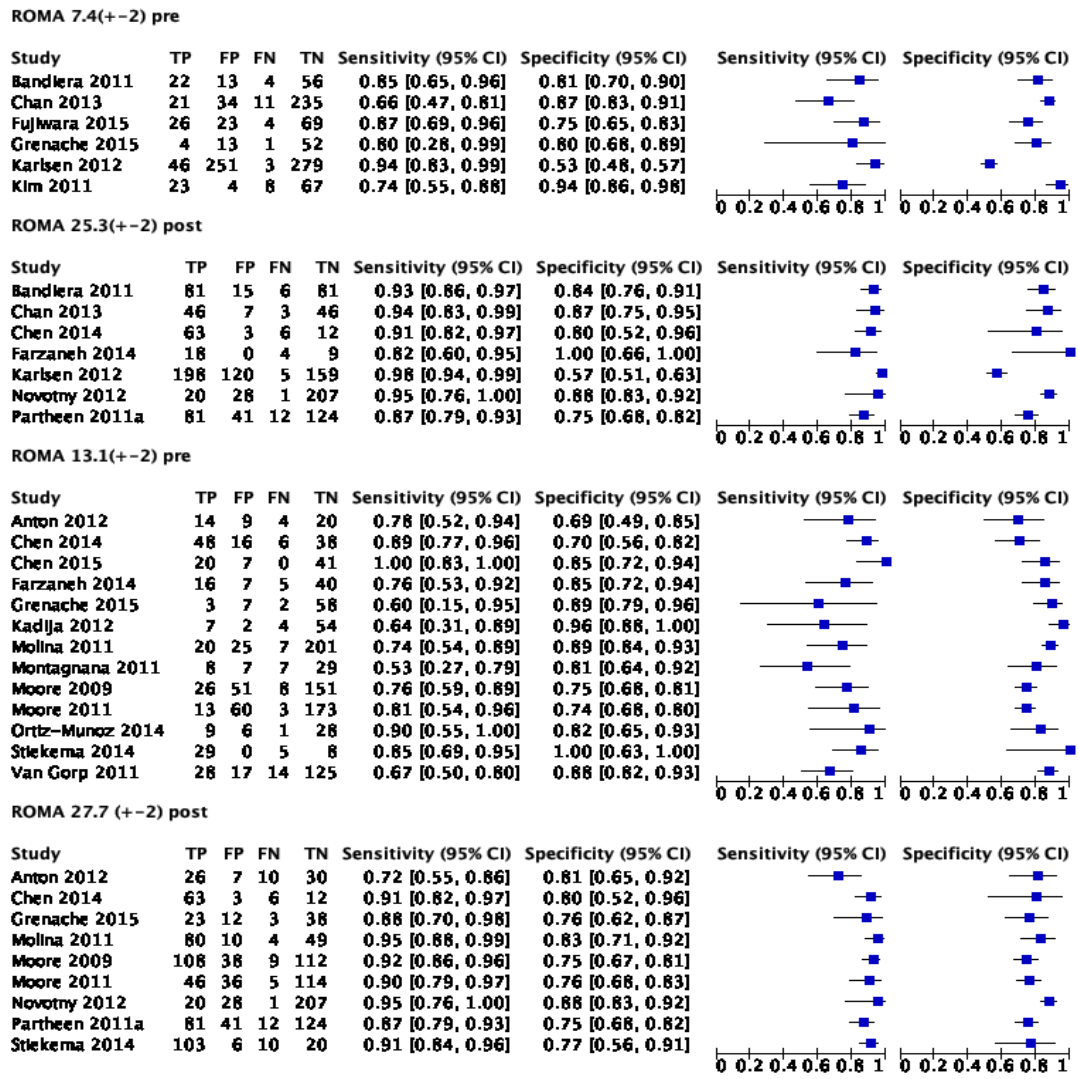


Figure 4. 4: Forest plot of tests: 1 ROMA in at (Thresholds of 7.4(+-2) and 13.1(+-2) in premenopausal women, and at thresholds of 25.3(+-2) and 27.7 (+-2) in postmenopausal women.

Accuracy in LR2 (See Figure 4.5)

All LR2 studies included a threshold to achieve a post test probability of malignancy of 10%. Only one study also included a threshold to achieve a post test probability of malignancy of 6.5% (Nunes 2013).

LR2 post test probability of malignancy 10% premenopausal women

LR2 post test probability of malignancy 10% in premenopausal women (3 studies, 2715 participants of whom 588 had ovarian cancer).

The sensitivity of LR2 in premenopausal women for a post test probability of malignancy of 10% was 83.6% (95% CI 78.6 to 87.6). For every 100 premenopausal women with an actual diagnosis of ovarian cancer an estimated 84 (79 to 88) would have their cancer identified whilst an estimated 16 (12 to 21) would have their cancer missed (test negative in error).

The specificity of LR2 in premenopausal women for a post test probability of malignancy was 90.4% (95% CI 81.3 to 95.3). For every 100 premenopausal women without an actual diagnosis of ovarian cancer, 90 (81 to 95) would correctly be identified as negative by the test whilst 10 (5 to 19) would receive a positive diagnosis of ovarian cancer in error (a false positive result).

LR2 post test probability of malignancy 10% postmenopausal women

LR2 post test probability of malignancy 10% in postmenopausal women (3 studies, 1861 participants of whom 1008 had ovarian cancer).

The sensitivity of LR2 in postmenopausal women for a post test probability of malignancy of 10% was 93.9% (95% CI 92.3 to 95.3). For every 100 postmenopausal women with an actual diagnosis of ovarian cancer an estimated 94 (92 to 95) would have their cancer identified whilst an estimated 6 (5 to 8) would have their cancer missed (test negative in error).

The specificity of LR2 in postmenopausal women for a post test probability of malignancy of 10% was 63.9% (95% CI 60.6 to 67.0). For every 100 postmenopausal women without an actual diagnosis of ovarian cancer, 64 (61 to 67) would correctly be identified as negative by the test whilst 36 (33 to 39) would receive a positive diagnosis of ovarian cancer in error (a false positive result).

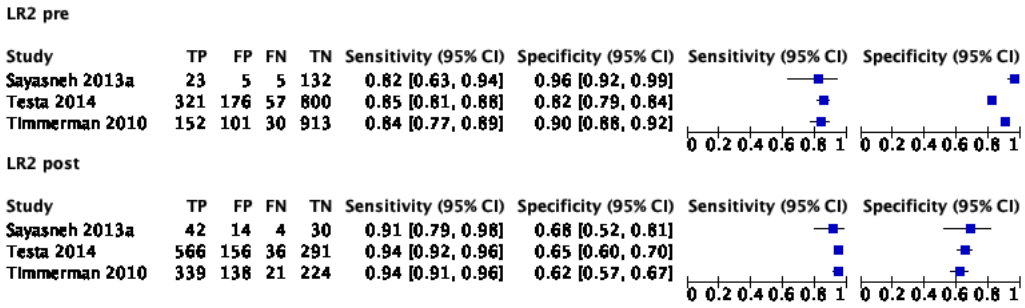


Figure 4. 5: Forest plot of tests: LR2 separately for premenopausal and postmenopausal women.

Accuracy in ACOG v3 (incorporating CA125). See Figure 4.6

ACOG vs (CA125) premenopausal women

ACOG vs (CA125) in premenopausal women (3 studies, 899 participants of whom 150 had ovarian cancer).

The sensitivity of ACOG v3 (CA125) in premenopausal women was 78.0% (95 CI 70.4 to 84.1). For every 100 premenopausal women with an actual diagnosis of ovarian cancer an estimated 78 (70 to 84) would have their cancer identified whilst an estimated 22 (16 to 30) would have their cancer missed (test negative in error).

The specificity of ACOG v3 (CA125) in premenopausal women was 83.9% (95% CI 65.1 to 92.6). For every 100 premenopausal women without an actual diagnosis of ovarian cancer, 84 (65 to 93) would correctly be identified as negative by the test whilst 16 (7 to 35) would receive a positive diagnosis of ovarian cancer in error (a false positive result).

ACOG vs 3 (CA125) in postmenopausal women

ACOG vs 3 (CA125) in postmenopausal women (3 studies, 1224 participants of whom 486 had ovarian cancer).

The sensitivity of ACOG v3 (CA125) in postmenopausal women was 84.6% (95% CI 76.0 to 90.5). For every 100 postmenopausal women with an actual diagnosis of ovarian cancer an

estimated 85 (76 to 91) would have their cancer identified whilst an estimated 15 (9 to 24) would have their cancer missed (test negative in error).

The specificity of ACOG v3 (CA125) in postmenopausal women was 78.1% (95% CI 71.0 to 83.9). For every 100 postmenopausal women without an actual diagnosis of ovarian cancer, 78 (71 to 84) would correctly be identified as negative by the test whilst 22 (16 to 29) would receive a positive diagnosis of ovarian cancer in error (a false positive result).

ACOG v3 (incorporating OVA1). See Figure 4.6

ACOG v3 (incorporating OVA1) premenopausal women

ACOG v3 (incorporating OVA1) in premenopausal women (one study, 235 participants of whom 45 had ovarian cancer).

The sensitivity of ACOG v3 (OVA1) in premenopausal women was 91.1% (95% CI 78.8 to 97.5). For every 100 premenopausal women with an actual diagnosis of ovarian cancer an estimated 91 (79 to 98) would have their cancer identified whilst an estimated 9 (2 to 21) would have their cancer missed (test negative in error).

The specificity of ACOG vs (OVA1) in premenopausal women was 43.2% (95% CI 36.0 to 50.5). For every 100 premenopausal women without an actual diagnosis of ovarian cancer, 43 (36 to 51) would correctly be identified as negative by the test whilst 57(49 to 64) would receive a positive diagnosis of ovarian cancer in error (a false positive result).

ACOG v3 (incorporating OVA1) postmenopausal women

ACOG v3 (incorporating OVA1) in postmenopausal women (one study, 281 participants of whom 116 had ovarian cancer.

The sensitivity of ACOG v3 (OVA1) in postmenopausal women was 94.8% (95% CI 89.1 to 98.1). For every 100 postmenopausal women with an actual diagnosis of ovarian cancer an estimated 95 (89 to 98) would have their cancer identified whilst an estimated 5 (2 to 11) would have their cancer missed (test negative in error).

The specificity of ACOG v3 (OVA1) in postmenopausal women was 25.5% (95 CI 19.0 to 32.8). For every 100 postmenopausal women without an actual diagnosis of ovarian cancer, 25 (19 to 33) would correctly be identified as negative by the test whilst 75 (67 to 81) would receive a positive diagnosis of ovarian cancer in error s(a false positive result).

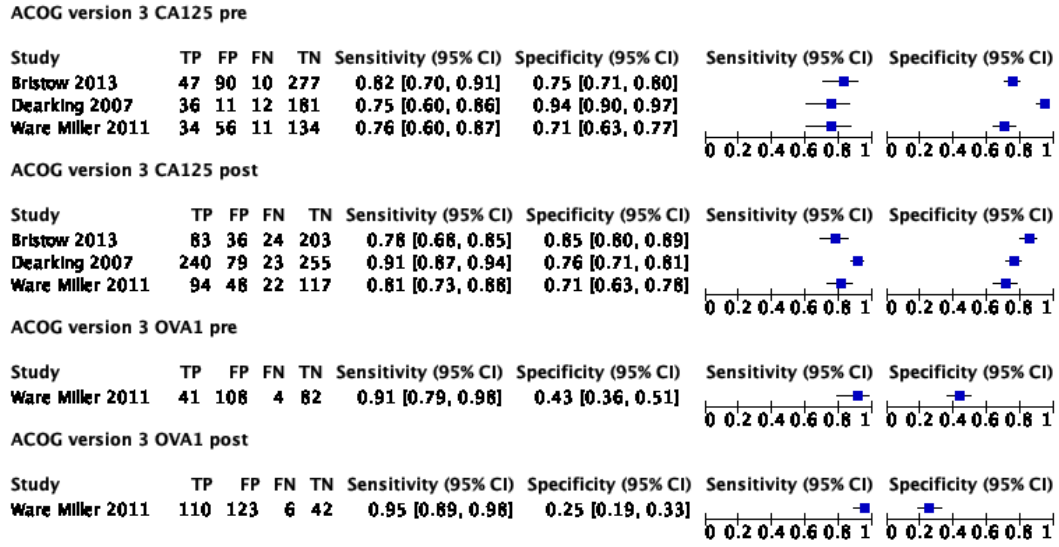


Figure 4. 6 Forest plot of tests: ACOG version 3 CA125 and ACOG version 3 OVA1 separately for premenopausal and postmenopausal women.

Accuracy in ADNEX See Figure 4.7

The test accuracy of ADNEX was investigated at four thresholds to achieve post test probabilities of malignancy of 3%, 5%, 10%, 15% (1 study). Results for post test probabilities of 5% and 10% are presented below.

ADNEX post test probability of malignancy 5% premenopausal women

ADNEX post test probability of malignancy 5% in premenopausal women (one study,1354 participants of whom 378 had ovarian cancer).

The sensitivity of ADNEX in premenopausal women at a threshold to achieve a post test probability of malignancy of 5% was 97.6% (95% CI 95.5 to 98.7). For every 100

premenopausal women with an actual diagnosis of ovarian cancer an estimated 98 (96 to 99) would have their cancer identified whilst an estimated 2 (1 to 4) would have their cancer missed (test negative in error).

The specificity of ADNEX in premenopausal women at a threshold to achieve a post test probability of malignancy of 5% was 69.5% (95% CI 66.5 to 72.3). For every 100 premenopausal women without an actual diagnosis of ovarian cancer, 70 (67 to 72) would correctly be identified as negative by the test whilst 30 (28 to 33) would receive a positive diagnosis of ovarian cancer in error (a false positive result).

ADNEX post test probability of malignancy 5% postmenopausal women

ADNEX post test probability of malignancy 5% in postmenopausal women (one study, 1049 participants of whom 602 had ovarian cancer).

The sensitivity of ADNEX in postmenopausal women at a threshold to achieve a post test probability of malignancy of 5% was 98.8% (95 CI 97.6 to 99.4). For every 100 postmenopausal women with an actual diagnosis of ovarian cancer an estimated 99 (98 to 99) would have their cancer identified whilst an estimated 1 (1 to 2) would have their cancer missed (test negative in error)

The specificity of ADNEX in postmenopausal women at a threshold to achieve a post test probability of malignancy of 5% was of 37.4% (95% CI 33.0 to 41.9). For every 100 postmenopausal women without an actual diagnosis of ovarian cancer, an estimated 38 (33 to 42) would correctly be identified as negative by the test whilst 62 (48 to 67) would receive a positive diagnosis of ovarian cancer in error (a false positive result).

ADNEX post test probability of malignancy 10% premenopausal women

ADNEX post test probability of malignancy 10% in premenopausal women (one study, 1354 participants of whom 378 had ovarian cancer).

The sensitivity of ADNEX in premenopausal women at a threshold to achieve a post test probability of malignancy of 10% was 94.7% (95% CI 92.0 to 96.6). For every 100 premenopausal women with an actual diagnosis of ovarian cancer an estimated 95 (92 to 97)

would have their cancer identified whilst an estimated 5 (3 to 8) would have their cancer missed (test negative in error).

The specificity of ADNEX in premenopausal women at a threshold to achieve a post test probability of malignancy of 10% was 78.6% (95% CI 75.9 to 81.1). For every 100 premenopausal women without an actual diagnosis of ovarian cancer, an estimated 79 (76 to 81) would correctly be identified as negative by the test whilst 21 (19 to 24) would receive a positive diagnosis of ovarian cancer in error (a false positive result).

ADNEX post test probability of malignancy 10% postmenopausal women

ADNEX post test probability of malignancy 10% in postmenopausal women (one study,1049 participants of whom 602 had ovarian cancer).

The sensitivity of ADNEX in postmenopausal women to achieve a post test probability of malignancy of 10% was 97.7% (95% CI 96.1 to 98.6). For every 100 postmenopausal women with an actual diagnosis of ovarian cancer an estimated 98 (96 to 99) would have their cancer identified whilst 2 (1 to 4) would have their cancer missed (test negative in error).

The specificity of ADNEX in postmenopausal women to achieve a post test probability of malignancy of 10% was 55.5% (95% CI 50.9 to 60.0). For every 100 postmenopausal women without an actual diagnosis of ovarian cancer, an estimated 56 (51 to 60) would correctly be identified as negative by the test whilst 46 (40 to 49) would receive a positive diagnosis of ovarian cancer in error (a false positive result).

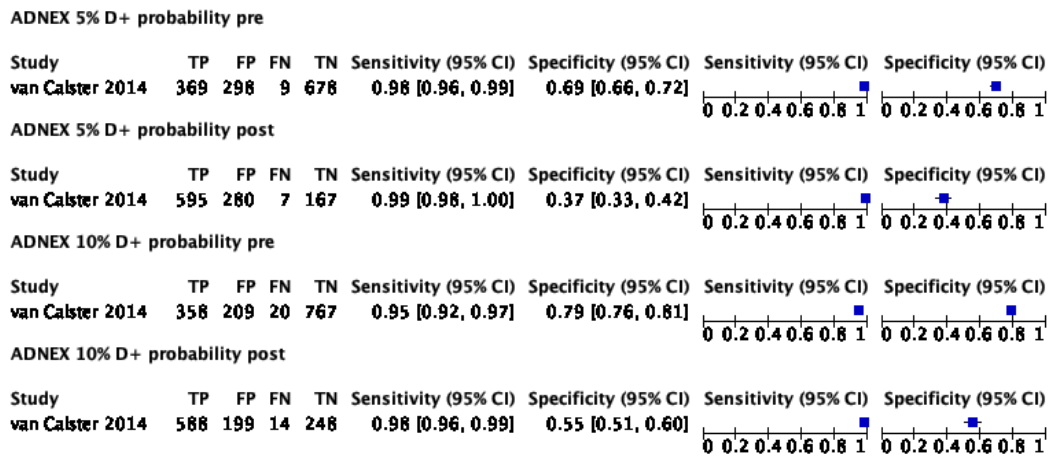


Figure 4. 7: Forest plot of tests: ADNEX at thresholds of 5% and 10% disease probability separately for premenopausal and postmenopausal women.

Comparison of index test accuracy: LR2, ROMA, ACOG v3 CA125 and ACOG v3 OVA1 compared to RMI1 at a fixed specificity of 90%

To maximise data for comparison, studies were included regardless of the test positivity threshold used and an indirect comparison of index tests was undertaken by fitting HSROC curves for premenopausal women and postmenopausal women separately. RMI1 was chosen as the baseline comparator as this is the test combination currently in routine clinical use in the UK. Differences in sensitivity between tests was estimated at fixed specificities of 80% and 90%. Table 4.7 and Figure 4.8 and Figure 4.9 illustrates the results of these comparisons for each version of each index test in premenopausal women and separately in postmenopausal women. 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. Here results at a fixed specificity of 90% for test versions considered most relevant to current practice are presented (ROMA, LR2, ACOG version 3 (CA125) and ACOG version 3 OVA1). Statistical comparison of the accuracy of ADNEX with RMI1 was not possible as only one ADNEX study was included in this review.

At a fixed specificity of 90%, for every 100 women without an actual diagnosis of ovarian cancer, 90 would correctly be identified as negative by the test whilst 10 would receive a

positive diagnosis of ovarian cancer in error (a false positive result). At this specificity threshold 90 of 100 women would be correctly reassured that they do not have ovarian cancer. A minority, 10 of 100 women, would be referred to secondary care unnecessarily, causing anxiety for the woman concerned as well as further inappropriate testing and risking unnecessary surgery with consequent risk to health and to preservation of fertility.

Comparison of sensitivity at a fixed specificity of 90%: LR2, ROMA, ACOG v3 Ca125 and ACOG v3 OVA1 compared to RMI1 in premenopausal women

At a fixed specificity of 90% RMI1 had a sensitivity of 0.59% (95% CI 0.54 to 0.65) in premenopausal women.

LR2 [0.82% (95% CI 0.79 to 0.86)] and ROMA [0.75% (95% CI 0.69 to 0.80)] were found to have a statistically significantly higher sensitivity compared to RMI1 in premenopausal women (relative DOR 7.33% [95% CI 2.34 to 22.94] p=0.0001 and 3.37% [9% CI 1.22 to 9.28] p=0.02 respectively). Although estimates of sensitivity appeared higher for both ACOG v3 (CA125) [0.73% (95% CI 0.64 to 0.81)] and ACOG v3 (OVA1) [0.82% (95% CI 0.57 to 0.94)] compared to RMI1 this was not statistically significant (relative DOR 2.95% (95% CI 0.91 to 9.61) p=0.07 and 6.73% (95% CI 0.49 to 91.75) p=0.15 respectively).

Based on this analysis, for every 100 premenopausal women with an actual diagnosis of ovarian cancer, 82 (79 to 86) women would test positive and be managed appropriately with LR2, 75 (69 to 80) with ROMA and 59 (54 to 65) with RMI1 whereas 18 (14 to 31) would test negative in error (false negative test result) and their cancer would be missed by LR2, 25 (20 to 31) with ROMA and 41 (35 to 46) with RMI.

Comparison of sensitivity at a fixed specificity of 90%: LR2, ROMA, ACOG v3 Ca125 and ACOG v3 OVA1 compared to RMI1 in postmenopausal women

There was no evidence for statistically significant differences in sensitivity for ROMA, LR2, ACOG v3 (CA125) or ACOG v3 (OVA1) at a fixed specificity of 90%.

Both pre and postmenopausal comparisons of accuracy should be interpreted with some caution as the analysis is limited by the volume of data available and the possible effects of confounding.

Table 4. 7: Comparison of the sensitivity of ROMA, LR2, ACOG V3 (CA125) and ACOG V3 (OVA1) to RMI I at fixed specificities of 80% and 90%.

PREMENOPAUSAL WOMEN								
TEST	**Studies** **N** **D+**	**Diagnostic Odds Ratio (95%CI)**	**ROR (95%CI)**	**P-value**	**Sensitivity at fixed specificity of 80% (95%CI)**	**Difference sensitivity at fixed specificity of 80% compared to RMI I (95% CI)**	**Sensitivity at fixed specificity of 90% (95% CI)**	**Difference sensitivity at fixed specificity of 90% compared to RMI I (95% CI)**
RMI1	6 2990 533	6.9 (4.0, 11.9)	-	-	0.66 (0.57, 0.74)	-	0.59 (0.54, 0.65)	-
ROMA	18 2564 445	23.3 (12.7, 42.5)	3.37 (1.22, 9.28)	0.02	0.80 (0.75, 0.84)	0.14 (0.05, 0.23)	0.75 (0.69, 0.80)	0.16 (0.07, 0.24)
LR2	3 2715 588	50.7 (25.0, 102.8)	7.33 (2.34, 22.94)	0.001	0.86 (0.83, 0.89)	0.20 (0.13, 0.28)	0.82 (0.79, 0.86)	0.23 (0.16, 0.30)
ACOG V3 CA125	3 899 150	20.4 (8.8, 47.2)	2.95 (0.91, 9.61)	0.07	0.78 (0.71, 0.85)	0.13 (0.02, 0.23)	0.73 (0.64, 0.81)	0.14 (0.03, 0.25)
ACOG V3 OVA1	1 235 45	46.5 (4.6, 469)	6.73 (0.49, 91.75)	0.15	0.86 (0.65, 0.95)	0.20 (0.01, 0.38)	0.82 (0.57, 0.94)	0.22 (0.02, 0.43)

Table 4.7 Continued...

TEST	Studies N D+	Diagnostic Odds Ratio (95%CI)	ROR (95% CI)	P-value	Sensitivity at fixed specificity of 80% (95%CI)	Difference sensitivity at fixed specificity of 80% compared to RMI I (95% CI)	Sensitivity at fixed specificity of 90% (95% CI)	Difference sensitivity at fixed specificity of 90% compared to RMI I (95% CI)
RMI1	7 2099 959	21.7 (14.7, 32.1)	-	-	0.84 (0.78, 0.89)	-	0.72 (0.64, 0.79)	-
ROMA	16 2594 1148	38.8 (23.7, 36.5)	1.79 (0.90, 3.55)	0.09	0.90 (0.87, 0.93)	0.06 (-0.01, 0.13)	0.82 (0.73, 0.89)	0.09 (0.00, 0.18)
LR2	3 1861 1008	40.4 (19.1, 85.2)	1.86 (0.82, 4.23)	0.13	0.90 (0.85, 0.94)	0.06 (-0.01, 0.13)	0.82 (0.69, 0.91)	0.10 (0.01, 0.19)
ACOG V3 (CA125)	3 1224 486	20.0 (11.8, 33.9)	0.92 (0.46, 1.85)	0.81	0.83 (0.76, 0.89)	-0.01 (-0.10, 0.08	0.71 (0.56, 0.82)	-0.02 (-0.15, 0.12)
ACOG V3 (OVA1)	1 281 116	14.5 (2.6, 58.9)	0.58 (0.11, 3.05)	0.5	0.76 (0.43, 0.93)	-0.08 (-0.37, 0.21)	0.61 (0.22, 0.90)	-0.11 (-0.50, 0.28)

Notes to table: Results are from fitting HSROC models to all available data taking one threshold from each stud. Where one study reported more than one threshold, the most commonly reported threshold across studies was chosen.

OC: Ovarian Cancer cases. ROR: Relative Diagnostic Odds Ratio.

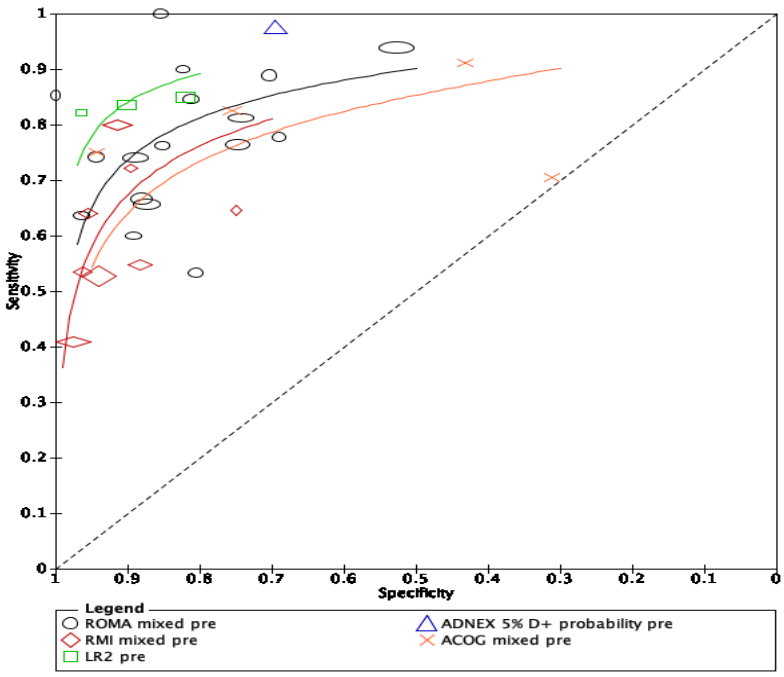


Figure 4. 8:Summary ROC Plot of tests: Premenopausal women: ROMA, RMI1, LR2, ACOG v3CA125 and ACOG v3 OVA1, ADNEX 5% Disease probability.

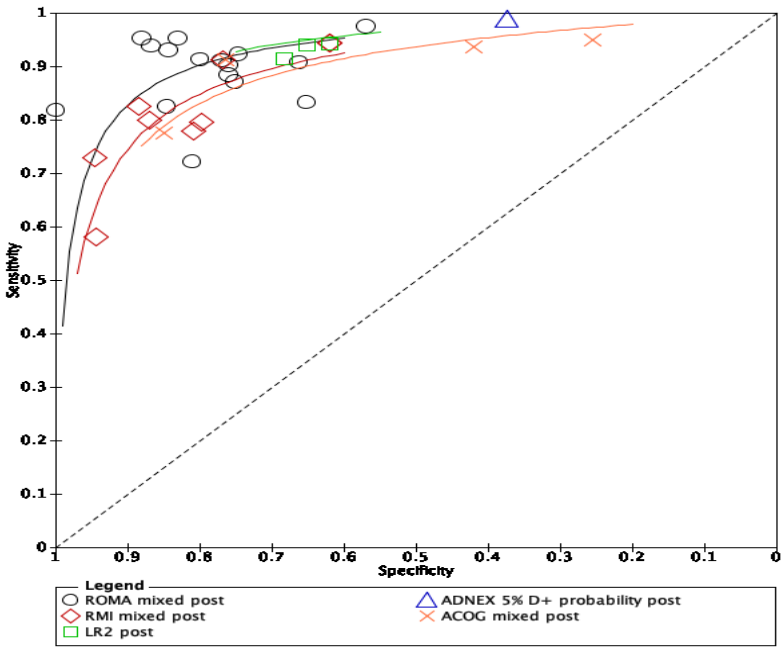


Figure 4. 9: Summary ROC Plot of tests: Postmenopausal women: ROMA, RMI1, LR2, ACOG v3 CA125, ACOG v3 OVA1, ADNEX 5% Disease probability.

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

In current clinical practice, ovarian tumours that are histologically borderline for malignancy are managed as malignant tumours. However, for estimation of test accuracy, disease state needs to be classified as either positive or negative. It was observed that included studies varied in the way borderline tumours were classified: as positive (grouped with malignant tumours) or negative (grouped with benign tumours) for the purposes of accuracy estimation. A comparison of test accuracy estimated by studies excluding borderline tumours from analysis or where the classification of borderline tumours for analysis was unclear, with studies where borderline tumours were classified as positive (grouped with malignant tumours) was performed.

Due to inadequate data this investigation was only possible for ROMA. In premenopausal women 10 ROMA studies either excluded borderline tumours for the purposes of accuracy estimation or the classification of borderline tumours was unclear and in 8 studies borderline tumours were classified as positive (grouped with malignant tumours). In postmenopausal women 9 ROMA studies either excluded borderline tumours for the purposes of accuracy estimation or the classification of borderline tumours was unclear and in 7 studies borderline tumours were classified as positive (grouped with malignant tumours). No evidence was found to suggest that the classification of borderline tumours influenced summary estimates of test accuracy in this review. However, the analysis is limited by the volume of data available and, because this is an indirect comparison, the possible effects of confounding

Table 4. 8: Summary of findings for test combinations in pre and postmenopausal women in secondary care (prevalence ovarian cancer 3% pre and 10% post).

Review question	Symptoms, ultrasound imaging and biochemical markers alone or 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	Approximately half of included studies were rated as at high risk of bias for the patient selection domain on the basis that sampling was not consecutive and the range of ovarian pathology included was limited. The majority of RMI studies were judged at unclear risk of bias in the index test domain on the basis it was unclear if the index test (including US) was interpreted blind to the reference standard. The majority of studies were judged as unclear applicability concern in the reference standard domain because length of followup for index test -ves (who did not undergo surgery) was not reported. The majority of included studies were judged as at high or unclear applicability in the patient selection domain on the basis that women were either asymptomatic or this was not clear and details of prior tests were not provided.					
Population	Premenopausal women - assume prevalence of ovarian cancer 3%					
Number of studies, premenopausal women, cases of ovarian cancer	A total of 31 studies were included in the comparative analysis. Prevalence in included studies varied between 9% to 28%, (median 19%) in premenopausal women.					
Index Test	Sensitivity (95% CI) at single threshold	Specificity at single threshold (95% CI)	Index test (multiple thresholds)	Sensitivity (95% CI) estimated across multiple thresholds at a fixed specificity of 0.9	Consequences in a hypothetical cohort of 1000 women assuming a specificity of 0.90 and a prevalence of 3%. Number of women who would have their cancer missed (false negatives).	Difference in sensitivity compared to RMI I at a fixed specificity Difference to RMI1 (p value)
RMI1 200	52.2 (45.9, 58.5)	95.4 (92.5, 97.3)	RMI 200 and 250	0.59 (0.54, 0.65)	12 (11 to 14)	----------
LR2	83.6 (78.6, 87.6)	90.4 (81.3, 95.3)	LR2	0.82 (0.79, 0.86)	5 (4 to 6)	p = 0.001
ROMA (13.1(+/-2)	78.3 (71.2, 84.0)	83.3 (78.2, 87.5)	ROMA 7.4 (+/-2);12.5; 13.1 (+/-2)	0.75 (0.69, 0.80)	8 (6 to 9)	p = 0.02
ACOG v3 (CA125)	78.0 (70.4, 84.1)	83.9 (65.1, 92.6)	ACOG v3 (CA125)	0.73 (0.64, 0.81)	8 (6-11)	p = 0.07
ACOG v3 (OVA1)	91.1 (78.8, 97.5)	43.2 (36.0, 50.5)	ACOG v3 (OVA1)	0.82 (0.57, 0.94	5 (2 to 13)	p = 0.15
Population	Postmenopausal women- assume prevalence of ovarian cancer 10%					

Table 4.8 Continued...

Number of studies, postmenopausal women, cases of ovarian cancer	A total of 31 studies were included in the comparative analysis. Prevalence in included studies varied between 40% to 57%, (median 44%) in postmenopausal women.					
Index Test	Sensitivity (95% CI) at single threshold	Specificity (95% CI) at single threshold	Index test	Sensitivity (95% CI) estimated across multiple thresholds at a fixed specificity of 0.9	Consequences in a hypothetical cohort of 1000 women assuming a specificity of 0.90 and a prevalence of 10%. Number of women who would have their cancer missed (false negatives).	Difference in sensitivity compared to RMI1 at a fixed specificity of 0.9 (p value)
LR2	93.9 (92.3, 95.3)	63.9 (60.6, 67.0)	LR2	0.82 (0.69, 0.91)	18 (9 to 31)	p = 0.13
ROMA 27.7 (+/-2)	90.7 (86.4, 93.7)	79.1(74.6, 83.1)	ROMA 25.3 (+/-2); 14.4; 27.7 (+/-2)	0.82 (0.73, 0.89)	18 (11 to 27)	p = 0.09
ACOG v3 (CA125)	84.6 (76.0, 90.5)	78.1 (71.0, 83.9)	ACOG v3 (CA125)	0.71 (0.56, 0.82)	29 (18 to 44)	p = 0.81
ACOG v3 (OVA1)	94.8 (89.1, 98.1)	25.5 (19.0, 32.8)	ACOG v3 (OVA1)	0.61 (0.22, 0.90)	39 (10 to 78	p = 0.76
RMI1 200	75.0 (69.5, 79.8)	90.1(83.1, 94.4)	RMI 200 and 250	0.72 (0.64, 0.79)	28 (21 to 36)	------

Table 4. 9: Summary of findings for test combinations for pre and postmenopausal women in tertiary care (prevalence ovarian cancer 30% pre and postmenopausal).

Review question	Symptoms, ultrasound imaging and biochemical markers alone or 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	Approximately half of included studies were rated as at high risk of bias for the patient selection domain on the basis that sampling was not consecutive and the range of ovarian pathology included was limited. The majority of RMI studies were judged at unclear risk of bias in the index test domain on the basis it was unclear if the index test (including US) was interpreted blind to the reference standard. The majority of studies were judged as unclear applicability concern in the reference standard domain because length of follow up for index test -ves (who did not undergo surgery) was not reported. The majority of included studies were judged as at high or unclear applicability in the patient selection domain on the basis that women were either asymptomatic or this was not clear and details of prior tests were not provided.					
Population	Premenopausal women - assume prevalence of ovarian cancer 30%					
Number of studies, premenopausal women, cases of ovarian cancer	A total of 31 studies were included in the comparative analysis. Prevalence in included studies varied between 9% to 28%, (median 19%) for premenopausal women.					
Index Test	Sensitivity (95% CI) at single threshold	Specificity (95% CI) at single threshold	Index test	Sensitivity (95% CI) estimated across multiple thresholds at a fixed specificity of 0.9	Consequences in a hypothetical cohort of 1000 women assuming a specificity of 0.90 and a prevalence of 10%	Difference in sensitivity compared to RMI I at a fixed specificity Difference to RMI1 (p value)
LR2	83.6 (78.6, 87.6)	90.4 (81.3, 95.3)	LR2	0.82 (0.79, 0.86)	54 (42 to 63)	p = 0.001
ROMA	78.3 (71.2, 84.0)	83.3 (78.2, 87.5)	ROMA 7.4 (+/-2); 12.5; 13.1 (+/-2)	0.75 (0.69, 0.80)	123 (105 to 138)	p = 0.02
ACOG v3 (CA125)	78.0 (70.4, 84.1)	83.9 (65.1, 92.6)	ACOG v3 (CA125)	0.73 (0.64, 0.81)	81 (57 to 108)	p = 0.07
ACOG v3 (OVA1)	91.1 (78.8, 97.5)	43.2 (36.0, 50.5)	ACOG v3 (OVA1)	0.82 (0.57, 0.94	54 (18, 129)	p = 0.15

Table 4.9 Continued...

RMI1	52.2 (45.9, 58.5)	95.4 (92.5, 97.3)	RMI 200 and 250	0.59 (0.54, 0.65)	123 (105 to 138)	----------
Population	Postmenopausal women - assume prevalence of ovarian cancer 30%					
Number of studies, postmenopausal women, cases of ovarian cancer	A total of 31 studies were included in the comparative analysis. Prevalence in included studies varied between 40% to 57%, (median 44%) in postmenopausal women.					
Index Test	Sensitivity (95% CI) at single threshold	Specificity (95% CI) at single threshold	Index test	Sensitivity (95% CI) estimated across multiple thresholds at a fixed specificity of 0.9	Consequences in a hypothetical cohort of 1000 women assuming a specificity of 0.90 and a prevalence of 10%	Difference in sensitivity compared to RMI I at a fixed specificity Difference to RMI1 (p value)
LR2	93.9 (92.3, 95.3)	63.9 (60.6, 67.0)	LR2	0.82 (0.69, 0.91)	54 (27 to 93)	p = 0.13
ROMA	90.7 (86.4, 93.7)	79.1(74.6, 83.1)	ROMA 25.3 (+/-2); 14.4; 27.7 (+/-2)	0.82 (0.73, 0.89)	54 (33 to 81)	p = 0.09
ACOG v3 (CA125)	84.6 (76.0, 90.5)	78.1 (71.0, 83.9)	ACOG v3 (CA125)	0.71 (0.56, 0.82)	87 (54 to 132)	p = 0.81
ACOG v3 s(OVA1)	94.8 (89.1, 98.1)	25.5 (19.0, 32.8)	ACOG v3 (OVA1)	0.61 (0.22, 0.90)	117 (30 to 234)	p = 0.76
RMI1	75.0 (69.5, 79.8)	90.1(83.1, 94.4)	RMI 200 and 250	0.72 (0.64, 0.79)	84 (63 to 108)	------

CHAPTER 5: RESULTS OF BIOMARKERS REVIEW

5.1 Screening and selection of studies

The original primary searches retrieved a total of 360005 records after de-duplication. After reviewing titles and abstracts, 34968 records were subsequently excluded. Full-text copies of 1037 potentially relevant reports were obtained and screened for inclusion, of which 45 were deemed eligible for inclusion in the biomarker review. Figure 4.1 shows the PRISMA diagram of the study selection process. Studies were excluded if they failed to meet one or more of the specified inclusion criteria with regard to study design, participants, target conditions or ability to derive a 2x2 table. (See Appendix 6 for reasons of exclusion).

The biomarkers included in the review were CA125, HE4 and OVA1. Ninety, nine studies met the original inclusion criteria. The spectrum of ovarian tumour varies with age and it is known that majority of EOC occurs in postmenopausal women. Apart from menopause being a risk factor, this variation in spectrum with different preponderance of histotypes with age affects the estimates of test accuracy due to the potential differential sensitivity of the test to different histotypes. Following our observation of consistent and mostly significant differences in test accuracy, only studies presenting results separately according to menopausal status were included in the synthesis post hoc (n=43). Of the included studies, 30 studies evaluated CA125, 13 studies evaluated HE4 and two studies evaluated OVA1. All the 13 studies estimating the test performance of HE4 also estimated the test performance of CA125 at various thresholds except one study.[138]

5.2 Study characteristics and quality of included studies

5.2.1 Study characteristics of included studies

In terms of pathways, common to all the biomarkers tested: the clinical pathway of patients from presentation to the decision to refer for surgical investigation were not detailed in any included studies. Testing prior to surgical investigation in this patient group will have included clinical history and examination, and may have also included ultrasound imaging or biomarker tests prior to the index test. These tests can be carried out in primary care, by gynaecologists, by gynaecological oncologists or a mixture of both. In clinical practice in the UK, clinical assessment and biomarker testing are most commonly initially performed in primary care. Ultrasound assessment may be initiated by a primary care practitioner but conducted in a secondary care setting. If the primary care practitioner decides to refer a woman for surgical investigation because of a suspicion of ovarian cancer, clinical assessment and ultrasound assessment may be repeated in secondary or tertiary care.

The characteristics of all included studies including the baseline characteristics of participants and the index test are described below under the headings CA125, HE4 and OVA1.

CA125 included studies

The study characteristics for 30 studies reporting the accuracy of CA125 separately in pre and/or postmenopausal women are described in (Table 5.1).

Results for both pre and postmenopausal women were described in 20 studies; only 2/22 studies[214], [137] describing results for premenopausal women did not include results for postmenopausal women. Eight studies included results only for postmenopausal women.[215, 216,217, 218, 219, 196,220,221]

Table 5. 1: Table of study characteristics: CA125 pre and postmenopausal studies.

Author year	Menopausal status	Study criteria and setting	Study design	Participant characteristics	Index test
Bandiera 2011	**pre, post**	**Study criteria:** Not reported **Clinical setting:** Tertiary **Prior test:** Unclear **Exclusions:** BOT excluded. ? Non EOC excluded	**Design:** Unclear **Country:** USA **Centre:** Single centre	**# participants:** 278 **% postmen (n):** 65.8 (183) **# ovarian cancer:** 113 **# benign:** 165 # **borderline:** excluded **Age mean+/-SD:** Pre endo-36.5, benign-41.5, malignant-44.7. post ben- 64, malignant-66.3 **Age median[range]:** Not reported[25-89]	**Prespecified threshold:** Yes-HE4, NO -CA125 **Time interval:** Unclear **Blinded to reference:** Unclear **Blinded to other index:** Not applicable
Benjapipal 2007	**pre, post**	**Study criteria:** Women with ovarian mass scheduled for elective surgery **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** Nil	**Design:** Prospective Cross sectional Study **Country:** Thailand Centre: Single centre	# **participants:** 120 % **postmen (n):** 41 (49) # **ovarian cancer:** 59 # **benign:** 61 # **borderline:** Not reported Age mean+/-SD: 46+/-14 **Age median[range]:** Not reported[18-91]	**Prespecified threshold:** Yes **Time interval:** 1 week **Blinded to reference:** Yes **Blinded to other index:** Not applicable
Botsis 1997	**post**	**Study criteria:** Women with suspected pelvic masses **Clinical setting:** Secondary **Prior test:** Unclear Exclusions: Nil	**Design:** Prospective Cross sectional Study **Country:** Greece Centre: Single centre	# **participants in study:** 62 % **postmen (n):** 100 (62) # **ovarian cancer**: 22 # **benign:** 40 # **borderline:** Not reported **Age mean+/-SD:** 54.4 [Not reported] **Age median[range]:** Not reported [47- 78]	**Prespecified threshold:** Yes **Time interval**: median 3 days (range 0-31) **Blinded to reference**: Unclear **Blinded to other index:** Unclear

Table 5.1 Continued...

Chan 2013	**pre, post**	**Study criteria:** Women over 18 years diagnosed with adnexal mass diagnosed by any Imaging method (US, CT or MRI) **Clinical setting:** Unclear **Prior test:** Unclear **Exclusions:** Nil	**Design:** Prospective Cross sectional Study **Country:** Acia- pacific **Centre:** Multicentre	# **participants:** 414 % **postmen (n):** 26%(108 # **ovarian cancer:** 74 # **benign:** 322 # **borderline:** 16 Age mean+/-SD: Not reported **Age median[range]:** Not reported[Not reported]	**Prespecified threshold**: yes **Time interval:** Unclear **Blinded to reference:** Unclear **Blinded to other index:** Unclear
Chen 2015	**pre, post**	**Study criteria:** Women with pelvic masses scheduled for surgery **Clinical setting**: Unclear **Prior test:** Unclear **Exclusions:** Unclear	**Design:** Unclear **Country:** China **Centre:** Single centre	# **participants:** 232 % **postmen (n):** Not reported # **ovarian cancer:** 60 # **benign:** 70 # **borderline:** Not reported **Age mean+/-SD:** benign-33, Malignant-53 **Age median[range]:** Not reported[17-81]	**Prespecified threshold:** Yes **Time interval:** Unclear **Blinded to reference:** Unclear **Blinded to other index:** Unclear
Dowd 1993	**pre, post**	**Study criteria:** Women with CA125 and suspected pelvic mass **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** Nil	**Design:** Retrospective Cross sectional Study **Country:** Australia **Centre:** Single centre	# **participants:** 264 % **postmen (n):** 56(143) # **ovarian cancer:** 100 # **benign:** 128 # **borderline:** 36 **Age mean+/-SD**: Benign- 47, Malignant- 54 **Age median[range]:** Not reported[15-89]	**Prespecified threshold:** yes **Time interval:** Unclear **Blinded to reference:** Unclear **Blinded to other index:** Unclear

Table 5.1 Continued...

Franchi 1995	**pre, post**	**Study criteria:** women with suspected pelvic mass on examination/ US **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** Nil	**Design:** Unclear **Country:** Italy **Centre:** Single centre	# **participants:** 129 % **postmen (n):** 36 (46) # **ovarian cancer:** 30 # **benign:** 92 # **borderline:** 7 **Age mean+/-SD:** 44 **Age median[range]:** Not reported[12-91]	**Prespecified threshold:** Yes **Time interval:** less than 90 days **Blinded to reference:** Unclear **Blinded to other index:** Not applicable
Fujiwara 2015	**pre, post**	**Study criteria:** Women with adnexal mass diagnosed on Ultrasound/CT/MRI/PET and scheduled for surgery **Clinical setting:** Unclear **Prior test:** Unclear **Exclusions:** Women with previous history of any gynaecological disease and NON EOC excluded	**Design:** Prospective Cross sectional Study **Country:** Japan **Centre:** Multicentre	# **participants:** 221 % **postmen (n):** 40 (90) # **ovarian cancer:** 71 # **benign: 131** # **borderline: 19** Age mean+/-SD: Benign-43, BOT-47, Malignant-54 **Age median[range]:** Not reported[20-79]	**Prespecified threshold:** Yes **Time interval:** less than 42 days **Blinded to reference:** Unclear **Blinded to other index:** Not applicable
Gadducci 1992	**pre, post**	**Study criteria:** Women undergoing laparotomy because of ovarian masses **Clinical setting:** Secondary Prior **test:** Unclear Exclusions: Non EOC excluded	**Design:** Unclear **Country:** Italy **Centre:** Single centre	# **participants in study:** 344 % **postmen (n):** 38 (131) # **ovarian cancer:** 90 # **benign:** 254 # **borderline:** Not reported Age mean+/-SD: Not reported [Not reported] **Age median[range]:** Not reported [Not reported]	**Prespecified threshold:** Yes **Time interval:** Not reported **Blinded to reference:** Unclear **Blinded to other index:** Not applicable

Table 5.1 Continued...

Gadducci 1996	**pre, post**	**Study criteria:** Women with clinical diagnosis of ovarian mass submitted to laparotomy **Clinical setting:** Secondary Prior **test:** Unclear Exclusions: Unclear	**Design:** Unclear **Country:** Italy **Centre:** Single centre	# **participants in study:** 124 % **postmen (n):** 52 (64) # **ovarian cancer:** 56 # **benign:** 65 **# borderline:** Not reported Age mean+/-SD: Not reported **Age median[range]:** malignant: 62 [28-81]; benign: 42 [17-73]	**Prespecified threshold:** Yes **Time interval:** Not reported **Blinded to reference:** Unclear **Blinded to other index:** Unclear
Kuesel 1992	**pre, post**	**Study criteria:** Women with pelvic mass who has undergone laparotomy **Clinical setting:** Secondary Prior **test:** Unclear Exclusions: Unclear	**Design:** Retrospective cross sectional **study Country: Canada Centre:** Single centre	# **participants in study:** 37 % **postmen (n):** 57 (21) # **ovarian cancer:** 13 # **benign:** 24 # **borderline:** Not reported **Age mean+/-SD:** Not reported **Age median[range]:** Not reported [Not reported]	**Prespecified threshold:** No Time **interval**: Not reported **Blinded to reference:** Unclear **Blinded to other index:** Not applicable
Kusnetzoff 1998	**post**	**Study criteria:** patients with clinical or sonographic diagnosis of adnexal masses Clinical setting: Secondary Prior test: Unclear **Exclusions:** Nil	**Design:** Prospective cross sectional study **Country: Argentina Centre:** Single centre	**# participants in study:** 130 % **postmen (n):** 44 (57) # **ovarian cancer:** 39 # **benign:** 91 # **borderline:** 2 (with malignant) **Age mean+/-SD:** All 41.5, pre:34, post:58.3 **Age median[range]:** Not reported [19-82]	**Prespecified threshold:** No **Time interval: 1** week Blinded to **reference:** Unclear Blinded to other index: Unclear

Table 5.1 Continued...

Molina 2011	**pre, post**	**Study criteria:** Not reported **Clinical setting:** Unclear **Prior test:** Unclear **Exclusions:** Unclear	**Design:** Retrospective Cross sectional Study **Country:** Spain **Centre:** Single centre	# **participants:** 422 **% postmen (n):** 34(143) # **ovarian cancer:** 111 **# benign:** 285 *benign gyn disease with 137 ovarian cysts # **borderline:** Not reported **Age mean+/-SD:** Not reported **Age median[range]:** Benign gyn disease-40+/- 0.8, Gyn cancer61+/-1.2[17-90]	**Prespecified threshold:** Yes **Time interval:** Unclear **Blinded to reference:** Unclear **Blinded to other index:** Unclear
Partheen 2011	**pre, post**	**Study criteria:** Women with complex cystic mass and suspicious of malignancy undergoing surgery Clinical setting: Tertiary **Prior test:** Unclear **Exclusions:** Solid anpelvic mass were not eligible	**Design:** Prospective Cross sectional Study **Country:** Sweden **Centre:** Single centre	# **participants:** 374 % **postmen (n):** 73.7 (276) # **ovarian cancer:** 108 **# benign:** 215 # **borderline:** 45 Age mean+/-SD: Not reported Age median[range]: Not reported[Not reported]	**Prespecified threshold:** Yes **Time interval:** 1 day **Blinded to reference:** Yes **Blinded to other index:** Unclear
Radosa 2011	**pre, post**	**Study criteria:** women with adnexal mass who subsequently underwent surgery were selected **Clinical setting:** Tertiary **Prior test:** Unclear **Exclusions:** Nil	**Design:** Unclear **Country:** Germany Centre: Single centre	# **participants:** 1362 % postmen (n): 32(442) # **ovarian cancer:** 79 # **benign:** 1259 **# borderline:** 19 **Age mean+/-SD:** 43.3 **Age median[range]:** Not reported[Not reported]	**Prespecified threshold:** Yes **Time interval:** Unclear **Blinded to reference:** Unclear **Blinded to other index:** Unclear

Table 5.1 Continued...

Stiekema 2014	**pre, post**	**Study criteria:** histologically confirmed EOC or benign ovarian disease referred to the institute **Clinical setting:** Tertiary **Prior test:** Unclear **Exclusions:** BOT excluded	**Design:** Retrospective Cross sectional Study **Country:** Netherlands **Centre:** Single centre	# **participants:** 181 % **postmen (n):** &9 (143) # **ovarian cancer:** 147 # **benign:** 34 # **borderline:** Excluded **Age mean+/-SD:** Benign- 47, malignant- 57 **Age median[range]:** Not reported[Not reported]	**Prespecified threshold:** Yes **Time interval:** Unclear **Blinded to reference:** Unclear **Blinded to other index:** Unclear
Strigini 1995	**pre, post**	**Study criteria:** Women undergoing laparotomy for adnexal mass **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** Nil	**Design:** Unclear **Country:** Italy **Centre:** Single centre	# **participants:** 109 % **postmen (n):** 31.2 (34) **# ovarian cancer:** 19 # **benign:** 75 # **borderline:** Not reported **Age mean+/-SD:** Not reported **Age median[range]:** 43[18-80]	**Prespecified threshold:** yes **Time interval:** Unclear **Blinded to reference:** Unclear **Blinded to other index:** Unclear
Van Calster 2007	**pre, post**	**Study criteria:** Women with persistent adnexal mass undergoing surgery within 120 days **Clinical setting:** Mixed **Prior test:** Unclear **Exclusions:** Nil	**Design:** Retrospective Cross sectional Study **Country**: Europe **Centre**: Multicentre	# **participants:** 809 % **postmen (n):** 40.4 (431), # **ovarian cancer**: 152 # **benign**: 567 # **borderline**: 52 **Age mean+/-SD:** Not reported **Age median[range]:** Not reported[Not reported]	**Prespecified threshold:** No **Time interval**: 120 days **Blinded to reference:** Unclear **Blinded to other index:** Unclear

Table 5.1 Continued...

Van Gorp 2011	**pre, post**	**Study criteria**: All patients diagnosed with pelvic mass undergoing surgery **Clinical setting**: Unclear **Prior test**: Unclear **Exclusions**: Nil	**Design**: Prospective Cross sectional Study **Country**: Belgium **Centre**: Single centre	# **participants**: 389 % postmen (n): 41.4%(161) # **ovarian cancer**: 161 # **benign**: 228 # **borderline**: Not reported **Age mean+/-SD: Benign**:46.3+/-16 **Malignant:**57.8+/-12.6 Age median[range]: Not reported[Not reported]	**Prespecified threshold**: Yes **Time interval:** 1 day **Blinded to reference:** Unclear **Blinded to other index:** Yes
Holcomb 2011	**pre**	**Study criteria**: Women with adnexal mass undergoing surgery **Clinical setting**: Unclear **Prior test**: Unclear **Exclusions**: postmenopausal women excluded ?non EOC excluded)	**Design**: Prospective Cross sectional Study **Country**: USA **Centre**: Multicentre	# **participants**: 229 % **postmen (n):** 53.6 (265) **# ovarian cancer:** 18 # **benign:** 195 # **borderline:** 16 **Age mean+/-SD:** Not reported **Age median[range]:** Not reported[Not reported]	**Prespecified threshold:** Yes* (but reported for diff cut off than the prespecified) **Time interval:** Less than 30 days **Blinded to reference:** Yes **Blinded to other index:** Not applicable
Anton 2012	**post**	**Study criteria:** Women referred with pelvic masses diagnosed by US or CT or MRI undergoing surgery or image-guided biopsy when they presented with signs of carcinomatosis **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** Nil	**Design:** Prospective Cross sectional Study **Country:** Brazil **Centre:** Single centre	# **participants:** 120 % **postmen (n):** 60.8 (73) **# ovarian cancer:** 30 # **benign:** 66 # **borderline:** 17 **Age mean+/-SD:** Benign- 50.7, BOT-56.4, malignant-54.7 **Age median[range]:** Benign- 51, BOT-58, malignant-54[Not reported]	**Prespecified threshold:** Yes **Time interval:** 1 day **Blinded to reference:** Unclear **Blinded to other index:** Unclear

Table 5.1 Continued...

Antonic and Rakar 1995	**post**	**Study criteria:** patients who underwent Doppler and CA125 evaluation of pelvic masses and undergoing surgery **Clinical setting:** Unclear **Prior test:** Unclear Exclusions: Age <35 years excluded	**Design:** Prospective cross sectional study **Country:** Slovenia **Centre**: Single centre	# **participants in study:** 71 % **postmen (n):** 41 (29) # **ovarian cancer:**18 # **benign:** 53 # **borderline:**1 (with malignant) **Age mean+/-SD:** Not reported **Age median[range]:** Not reported [35-82]	**Pre-specified threshold:** No **Time interval:** 1 **week Blinded to reference:** unclear **Blinded to other index:** Unclear
Avsar 2008	**post**	**Study criteria:** Women presenting with adnexal mass scheduled for surgery **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** premenopausal women were excluded	**Design:** Unclear **Country:** Turkey **Centre:** Single centre	**# participants:** 46 % **postmen (n):** 100 (46) # **ovarian cancer:** 33 # **benign:** 13 # **borderline:** Not reported **Age mean+/-SD:** 46±8.4 **Age median[range]:** Not reported[Not reported]	**Prespecified threshold**: No **Time interval**: Unclear **Blinded to reference**: Unclear **Blinded to other index:** Unclear
Engelen 2008	**post**	**Study criteria:** Women with pelvic mass scheduled for surgery **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** Nil	**Design:** Prospective Cross sectional Study **Country:** Netherlands Centre: Single centre	**# participants:** 138 % **postmen (n):** 51.4 (71) # ovarian cancer: 30 # benign: 88 # borderline: 14 Age mean+/-SD: 51-all, Benign-47, Malignant-61 Age median[range]: [15-87]	**Prespecified threshold:** Yes **Time interval:** Unclear **Blinded to reference:** Unclear Blinded to other index: Unclear
Maggino 1993	**post**	**Study criteria:** Postmenopausal Women with pelvic mass **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** premenopausal women excluded	**Design:** Prospective Cross sectional Study **Country:** Italy **Centre:** Single centre	# **participants:** 335 % **postmen (n):** 100 (335) # **ovarian cancer:** 106 # **benign:** 229 # **borderline:** Not reported **Age mean+/-SD:** benign-62, Malignant-63 **Age median[range]:** Not reported[40-91]	**Prespecified threshold:** Yes **Time interval:** Unclear **Blinded to reference:** Unclear **Blinded to other index:** Unclear

Table 5.1 Continued...

Novotny 2012	**post**	**Study criteria:** Women with pelvic abnormalities **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** premenopausal women excluded	**Design:** Unclear **Country:** Czech republic **Centre:** Single centre	# **participants:** 256 % **postmen (n):** 100 (256) # **ovarian cancer:** 21 **# benign:** 256 # **borderline:** Not reported **Age mean+/-SD:** Benign-65.28, Malignant-64.37 **Age median[range]:** Benign-64, malignant-63[47-93]	**Prespecified threshold:** No **Time interval:** Unclear **Blinded to reference:** Unclear **Blinded to other index:** Not applicable
Ortiz-Munoz 2014	**post**	**Study criteria:** Women with gynaecological symptoms, diagnosed with primary ovarian cancer **Clinical setting:** Tertiary **Prior test:** Symptoms **Exclusions:** Nil	**Design:** Retrospective Cross sectional Study **Country:** Spain **Centre:** Single centre	# **participants:** 148 % **postmen (n):** 70% (104) # **ovarian cancer:** 29 # **benign:** 119 # **borderline:** not reported **Age mean+/-SD:** not reported **Age median[range]:** Benign premenopausal 39.5 +/- 8.4, postmenopausal 56 +/-11.5. Malignant premenopausal 40.5 +/- 5.8, postmenopausal 57 +/- 9.4[not reported]	**Prespecified threshold:** Yes **Time interval:** 1 day **Blinded to reference:** Unclear **Blinded to other index:** Unclear
Schutter 1994	**post**	**Study criteria:** Women presenting with adnexal mass and undergoing surgery. 199 women were detected on physical examination and the rest on ultrasound **Clinical setting:** Secondary **Prior test:** Unclear Exclusions: <45 years of age were excluded Women with Borderline tumours were excluded	**Design:** Prospective Cross sectional Study **Country:** Netherlands **Centre:** Multicentre	# **participants:** 228 % **postmen (n):** !00 (228) # **ovarian cancer:** 95 # **benign:** 127 # **borderline:** (6) Excluded **Age mean+/-SD:** 63 **Age median[range]:** Not reported[45-88]	**Prespecified threshold:** Yes **Time interval:** Unclear **Blinded to reference:** Unclear **Blinded to other index:** Unclear

Of the 30 studies, 17 were conducted in Europe[216, 217, 222, 223, 224, 219, 193, 196, 194, 192, 165, 221, 137, 225, 226, 139,] 3 in USA,[201, 214, 227, 220] one in China,[202] one in Australia,[228] one in Canada,[229] one in Argentina,[218] one in Brazil,[136] one in Japan,[187] one in Iran,[200] one in Turkey,[215] one in Thailand,[230] and one in Asia-Pacific.[186] Five studies were multicentre studies.[186,187,214, 221, 226] Women were recruited prospectively in less than 50% of the studies (13/29), the rest being retrospective (7) and unclear (9).

CA125 35U/ml (+/-10) was the most commonly used threshold reported in 26 of the included studies. Three of these studies also reported using CA125 at thresholds of 65 (+/-10).[228, 226, 219] Other three studies used CA125 at threshold of 65U/ml (+/-10).[225, 223, 224] One study used CA125 at threshold of 65U/ml (+/-10) for premenopausal women and at threshold of 35U/ml (+/-10) for postmenopausal women.[187]

Spectrum: presentation

All included studies recruited patients undergoing surgical investigation of an adnexal mass. Participants were mostly recruited from secondary care settings (gynaecology units or non-gynaecology departments in the hospital) (17 studies) or a tertiary care setting (specialised gynaecology cancer units) (5 studies). One study recruited its participants from a mixed setting that included both general gynaecology and specialist gynaecology oncology units. None of the included studies recruited participants from primary care setting. These highly selected populations are not representative of unselected, symptomatic women presenting in primary care where testing is initiated because of a suspicion of ovarian cancer.

Spectrum: target condition

Tests such as CA125 and HE4 are less sensitive in non EOC and 25% of the tumours are non EOC in women under 40 years;[4] therefore excluding non EOC tumours can potentially affect test performance and accuracy measures due to the altered disease spectrum. Five studies included only epithelial ovarian cancer (EOC).[201, 200, 187, 214,223] One study excluded solid pelvic masses.[192] Fifteen of the included studies did not report borderline tumours[230, 202, 200, 193, 227, 225, 139, 215, 219, 196, 194, 216, 223, 224, 229] and three excluded borderline tumours.[201, 137, 221]

Spectrum: age and menopausal status

The age of included participants ranged from 12[222] to 93 years old.[196] Information on participant's age by benign and malignant group was reported in 10/30 studies. The mean age for the benign group ranged from 33[202] to 65.3 years[196] and for the malignant group it varied from 51[200] to 64.4 years.[196] Four studies excluded premenopausal women from their study population[215, 219, 196, 221] and one study restricted inclusion to women over 35 years.[220] Postmenopausal women comprised 4100/7771 (52.8%) of the included review study population and across individual studies the proportion ranged from 108/414 (26%)[186] to 143/181 (79%).[137]

HE4 included studies

The study characteristics for 13/22 using HE4 in pre and or postmenopausal women are described in (Table 5.2).

Results for both pre and postmenopausal women were described in 13 studies. Of 12 studies in premenopausal women, all except one study[187] is also included in the 10 studies reporting results for postmenopausal women. Three studies include results from premenopausal women but not from postmenopausal women.[200, 214, 192] One study included results from postmenopausal women only.[187]

Twelve of the included studies used HE4 at threshold of 70pMol/L (+-10). Of the 12 studies, six of them also included HE4 140pMol/L (+-10). The one remaining study used only HE4 140pMol/L (+-10).[193]

Of the 13 studies reporting either pre or postmenopausal results for HE4, six were conducted in Europe,[138,193, 194, 192,137, 139] two in USA,[201, 214] one in China,[202] one in Brazil,[136] one in Iran,[200] one in Japan[187] and one in Asia-Pacific[186]. Three studies were multicentre studies.[214, 186, 187]

Table 5. 2: Table of study characteristics: HE4 pre and postmenopausal studies.

Author year	Menopausal status	Study criteria and setting	Study design	Participant characteristics	Index test
Anton 2012	**pre, post, all**	**Study criteria:** Women referred with pelvic masses diagnosed by US or CT or MRI undergoing surgery or image-guided biopsy when they presented with signs of carcinomatosis **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** Nil	**Design:** Prospective Cross sectional Study **Country:** Brazil **Centre:** Single centre	# **participants in study:** 120 % **postmen (n):** 60.8 (73) # **ovarian cancer:** 30 # **benign:** 66 # **borderline:** 17 **Age mean+/-SD:** Benign- 50.7, BOT-56.4, malignant-54.7 **Age median[range]:** Benign- 51, BOT-58, malignant-54[Not reported]	**Prespecified threshold:** Yes **Time interval:** 1 day **Blinded to reference:** Unclear **Blinded to other index:** Unclear
Bandiera 2011	**pre, post, all**	**Study criteria:** Not reported **Clinical setting:** Tertiary **Prior test:** Unclear **Exclusions:** BOT excluded. ? Non EOC excluded	**Design:** Unclear **Country:** USA **Centre:** Single centre	# **participants in study:** 278 % **postmen (n):** 65.8 (183) # **ovarian cancer:** 113 # **benign:** 165 # **borderline:** excluded **Age mean+/-SD:** Pre endo-36.5, benign-41.5, malignant-44.7. post ben- 64, malignant-66.3 **Age median[range]:** Not reported[25-89]	**Prespecified threshold:** Yes - HE4, No - CA125 **Time interval:** Unclear **Blinded to reference:** Unclear **Blinded to other index:** Not applicable
Chen 2015	**pre, post, all**	**Study criteria:** Women with pelvic masses scheduled for surgery **Clinical setting:** Unclear **Prior test:** Unclear **Exclusions:** Unclear	**Design:** Unclear **Country:** China **Centre:** Single centre	# **participants in study:** 232 % **postmen (n):** Not reported # **ovarian cancer:** 60 # **benign:** 70 # **borderline:** Not reported Age mean+/-SD: benign-33, Malignant-53 **Age median[range]:** Not reported[17-81]	**Prespecified threshold:** Yes **Time interval:** Unclear **Blinded to reference:** Unclear **Blinded to other index:** Unclear

Table 5.2 Continued...

Karlsen 2012	**pre, post, all**	**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 **Exclusions:** Nil	**Design:** Prospective Cross sectional Study **Country:** Denmark **Centre:** Single centre	# **participants in study:** 1218 % **postmen (n):** 51(621) # **ovarian cancer:** 261 # **benign:** 809 # **borderline:** 79 **Age mean+/-SD:** Not reported **Age median[range]:** 51[16-90]	**Prespecified threshold:** Yes **Time interval:** 2 weeks **Blinded to reference:** Unclear **Blinded to other index:** Unclear
Ortiz-Munoz 2014	**pre, post, all**	**Study criteria:** Women with gynaecological symptoms, diagnosed with primary ovarian cancer **Clinical setting:** Tertiary **Prior test:** Symptoms **Exclusions**: Nil	**Design:** Retrospective Cross sectional Study **Country:** Spain **Centre:** Single centre	# **participants in study:** 148 % **postmen (n):** 70% (104) # **ovarian cancer:** 29 # **benign:** 119 # **borderline:** not reported **Age mean+/-SD:** not reported **Age median[range]:** Benign premenopausal 39.5 +/- 8.4, postmenopausal 56 +/-11.5. Malignant premenopausal 40.5 +/- 5.8, postmenopausal 57 +/- 9.4[not reported]	**Prespecified threshold:** Yes **Time interval:** 1 day **Blinded to reference:** Unclear **Blinded to other index:** Unclear
Stiekema 2014	**pre, post, all**	**Study criteria**: histologically confirmed EOC or benign ovarian disease referred to the institute **Clinical setting:** Tertiary **Prior test:** Unclear **Exclusions:** BOT excluded	**Design:** Retrospective Cross sectional Study **Country:** Netherlands **Centre:** Single centre	# **participants in study:** 181 % **postmen (n):** &9 (143) # **ovarian cancer:** 147 # **benign:** 34 # **borderline:** Excluded Age mean+/-SD: Benign- 47, malignant- 57 **Age median[range]:** Not reported[Not reported]	**Prespecified threshold:** Yes **Time interval:** Unclear **Blinded to reference:** Unclear **Blinded to other index:** Unclear

Table 5.2 Continued...

Van Gorp 2011	**pre, post, all**	**Study criteria:** All patients diagnosed with pelvic mass undergoing surgery **Clinical setting:** Unclear **Prior test:** Unclear **Exclusions**: Nil	**Design:** Prospective Cross sectional Study **Country:** Belgium **Centre:** Single centre	# **participants in study:** 389 % **postmen (n):** 41.4%(161) # **ovarian cancer:** 161 # **benign:** 228 # **borderline:** Not reported Age mean+/-SD: Benign:46.3+/-16 Malignant:57.8+/-12.6 Age median[range]: Not reported[Not reported]	**Prespecified threshold:** Yes **Time interval: 1** day **Blinded to reference:** Unclear **Blinded to other index:** Yes
Molina 2011	**pre, post**	**Study criteria:** Not reported **Clinical setting:** Unclear **Prior test:** Unclear **Exclusions**: Unclear	**Design:** Retrospective Cross sectional Study **Country:** Spain **Centre:** Single centre	# **participants in study:** 422 % **postmen (n):** 34(143) # **ovarian cancer:** 111 # **benign**: 285 *benign gyn disease with 137 ovarian cysts # **borderline:** Not reported Age mean+/-SD: Not reported **Age median[range]:** Benign gyn disease-40+/- 0.8, Gyn cancer61+/-1.2[17-90]	**Prespecified threshold:** Yes **Time interval:** Unclear **Blinded to reference:** Unclear **Blinded to other index:** Unclear
Chan 2013	**pre, all**	**Study criteria:** Women over 18 years diagnosed with adnexal mass diagnosed by any Imaging method (US, CT or MRI) **Clinical setting:** Unclear **Prior test:** Unclear **Exclusions:** Nil	**Design:** Prospective Cross sectional Study **Country:** Asia- pacific **Centre:** Multicentre	# **participants in study:** 414 % **postmen (n):** 26%(108 # **ovarian cancer:** 74 # **benign:** 322 # **borderline:** 16 **Age mean+/-SD:** Not reported **Age median[range]:** Not reported[Not reported]	**Prespecified threshold:** yes **Time interval:** Unclear **Blinded to reference:** Unclear **Blinded to other index:** Unclear

Table 5.2 Continued...

Farzaneh 2014	**pre, all**	**Study criteria:** Women with adnexal mass undergoing surgery **Clinical setting:** Secondary **Prior test:** Unclear **Exclusions:** Non epithelial Ovarian cancer was excluded	**Design:** Prospective Cross sectional Study **Country:** Iran **Centre:** Single centre	# **participants in study:** 100 % **postmen (n):** 31(31) # **ovarian cancer:** 43 # **benign:** 56 # **borderline:** Not reported **Age mean+/-SD:** 44+/-16 **Age median[range]:** Not reported[Not reported]	**Prespecified threshold:** No **Time interval:** Unclear **Blinded to reference:** Unclear **Blinded to other index:** Unclear
Holcomb 2011	**pre**	**Study criteria:** Women with adnexal mass undergoing surgery **Clinical setting:** Unclear **Prior test:** Unclear **Exclusions:** postmenopausal women excluded ?non EOC excluded)	**Design:** Prospective Cross sectional Study **Country:** USA **Centre:** Multicentre	# **participants in study:** 229 % **postmen (n):** 53.6 (265) # **ovarian cancer:** 18 # **benign:** 195 # **borderline:** 16 **Age mean+/-SD:** Not reported **Age median[range]:** Not reported[Not reported]	**Prespecified threshold:** Yes* (but reported for diff cut off than the prespecified) **Time interval:** Less than 30 days **Blinded to reference:** Yes **Blinded to other index:** Not applicable
Partheen 2011	**Pre**	**Study criteria:** Women with complex cystic mass and suspicious of malignancy undergoing surgery Clinical setting: Tertiary Prior test: Unclear **Exclusions**: Solid anpelvic mass were not eligible	**Design:** Prospective Cross sectional Study Country: Sweden Centre: Single centre	**# participants in study:** 374 % **postmen (n):** 73.7 (276) # **ovarian cancer:** 108 # benign: 215 # borderline: 45 Age mean+/-SD: Not reported **Age median[range]:** Not reported[Not reported]	**Prespecified threshold:** Yes **Time interval:** 1 day **Blinded to reference:** Yes **Blinded to other index:** Unclear
Fujiwara 2015	**post**	**Study criteria:** Women with adnexal mass diagnosed on Ultrasond/CT/MRI/PET and scheduled for surgery **Clinical setting:** Unclear **Prior test:** Unclear **Exclusions:** Women with previous history of any gynaecological disease and NON EOC excluded	**Design:** Prospective Cross sectional Study **Country:** Japan **Centre:** Multicentre	# **participants:** 221 % **postmen (n):** 40 (90) # **ovarian cancer:** 71 # **benign:** 131 # **borderline:** 19 **Age mean+/-SD:** Benign-43, BOT-47, Malignant-54 **Age median[range]:** Not reported[20-79]	**Prespecified threshold:** Yes **Time interval:** less than 42 days **Blinded to reference:** Unclear **Blinded to other index:** Not applicable

Spectrum: presentation

All included studies recruited patients undergoing surgical investigation of an adnexal mass. Participants were recruited from either secondary care setting – 3 studies (gynaecology units or non-gynaecology departments in the hospital) or tertiary care setting – 4 studies (specialised gynaecology cancer units) and 6 unclear if it was secondary or tertiary care setting. None of the included studies recruited participants from primary care setting.

Spectrum: target condition

Four studies included only epithelial ovarian cancer (EOC).[201, 200, 214, 187] One study excluded solid pelvic masses.[192] Five of the included studies did not report borderline tumours[202, 194, 139, 193, 200] and two excluded borderline tumours.[201, 137]

Spectrum: age and menopausal status

The age of included participants ranged from 16[138] to 90 years old.[138, 193] The mean age for the benign group ranged from 33[202] to 52.8 years[201] and for the malignant group it varied from 53[202] (Chen 2015) to 57.8 years[139] as reported in six studies. One study excluded postmenopausal women from the study population[214]. The proportion of postmenopausal women in the included studies that report both pre and postmenopausal women varied from 108/414 (26%)[186] to 143/181 (79%)[137]. Overall the postmenopausal women were 2189/4326 and represented 50.6% of the included study population.

OVA1 included studies

The study characteristics of two included studies using OVA1 in pre and postmenopausal women separately[188, 205] are described in Table 5.3

Table 5. 3: Table of study characteristics: OVA1.

Marker	Author year	Menopausal status	Study criteria and setting	Study design	Participant characteristics	Index test
OVA1	**Grenache 2015**	Pre, post	**Study criteria:** Women with abnormal adnexal mass detected on physical examination and Imaging Included Ultrasound, CT or MRI) followed by surgery **Clinical setting:** Unclear **Prior test:** Unclear **Exclusions:** Unclear	**Design:** Cross sectional study based on two previous clinical trials **Country:** USA **Centre:** Multicentre	# **participants:** 146 % **postmen (n):** 52 (76) # **ovarian cancer:** 19 # **benign:** 115 # **borderline:** 7 **Age mean+/-SD:** 52 **Age median[range]:** Not reported[18-89]	**Prespecified threshold:** Yes **Time interval:** Unclear **Blinded to reference:** Unclear **Blinded to other index:** Unclear
OVA1	**Bristow 2013**	Pre, post	**Study criteria:** Women undergoing surgery within 3 months of a documented pelvic mass on imaging (CT, US or MRI) **Clinical setting:** Mixed **Prior test:** Unclear **Exclusions:** Nil	**Design:** Prospective cross sectional study **Country:** USA **Centre:** Multicentre	# **participants:** 490 % **postmen (n):** 43.9 (217) # **ovarian cancer:** 65 # **benign:** 402 # **borderline:** 17 **Age mean+/-SD:** 48.6 (SD 14) **Age median[range]:** 48[18-87]	**Prespecified threshold:** Yes Time interval: less than or equal to 3 months Blinded to reference: Unclear Blinded to other index: Unclear

Both studies were conducted in USA at multiple centres.

Spectrum: presentation

Both OVA1 studies recruited patients undergoing surgical investigation of adnexal masses and recruited participants from centres that included both general gynaecology and specialist gynaecology oncology units.[205]

Spectrum: target conditions

Both studies reported a range of ovarian pathology including non EOC and borderline tumours.

Spectrum: age and menopausal status

The age of the participants in these studies ranged from 18-89 years old. Both studies presented results for pre and postmenopausal women separately. The proportion of postmenopausal women was 76/146 (52%)[188] and 505/1024 (44%).[205]

Only one study was identified that presented results for the biomarkers HCG, LDH and AFP. However, this study only reported results using a combination of all these biomarkers with CA125, so these results were not eligible to be included in this review, but were considered for the review of combination tests.

5.2.2 Methodological quality of included studies

The methodological quality for CA125, HE4 and OVA1 for studies reporting separately results for pre and postmenopausal women are summarised in (Figure 5.1, Figure 5.2, Figure 5.3) and graphs (Appendix 13 and 14). The data was extracted using a QUADAS-2 criteria proforma tailored to the review topic (Appendix 3) focusing on four domains of methodological quality: patient selection; index test; reference standard; and flow and timing. Tailoring of QUADAS-2 to the clinical topic is already explained in detail in the test combination review chapter.

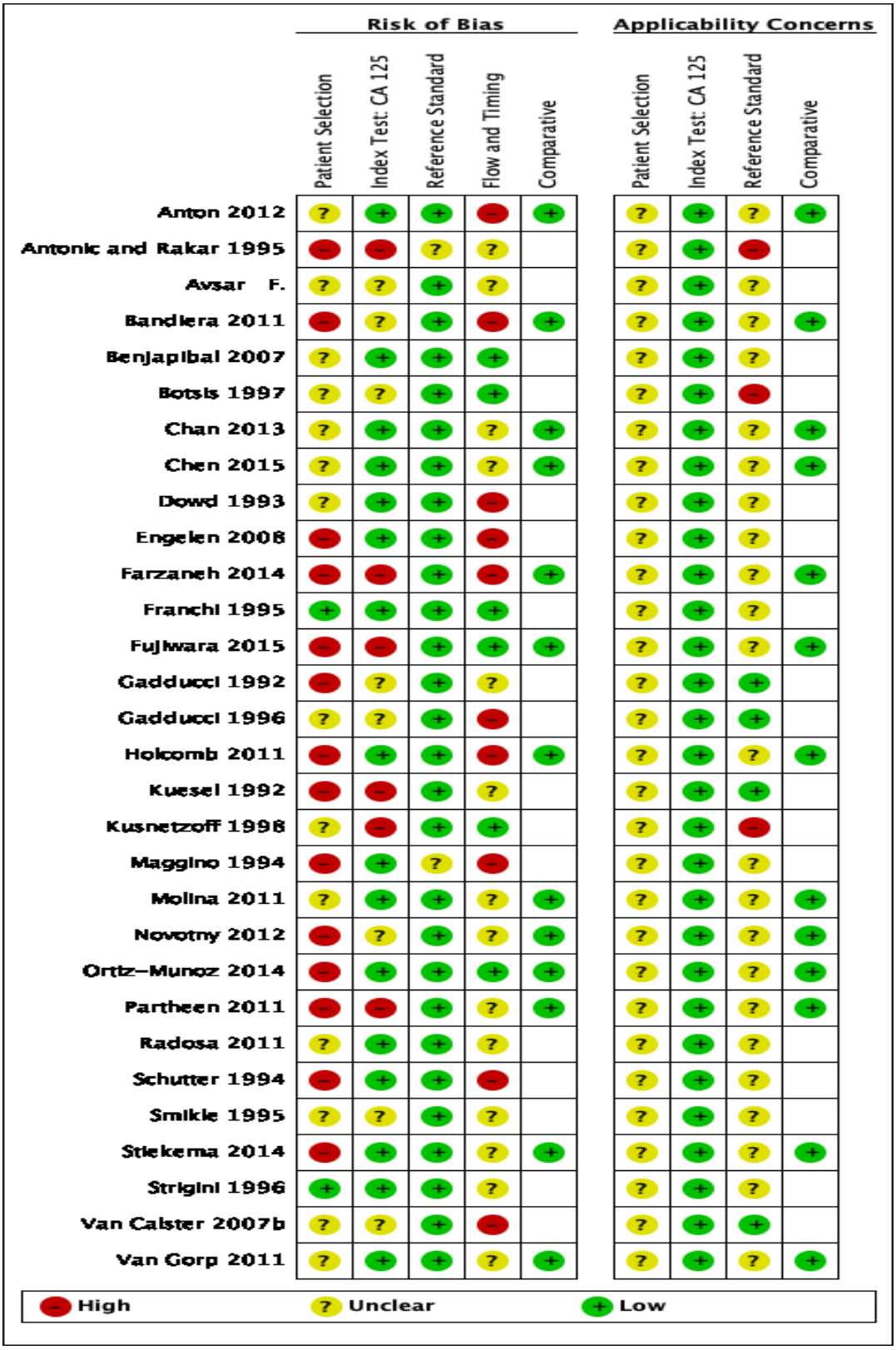


Figure 5. 1: CA125 QUADAS-2: Pre and postmenopausal studies bias applicability concerns summary.

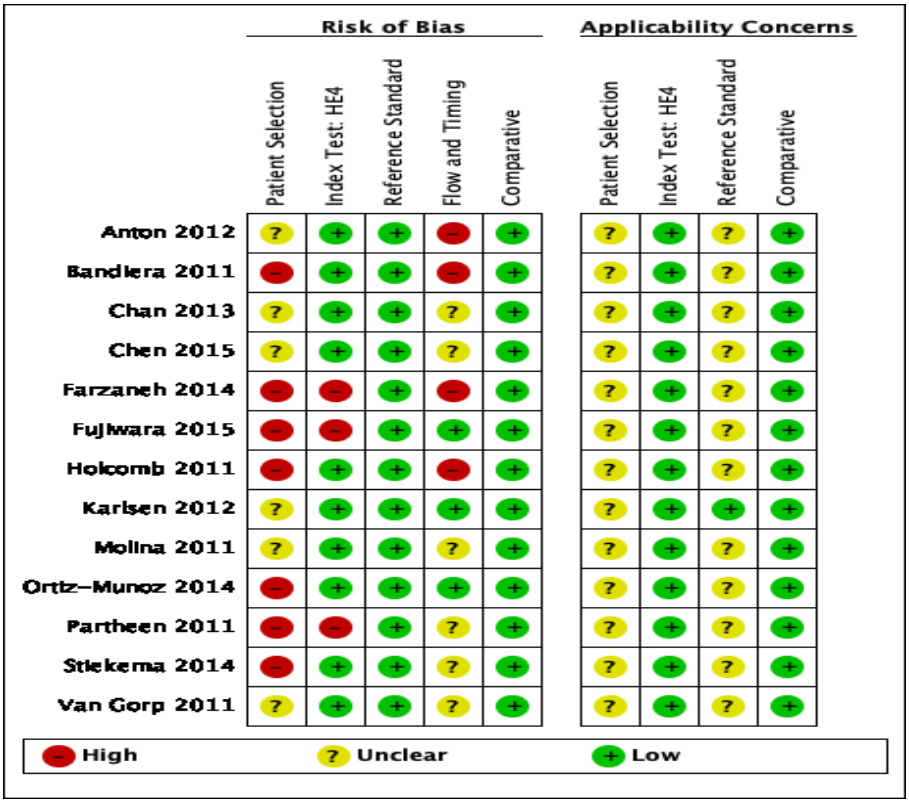


Figure 5. 2: HE4 QUADAS-2: Pre and postmenopausal studies risk of bias and applicability concerns summary.

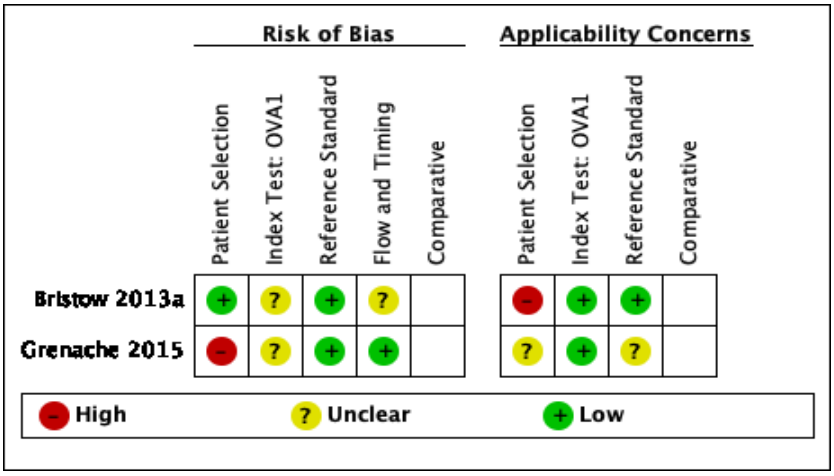


Figure 5. 3: OVA1 QUADAS-2: Pre and postmenopausal studies risk of bias and applicability concerns summary.

Methodological quality of studies including CA125

The prevalence of ovarian cancer in the studies was high, indicating an advanced position of the study in the patient pathway. The median prevalence of ovarian cancer in premenopausal women was 22% (IQR 11% to 31%, range 3 to 81%). The median prevalence of ovarian cancer in postmenopausal women was 46% (IQR 32% to 58%, range 8 to 71%).

QUADAS data for the included pre and postmenopausal studies (30) only is shown in Figure 5.1 and appendix 13

Patients were recruited prospectively in 14 studies[136, 186, 187, 192, 200, 214, 216, 217, 220, 230, 219, 221, 139, 218] retrospectively in 7 studies [229, 137, 193, 194, 226-228] and recruitment was unclear in 9 studies.[201, 202, 215, 222-224, 165, 196, 225] 21 studies used a cross sectional study design, where in 9 studies the study was unclear.[223, 224 165, 196, 201, 202, 215, 222]

Risk of bias

For the patient selection domain all the studies were judged as unclear or high risk of bias.[222, 225] For studies rated as unclear or high risk of bias this was most often because of a lack of an explicit mention or absence of consecutive sampling in combination with a limited range of tumour types and benign pathology presented and lack of clarity on prior testing.

For the index test domain nearly 50% of the studies (14/30) were judged as unclear or high risk of bias, the most common reason for was lack of clarity regarding prespecified threshold for test positivity

For the reference standard domain all studies were judges as low risk of bias.

In the flow and timing domain only 9/30 studies were judged at low risk; the commonest reason for unclear or high risk of bias was lack of clarity of the interval between the conduct of the index test and the reference standard whereas those studies (12) judged at high risk of bias in this domain did not include all patients in their analysis.

Applicability concerns

Participant selection, all studies were judged as unclear risk due to lack of clarity on patients presenting symptomatology. The prevalence of ovarian cancer in the studies was high, indicating an advanced position of the study in the patient pathway. The median prevalence of ovarian cancer in premenopausal women was 22% (IQR 11% to 31%, range 3 to 81%, 19 studies). The median prevalence of ovarian cancer in postmenopausal women was 46% (IQR 32% to 58%, range 8 to 71%, 22 studies).

All studies were judged as low applicability concern for the index test domain because the index test is objective and not dependent on operator skill.

For the reference standard domain 4/30 studies[223, 224, 226, 229] were judged as low applicability concern and 3 as high applicability concern[216, 218, 220] as it was unclear how borderline tumours had been handled for the paper's analysis.

Methodological quality of studies including HE4

The prevalence of ovarian cancer in the studies was high, indicating an advanced position of the study in the patient pathway. The median prevalence of ovarian cancer in these premenopausal women was 25% (IQR 13% to 30%, range 8 to 81%). The median prevalence of ovarian cancer in these postmenopausal women was 50% (IQR 45% to 59%, range 18 to 81%).

QUADAS-2 data was for 13 HE4 studies, is shown in Figure 5.2 and Appendix 14.

Risk of bias

Patients were recruited prospectively in 8 studies,[136, 138, 139, 186, 187, 192, 200, 214] retrospectively in 3 studies[137, 193, 194] and recruitment was unclear in 2 studies.[201, 202] 11 studies used a cross sectional study design, and for 2 studies the study design was unclear.[201, 202]

For the patient selection domain, all studies had inappropriate exclusions as summarised in the table of study characteristics (Table 5.2) and were judged as unclear or high risk of bias for patient selection

10/13 studies were judged as low risk of bias for the index test domain. 3 studies[187, 192, 200] were judged as high risk as the threshold was not specified.

All the included studies were judged as low risk of bias for the reference standard domain as only surgical patients were included in all the included studies.

For the flow and timing domain 6 studies were judged at high risk of bias because not all patients were included in the analysis.

Applicability concerns

No studies were judged as having a low applicability concern for the participant selection domain due to the fact that study participants did not obviously represent symptomatic women presenting in primary care. The prevalence of ovarian cancer in the studies was high, indicating an advanced position of patients in the clinical pathway. The median prevalence of ovarian cancer in premenopausal women was 25% (IQR 13% to 30%, range 8 to 81%, 12 studies). The median prevalence of ovarian cancer in postmenopausal women was 50% (IQR 45% to 59%, range 18 to 81%) from 10 studies.

All the studies were judged as low applicability concern for the index test domain because the index test is objective and not dependent on operator skill. For the reference standard domain all studies except 1[138] were judged as having an unclear applicability concern as it was unclear how borderline tumours had been handled for the paper's analysis. (Figure 5.2)

Methodological quality of studies including OVA1

Prevalence of ovarian cancer in the 2 studies was 17% [93] and 18%.[188]

Both OVA1 studies presented test accuracy data separately for pre and postmenopausal women. Figure 5.3 is the QUADAS-2 plot for these studies.

Risk of bias

One study[188] was at high risk of bias in the patient selection domain as patients were not recruited randomly/consecutively and the other unclear as it was unclear if there was consecutive recruitment of patients. In both studies patients were recruited prospectively. One study used a cross sectional design[93] and in the other study accuracy data was obtained from

clinical trial data.[188] Risk of bias was low for the index test domain for both studies due to objectivity of standardised serum testing.

Applicability concerns

Both studies were judged as having as unclear applicability concern for the participant selection domain due to the fact that study participants did not obviously represent symptomatic women presenting in primary care. Applicability concerns were low for the index test domain for both studies due to objectivity of standardised serum testing. Both studies grouped borderline results with malignant and were therefore classified as low applicability concern in the reference standard domain.

For CA125, HE4, and OVA1 results were examined at commonly used and where these existed clinically relevant test thresholds. Results of test performance compared at these thresholds demonstrate the effect of test threshold on test performance (e.g. Figure 5.4, Figure 5.5).

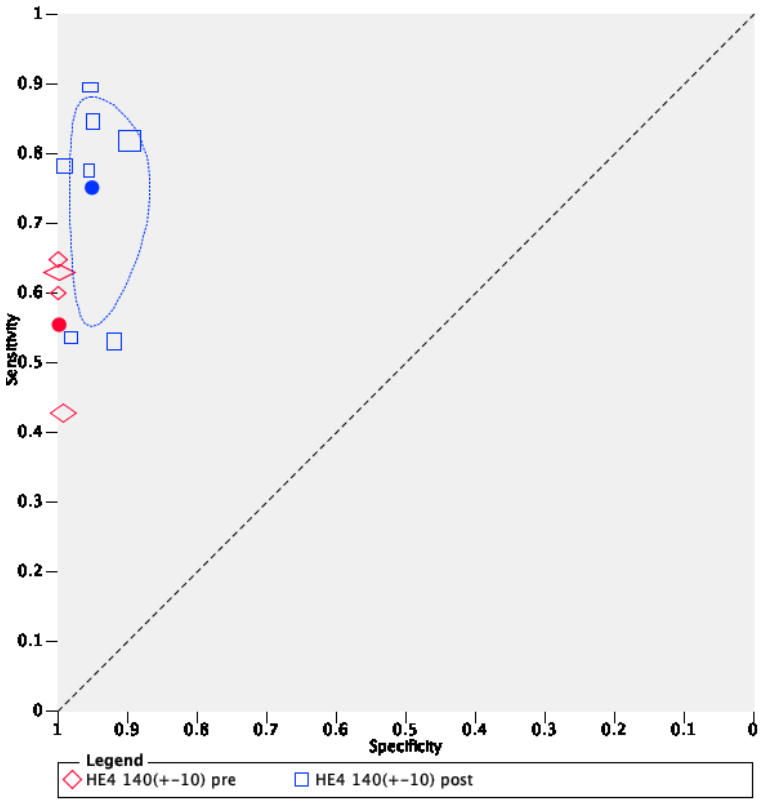


Figure 5. 4: Summary ROC Plot of tests: HE4 140 (+-10) pMol/L for pre and postmenopausal women.

5.3 Test positivity thresholds used in included studies

For CA125, the thresholds investigated were 35U/ml (RCOG recommended for postmenopausal women), 65U/ml (modified ACOG) and 200U/ ml (recommended by RCOG for premenopausal women). In addition, results within a 5U range of the test thresholds 35U/ml (30-40U/ml) and 65U/ml (60-70U/ml) were investigated.

For HE4 there is no consensus on the clinically agreed threshold cut off for test positivity. Therefore, thresholds recommended by the manufacture and most commonly reported in the literature (70pMol/L and 140pMol/L) were analysed. In addition, two analyses including results within a 10U range of these test thresholds (60-80pMol/L and 130-150pMol/L) were conducted. Results of all test positivity thresholds are reported in Table 5.9, forest plot in Appendix 24 and 25.

Well documented differences in spectrum in pre compared to postmenopausal women exist which means that differences in the percentage of postmenopausal women across studies is a source of between study heterogeneity; the most accurate test in premenopausal women might not be the most accurate test in postmenopausal women. The observed effect of menopausal status on test performance is shown in Tables 5.5 and Table 5.7 (e.g. Figure 5.4, Figure 5.5, Figure 5.6 and Figure 5.7). A significant difference in test performance was demonstrated for pre compared to postmenopausal women. Therefore, the results are presented separately for premenopausal and postmenopausal women, based on the 43 studies presenting results stratified by menopausal status

In addition to threshold and menopausal status, additional analyses were conducted to investigate the effects on test accuracy of management of borderline tumours in individual study analyses and comparison of tests based on direct (within study) comparisons and indirect (across study) comparisons (see Appendix 15, Table 5.5 and 5.7). Within study comparisons are more valid than between study comparisons as they are less prone to confounding resulting from variation in population characteristics, index test conduct and study design. The management of borderline tumours in included studies was variable. In this review grouping of borderline tumours with malignant tumours was considered an appropriate analysis strategy reflecting management of borderline tumours in clinical practice. Studies excluding borderline tumours from analysis were considered clinically inappropriate and in some studies

management of borderline tumours for analyses was unclear. For CA125, only at a threshold of 35 U/ml there was sufficient evidence to give estimates of sensitivity and specificity separately in pre and postmenopausal women for studies using clinically appropriate and inappropriate or unclear grouping of borderline results (see Appendix 15). This analysis demonstrated that estimates of sensitivity in CA125 35U/ml studies using appropriate grouping of borderline tumours were lower than in studies using inappropriate grouping or where the analysis strategy was unclear. Estimates of sensitivity and specificity for other tests and thresholds are therefore likely to be overestimates, and this should be kept in mind when interpreting the results.

5.3.1 Accuracy of CA125

Diagnostic accuracy was investigated at three commonly used thresholds of 35U/ml, 65U/ml and 200U/ml with two sensitivity analyses including results within a 5U range of the test thresholds (30- 40U/ml and 60-70U/ml). Average sensitivity and specificity values for thresholds within a 5 U range were unchanged compared to analysis based on exact threshold values (refer to table illustrating sensitivity and specificity at both point thresholds and threshold ranges). The discussion below is therefore focussed on results at two threshold ranges of 30U/ml (+/-5) and 65U/ml (+/-5). Table 5.4 presents diagnostic accuracy results for CA125 at each prespecified test threshold by menopausal status. The table reports the number of studies, the total number of women and women with ovarian cancer contributing to each summary estimate of diagnostic accuracy.

Table 5. 4: Test performance: CA125 at common thresholds.

Menopausal status	Test	Threshold	Number studies (N total, N OC)	Sensitivity (95% CI)	Specificity (95% CI)
Pre	CA125	35	15 (2514, 376)	80 (72 to 86)	66 (60 to 72)
Post	CA125	35	16 (2526, 993)	84 (79 to 87)	83 (81 to 85)
Pre	CA125	35ish	17 (3042, 473)	78 (71 to 84)	66 (61 to 71)
Post	CA125	35ish	21 (3303, 1259)	84 (80 to 87)	83 (81 to 84)
Pre	CA125	35ish	7	71 (58 to 82)	70 (62 to 77)
Post	CA125	35ish	9	78 (74 to 82)	82 (80 to 84)
Pre	CA125	65	3 (641, 135)	NA	NA
Post	CA125	65	4 (876, 368)	NA	NA
Pre	CA125	65ish	4 (763, 165)	60 (49 to 71)	85 (79 to 89)
Post	CA125	65ish	4 (876, 368)	73 (68 to 77)	92 (89 to 94)
Pre	CA125	200	1 (445, 86)	29 (20 to 40)	98 (96 to 99)
Post	CA125	200	1 (364, 156)	54 (46 to 62)	97 (94 to 99)

Table 5. 5: Test performance: CA125 comparisons.

Menopausal status	Test 1 (Threshold)	Test 2 (Threshold)	Number studies (N total, N OC)	Sensitivity (95% CI)	Specificity (95% CI)	Notes
Post- pre	CA125 35ish	CA125 35ish	102 (17 pre and 21 post, 64 all)	5.0 (-2.6 to 12.7)	16.4 (10.4 to 22.5)	Univariate, 1 RE are zero. Estimates and 95% CI within 2% of single models
Post- pre	CA125 35ish	CA125 35ish	15	5.5 (-2.9 to 14.0)	15.5 (9.2 to 21.8)	Direct studies
Post- pre	CA125 65ish	CA125 65ish	32 (4 pre and 4 post, 24 all)	13.0 (1.0 to 24.9)	6.8 (1.3 to 12.3)	Univariate, 2 RE are zero
Post- pre	CA125 65ish	CA125 65ish	3	18.6 (7.4 to 29.8)	4.6 (0.3 to 8.8)	Direct studies
Pre (35i[SM1] sh minus 65ish)	CA125 65ish	CA125 35ish	21	18.1 (4.0 to 32.2)	-18.5 (-25.9 to -11.2)	Bivariate
Pre (35ish minus 65ish)	CA125 65ish	CA125 35ish	2 (566,129)	13.5 (0.1 to 26.8)	-19.5 (-24.8 to -14.1)	Direct studies
Post (35ish minus 65ish)	CA125 65ish	CA125 35ish	25	10.3 (4.6 to 16.0)	-9.3 (-12.2 to -6.4)	Univariate, 2 zero RE estimates, all within 1% of single models
Post (35ish minus 65ish)	CA125 65ish	CA125 35ish	3	7.9 (1.8 to 14.0)	-11.3 (-15.5 to -7.0)	Direct studies

CA125: Threshold 30U/ml to 40U/ml

Figure 5.5 presents study results for pre and postmenopausal women with CA125 at a threshold of within 5U/ml of 35U/ml in ROC space. The corresponding forest plot is included in the Appendix 16.

For premenopausal women in studies with appropriate grouping of borderline tumours (borderline tumours grouped with malignant tumours), the average sensitivity was 71% (95% CI 58 to 82) with an average specificity was 70% (95% CI 62 to 77), based on 9 studies (Appendix 15). In premenopausal women for all studies sensitivity was higher [78% (95% CI 71 to 84)] and specificity lower [(66% (95% CI 61 to 71) (Appendix 15).

For postmenopausal women in studies with appropriate grouping of borderline tumours the average sensitivity was 78% (95% CI 74 to 82) with an average specificity of 82% (95% CI 80 to 84), based on 9 studies (Appendix 15). In postmenopausal women for all studies sensitivity was higher [84% (95% CI 80 to 87)] and specificity lower [83% (95% CI 81 to 84)] (Appendix 15).

Results in postmenopausal women are less heterogeneous than results in premenopausal women, and heterogeneity in sensitivity is greater in both pre and postmenopausal women when all studies are included compared to estimates based on sub groups of studies using appropriate grouping of borderline tumours in study analyses (see ROC plots Figure 5.5 and Forest plots Appendix 16 referred to above).

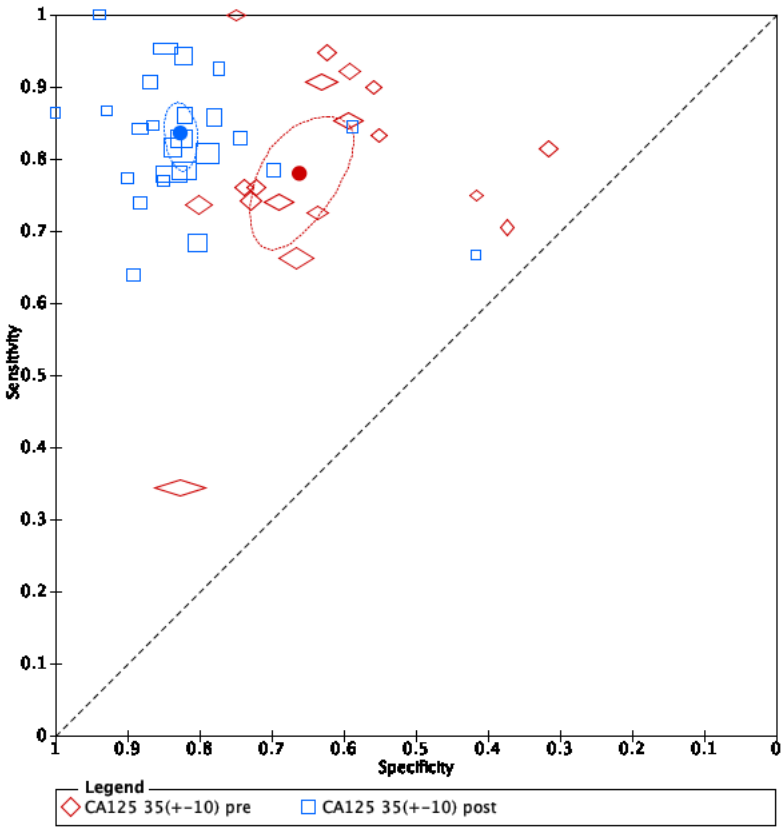


Figure 5. 5: Summary ROC Plot of tests: CA125 35(+-10) U/ml for pre and postmenopausal women.

CA125: Threshold 60U/ml to 70U/ml

Figure 5.6 presents study results for pre and postmenopausal women with CA125 at a threshold of within 5U/ml of 65U/ml in ROC space. The corresponding forest plot is included in the Appendix 17.

At a threshold of within 5U/ml of 65U/ml, in premenopausal women the average sensitivity was 60% (95% CI 49 to 71) with an average specificity of 85% (95% CI 79 to 89), based on 4 studies with 763 women, 165 with ovarian cancer (Table 5.4). Confidence intervals for summary estimates are wide as meta-analysis is based on only 4 studies; individual study results for sensitivity range from (95% CI 49 to 71) 50% to 73% and specificity results vary from 75% to 90%.

For postmenopausal women the average sensitivity was 73% (95% CI 68 to 77) with an average specificity of 92% (95% CI 89 to 94), based on 4 studies with 876 women, 368 with ovarian cancer (Table 5.4).

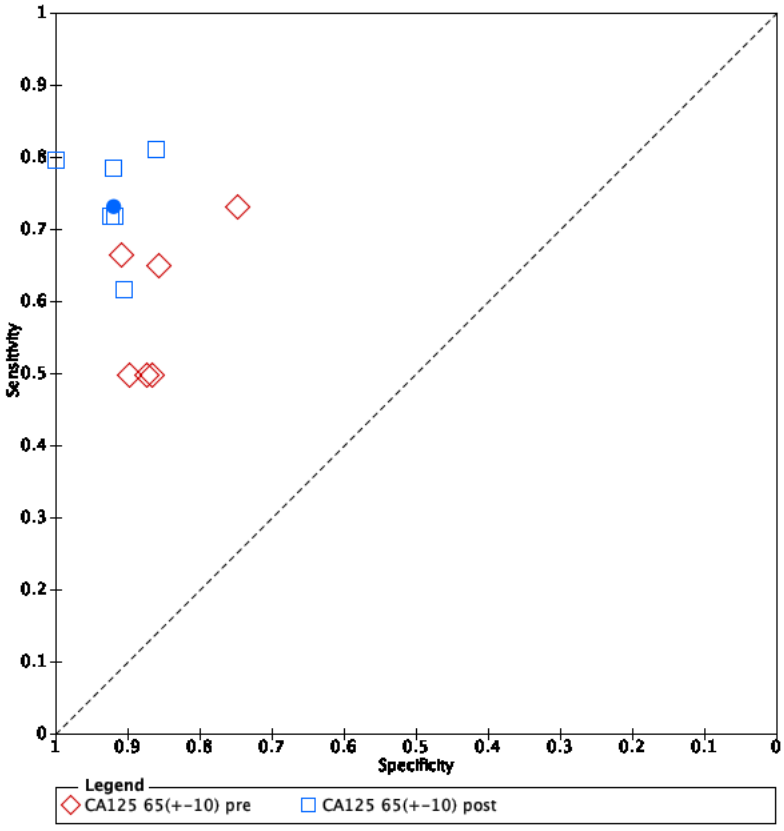


Figure 5. 6: Summary ROC Plot of tests: CA125 65(+-10) U/ml for pre and postmenopausal women.

CA125: Threshold 200U/ml

Only 1 study reported results for CA125 at a threshold of 200U/ml for pre and postmenopausal results (Table 5.4). For premenopausal women sensitivity was 29% (95% CI 20 to 40) and specificity was 98% (95% CI 96 to 99). For postmenopausal women sensitivity was 54% (95% CI 46 to 62) and specificity was 97% (95% CI 94 to 99). This study undertook appropriate grouping of borderline tumours in the analysis, is at low risk of bias and therefore estimates of sensitivity are unlikely to be overestimated.

5.3.2 Accuracy of HE4

Average sensitivity and specificity values for thresholds within a 10pMol/L range of 70pMol/L and 140pMol/L (60-80pMol/L and 130-150pMol/L) were unchanged compared to analysis based on an exact threshold values (see Table 5.6 illustrating sensitivity and specificity at both point thresholds and thresholds ranges). The discussion below is therefore focussed on results at two threshold ranges of 70pMol/L (+/-10) and 140pMol/L (+/-10). Table 5.6 presents diagnostic accuracy results for HE4 at each prespecified test threshold by menopausal status. The table reports the number of studies, the total number of women and women with ovarian cancer contributing to each summary estimate of diagnostic accuracy.

Table 5. 6: Test performance: HE4 at common thresholds.

Menopausal status	Test	Threshold	Number studies (N total, N OC)	Sensitivity (95% CI)	Specificity (95% CI)
Pre	HE4	70	7 (1467, 225)	73 (67 to 79)	91 (86 to 94)
Post	HE4	70	3 (417, 268)	85 (76 to 91)	77 (48 to 93)
Pre	HE4	70ish	11 (1718, 297)	74 (68 to 79)	91 (87 to 94)
Post	HE4	70ish	5 (601, 328)	84 (76 to 89)	78 (63 to 88)
Pre	HE4	140	2 (112, 30)	NA	NA
Post	HE4	140	5 (927, 392)	77 (66 to 84)	96 (91 to 98)
Pre	HE4	150	2 (437, 69)	?	?
Post	HE4	150	2 (348, 203)	?	?
Pre	HE4	140ish	4 (549, 99)	55 (43 to 67)	100 (98 to 100)
Post	HE4	140ish	7 (1275, 595)	75 (64 to 83)	95 (91 to 97)

Table 5. 7: Test performance: HE4 comparisons.

Menopausal status	Test 1 (Threshold)	Test 2 (Threshold)	Number studies (N total, N OC)	Difference in Sensitivity (95% CI) (Test1-Test2)	Difference in Specificity (95% CI) (Test1-Test2)	Notes
Post- Pre	HE4 70ish	HE4 70ish	16 (11 pre, 5 post)	10.2 (1.4 to 19.0)	-14.0 (-26.8 to -1.1)	Univariate, all studies
Post- Pre	HE4 70ish	HE4 70ish	4	13.2 (2.1 to 24.3)	-11.5 (-27.6 to 4.7)	Univariate, direct studies only *RE at zero for some variables, but sens and spec within 2% of bivariate above for pre
Post- Pre	HE4 140ish	HE4 140ish	20	18.9 (3.0 to 34.8)	-4.8 (-7.8 to -1.9)	Bivariate
Post- Pre	HE4 140ish	HE4 140ish	8	21.5 (3.1 to 39.8)	-5.7 (-8.7 to -2.8)	Direct only Univariate. Estimates within 1% of single test values
Pre	HE4 70ish	HE4 140ish	12	18.9 (5.7 to 32.1)	-8.0 (-11.7 to -4.3)	Bivariate, 1 RE of zero. All values within 1% estimates and 2% 95% CI from single tests
Table 5.7 Continued...			3 (296, 72)	23.0 (3.0 to 43.0)	-7.1 (-10.7 to -3.6)	Direct only, univariate
Post	HE4 70ish	HE4 140ish	10 (5 70ish, 7 140ish, 2 overlap)	9.4 (-2.3 to 21.1)	-17.2 (-29.9 to -4.6)	Univariate, all values within 1% estimates and 2% 95% CI from single tests
Post	HE4 70ish	HE4 140ish	2 (309, 138)	7.4 (-22.5 to 37.3)	-15.3 (-25.7 to -4.9)	Direct only, univariate

HE4: Threshold 60 pMol/L to 80 pMol/L

Figure 5.7 presents individual study results for pre and postmenopausal women with HE4 at a threshold of within +/- 10pMol/L of 70pMol/L in ROC space. The corresponding forest plot is included in the Appendix 18.

For premenopausal women the average sensitivity was 74% (95% CI 68 to 79) with an average specificity of 91% (95% CI 87 to 94), based on 11 studies with 1718 women, 297 with ovarian cancer, at a threshold of 60-80pMol/L. These estimates of sensitivity may be inflated by grouping of borderline results (Figure 5.8 and corresponding forest plot Appendix 22).

For postmenopausal women the average sensitivity was 84% (95% CI 76 to 89) with an average specificity of 78% (95% CI 63 to 88) based on 5 studies with 601 women, 328 with ovarian cancer, for a threshold of 60-80pMol/L. The 95% CI for specificity is wide (78% specificity, 95% CI 63 to 88; Table 2), indicating the high heterogeneity in these studies. These estimates of sensitivity may be inflated by grouping of borderline results. (e.g. Figure 5.9 and corresponding forest plot Appendix 23).

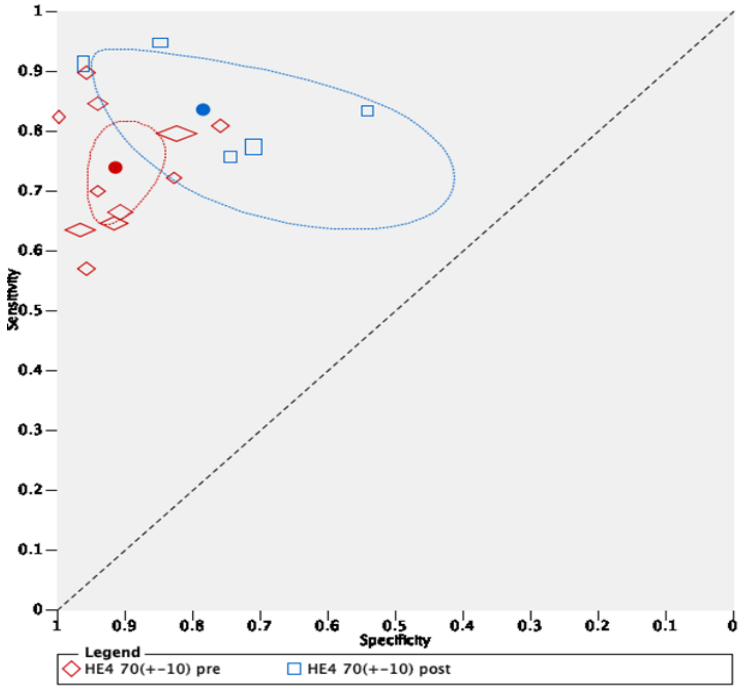


Figure 5. 7: Summary ROC Plot of tests: 11 HE4 70(+-10) pMol/L for pre and postmenopausal women.

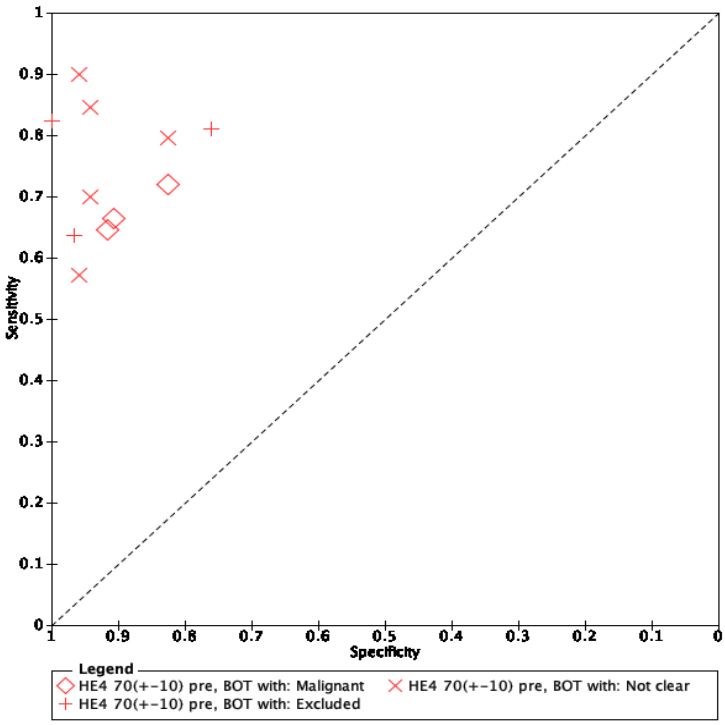


Figure 5. 8: Summary ROC Plot of 11 HE4 70(+-10) BOT for premenopausal women.

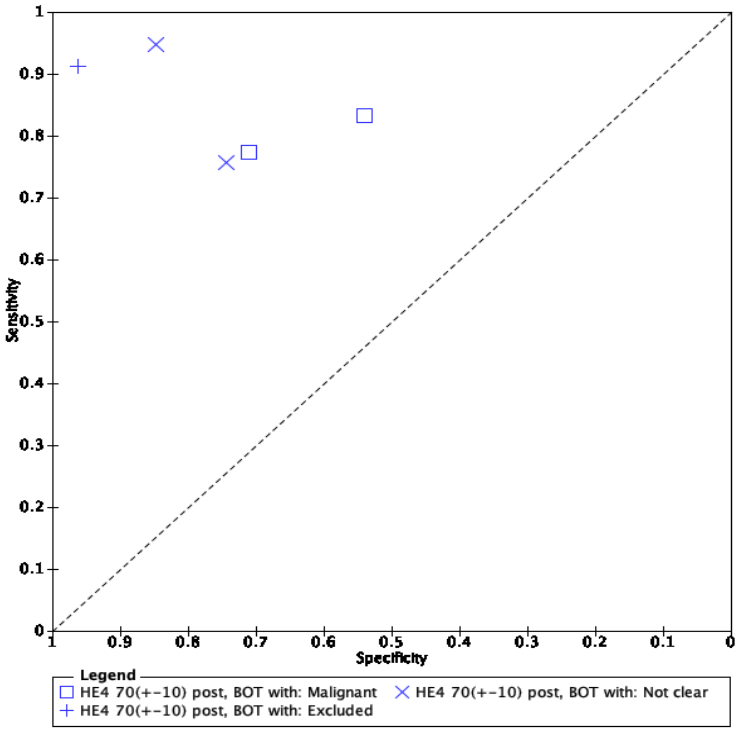


Figure 5. 9: Summary ROC Plot of 12 HE4 70(+-10) BOT for postmenopausal women.

HE4: Threshold 130pMol/L to 150pMol/L

Figure 5.10 presents individual study results for pre and postmenopausal women with HE4 at a threshold of within +/- 10pMol/L of 70pMol/L in ROC space. The corresponding forest plot is included in the Appendix 19.

For premenopausal women using HE4 thresholds between 130-150pMol/L, the average sensitivity was 55% (95% CI 43 to 67) with an average specificity of 100% (95% CI 98 to 100), based on 4 studies with 549 women, 99 with ovarian cancer (Table 5.6). The estimates of sensitivity may be inflated by grouping of borderline results; the one study with appropriate grouping of borderline results had a sensitivity of 43% compared to 60%, 63% and 65% in the other three studies where borderline results were excluded or grouping was unclear

For postmenopausal women using HE4 thresholds between 130-150pMol/L, the average sensitivity was 75% (95% CI 64 to 83) with an average specificity of 95% (95% CI 91 to 97) based on 7 studies with 1275 women, 595 with ovarian cancer (Table 5.6). Estimates of sensitivity may be inflated by grouping of borderline results, as the one study with correct grouping of borderline results had a sensitivity of 53% compared to values ranging from 78% to 89% in the other 6 studies where borderline results were excluded or grouping was unclear. For postmenopausal women the 95% CI ellipse for specificity covers roughly 40% of ROC space indicating the high heterogeneity in these studies.

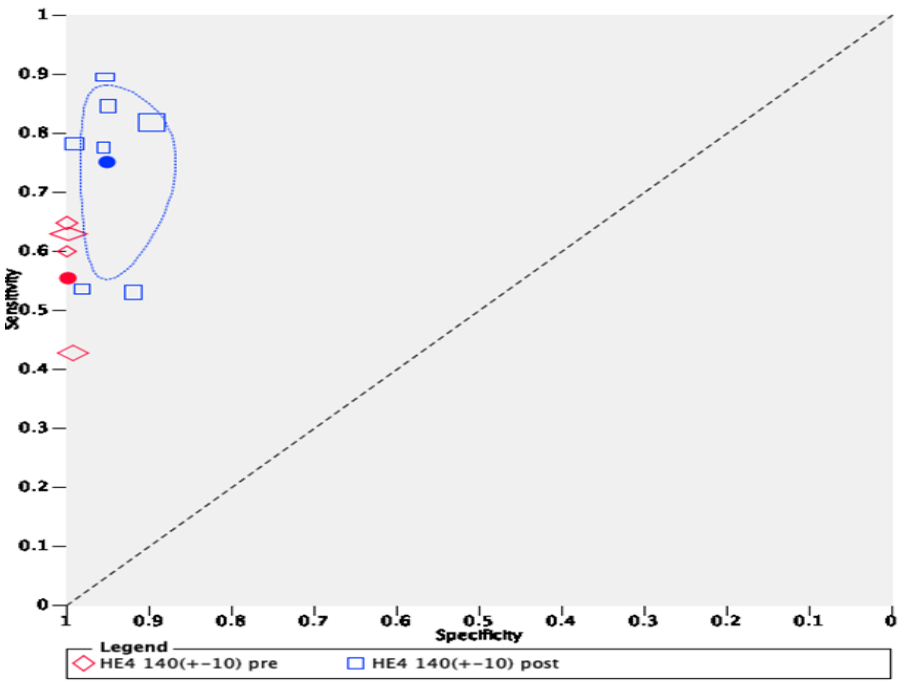


Figure 5. 10: Summary ROC Plot of tests: HE4 140pmol/L for premenopausal and postmenopausal women.

5.3.3 Accuracy of OVA1

Only 2 studies reported results for OVA1 at a threshold of 5 for premenopausal women and 4.4 for postmenopausal women. Both studies appropriately grouped borderline tumours with malignant for analysis (Table 5.8). The sensitivity for premenopausal women was 94% (95 CI 80 to 99) and specificity was 62% (95% CI 57 to 67). The sensitivity for postmenopausal women was 93% (95 CI 85 to 97) and specificity was 41 (95% CI 35 to 48).

Table 5. 8: Test performance of OVA1

Menopausal status	Test	Threshold	Number studies (N total, N OC)	Sensitivity (95% CI)	Specificity (95% CI)
Pre	OVA1	5	2	94 (80 to 99)	62 (57 to 67)
Post	OVA1	4.4	2	93 (85 to 97)	41 35 to 48)

5.4 Heterogeneity analysis

Heterogeneity analyses were conducted for menopausal status, thresholds of test positivity and appropriate BOT grouping (BOT combined with malignant tumours vs BOT combined with benign or unclear grouping for analyses).

The following planned heterogeneity analyses were not performed 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 of EOC vs Non EOC

-Reference standard QUADAS-2 domain: High/unclear risk of bias versus low risk of bias in the reference standard domain where low risk of bias is a minimum of 12 months follow-up for study participants not receiving surgery initially following a negative index test result.

5.4.1 Effect of menopausal status on test performance

The importance of using menopausal status when interpreting test results are demonstrated for HE4 and CA125 at all thresholds examined.

At a threshold of 30-40U/ml based on 15 direct within study comparisons, specificity of CA 125 was higher in postmenopausal women compared to premenopausal women by 15.5% (95% CI 9.2 to 21.8) (Table 5.5, Figure 5.5). Specificity results were more heterogeneous for premenopausal women (range 32% to 83%) than for postmenopausal women (59% to 87%) (see Forest plots in Appendix16).

At a threshold of 60-70U/ml based on 3 direct comparison studies, for postmenopausal women sensitivity was significantly higher by 18.6% (95% CI 7.4 to 29.8) and specificity was significantly higher by 4.6% (95% CI 0.3 to 8.8) than for premenopausal women's (Table 5.5, Figure 5.6). Study results were heterogeneous, particularly for sensitivity.

For the single CA125 study reporting accuracy at a threshold of 200U/ml sensitivity was statistically higher in postmenopausal women (25% higher) based on non-overlapping 95% confidence intervals, but specificity was within 1% of the estimate for premenopausal women (Table 5.5).

For HE4, diagnostic test accuracy was statistically different between pre and postmenopausal women for both sensitivity and specificity, based on both direct and indirect test comparisons (Table 5.7). At a threshold of 60-80pMol/L based on direct comparisons, for postmenopausal women sensitivity was higher by 13.2% (95% CI 2.1 to 24.3) and specificity was lower by 11.5% (95% CI 4.7 to 27.6) compared to premenopausal women (Table 5.7, Figure 5.7). At a threshold of 130 to 150 pMol/L based on direct comparisons, for postmenopausal women sensitivity was higher by 21.5% (95% CI 3.1 to 39.8) and specificity was lower by 5.7% (95% CI 2.8 to -8.7) than for premenopausal women (Table 5.7, Figure 5.10).

5.4.2 Effect of test threshold on test performance

For both HE4 and CA125 tests, statistically different test performance was found when comparing the most commonly used test thresholds: for HE4 60-70pMol/L compared to 130-150pMol/L and for CA125 30-40U/ml compared to 60–70U/ml. This was based on both direct and indirect test comparisons.

For, CA125 in both pre and postmenopausal women, both sensitivity and specificity were significantly different when thresholds between 30-40U/ml and 60-70U/ml were compared. For CA125 in premenopausal women, sensitivity was significantly higher using a threshold of 30-40U/ml by 18.1% (95% CI 4.0 to 32.2) and specificity was significantly lower by 18.5% (95% CI -11.2 to -25.9) compared to a threshold of 60-70U/ml, based on 21 indirect and direct comparison studies (Table 5.5, Figure 5.5 compared to Figure 5.6). A similar result was found based on 2 direct comparison studies (Table 5.5). For CA125 in postmenopausal women, sensitivity was significantly higher using a threshold of between 30-40U/ml by 10.3% (95% CI 4.6 to 16.0) and specificity was significantly lower by 9.3% (95% CI -6.4 to -12.2) compared to a threshold of 60–70U/ml, based on 25 indirect and direct comparison studies (Table 5.5, Figure 5.5 compared to Figure 5.6). A similar result was found based on 3 direct comparison studies (Table 5.5). For both pre and postmenopausal women the optimum threshold will depend on whether the increase in sensitivity achieved with a threshold of 30-40U/ml is worth

the decrease in specificity compared to 60-70U/ml. This in turn is dependent on the absolute magnitude of test errors determined by the prevalence of ovarian cancer in the testing setting.

At a prevalence of 0.23% typical of a primary care presentation, CA125 at a threshold of 60-70U/ml may be the preferred test as the higher sensitivity of CA125 at 30-40U/ml would on average identify no additional women with ovarian cancer, but result in 185 fewer women with a correct true negative result (185 more women with false positive result). However, at a point in the clinical pathway with a prevalence of 10%, an average additional 18 women with true positive results might be considered to outweigh the disadvantages of an average of 167 more women with false positive results. Table 5.10 presents results for four prevalence scenarios.

For HE4 in premenopausal women, at thresholds between 60-80pMol/L sensitivity was higher by 23.0% (95% CI 3.0 to 43.0) and specificity was lower by 7.1% (95%CI -3.6 to - 10.7) compared to a threshold of 130-150pMol/L (Table 5.7) based on 3 direct comparison studies. A similar result was found based on 12 indirect and direct comparison studies (Table 5.7, Figure 5.7 compared to Figure 5.10). For HE4 in postmenopausal women specificity was significantly lower (17.2% (95% CI 4.6 to 29.9) at a threshold of 60-80pMol/L compared to a threshold of 130-150pMol/L based on 10 indirect comparison studies. Sensitivity was higher but not significantly (9.4% (95% CI -2.3 to 21.1) (Table 5.7, Figure 5.7 and Figure 5.10). A similar result was found based on 2 direct comparison studies (Table 5.7). In postmenopausal women HE4 at a threshold of 130-150pMol/L is therefore preferred as it offers a higher specificity (no significant difference in sensitivity). In premenopausal women the evidence is less clear as an increase in sensitivity at a threshold of 60-70pMol/L has to be balanced against a decrease in specificity compared to a threshold of 130-150pMol/L. The magnitude of test errors determined by the prevalence of ovarian cancer in the testing setting will determine whether the increase in sensitivity is worth the trade off of a decrease in specificity.

Table 5. 9: Summary of findings: HE4, CA125 and OVA1 at common thresholds.

Test	Test threshold	Menopausal status	Dir/ All/ BOT	N Studies (N Participants,DP)	Sensitivity (95%	Specificity(95% CI)		Prevalence 0.23%	Prevalence 4%	Prevalence 10%	Prevalence 30%
HE4	70ish	Pre	All	11 (1718, 297)	74 (68 to 79)	91 (87 to 94)	True postives	2 (2 to 2)	30 (27 to 32)	74 (68 to 79)	222 (204 to 237)
							False negatives	1 (0 to 1)	10 (8 to 13)	26 (21 to 32)	78 (63 to 96)
							False positives	90 (60 to 130)	86 (58 to 125)	81 (54 to 117)	63 (42 to 91)
HE4	70ish	post	All	6 (601, 328)	84 (76 to 89)	78 (63 to 88)	True positives	2 (2 to 2)	34 (30 to 36)	84 (76 to 89)	252 (228 to 267)
							False negatives	0 (0 to 1)	6 (4 to 10)	16 (11 to 24)	48 (33 to 72)
							False positives	219 (120 to 369)	211 (115 to 355)	198 (108 to 333)	154 (84 to 259)
							True negatives	778 (629 to 878)	749 (605 to 845)	702 (567 to 792)	546 (441 to 616)
HE4	140ish	Pre	All	4 (549, 99)	55 (43 to 67)	100 (98 to 100)	True positives	1 (1 to 2)	22 (17 to 27)	55 (43 to 67)	165 (129 to 201)
							False negatives	1 (1 to 1)	18 (13 to 23)	45 (33 to 57)	135 (99 to 171)
							False positives	0 (0 to 20)	0 (0 to 19)	0 (0 to 18)	0 (0 to 14)
							True negatives	998 (978 to 998)	960 (941 to 960)	900 (882 to 900)	700 (686 to 700)
HE4	140ish	post	all	7 (1275, 595)	75 (64 to 83)	95 (91 to 97)	True positives	2 (1 to 2)	30 (26 to 33)	75 (64 to 83)	225 (192 to 249)
							False negatives	1 (0 to 1)	10 (7 to 14)	25 (17 to 36)	75 (51 to 108)
							False positives	50 (30 to 90)	48 (29 to 86)	45 (27 to 81)	35 (21 to 63)
							True negatives	948 (908 to 968)	912 (874 to 931)	855 (819 to 873)	665 (637 to 679)
CA125	35ish	Pre	BOT	7 (,)	71 (58 to 82)	70 (62 to 77)	True positives	2 (1 to 2)	28 (23 to 33)	71 (58 to 82)	213 (174 to 246)
							False negatives	1 (0 to 1)	12 (7 to 17)	29 (18 to 42)	87 (54 to 126)
							False positives	299 (229 to 379)	288 (221 to 365)	270 (207 to 342)	210 (161 to 266)
							True negatives	698 (619 to 768)	672 (595 to 739)	630 (558 to 693)	490 (434 to 539)
CA125	35ish	post	BOT	9 (,)	78 (74 to 82)	82 (80 to 84)	True positives	2 (2 to 2)	31 (30 to 33)	78 (74 to 82)	234 (222 to 246)
							False negatives	1 (0 to 1)	9 (7 to 10)	22 (18 to 26)	66 (54 to 78)
							False positives	180 (160 to 200)	173 (154 to 192)	162 (144 to 180)	126 (112 to 140)
							True negatives	818 (798 to 838)	787 (768 to 806)	738 (720 to 756)	574 (560 to 588)
CA125	65ish	Pre	All	4 (763, 165)	60 (49 to 71)	85 (79 to 89)	True positives	1 (1 to 2)	24 (20 to 28)	60 (49 to 71)	180 (147 to 213)
							False negatives	1 (1 to 1)	16 (12 to 20)	40 (29 to 51)	120 (87 to 153)
							False positives	150 (110 to 210)	144 (106 to 202)	135 (99 to 189)	105 (77 to 147)

Table 5.9 Continued...

CA125	65ish	Post	all	4 (876, 368)	73 (68 to 77)	92 (89 to 94)	True	positives	2	(2	to	2)	29	(27	to	31)	73	(68	to	77)	219	(204	to	231)
							False	negatives	1	(1	to	1)	11	(9	to	13)	27	(23	to	32)	81	(69	to	96)
							False	positives	80	(60	to	110)	77	(58	to	106)	72	(54	to	99)	56	(42	to	77)
							True negatives		918 (888 to 938)				883 (854 to 902)				828 (801 to 846)				644 (623 to 658)			
CA125	200	pre	All	1 (445, 86)	29 (20 to 40)	98 (96 to 99)	True	positives	1	(0	to	1)	12	(8	to	16)	29	(20	to	40)	87	(60	to	120)
							False	negatives	2	(1	to	2)	28	(24	to	32)	71	(60	to	80)	213	(180	to	240)
							False	positives	20	(10	to	40)	19	(10	to	38)	18	(9	to	36)	14	(7	to	28)
							True negatives		978 (958 to 988)				941 (922 to 950)				882 (864 to 891)				686 (672 to 693)			
1212	200	Post	All	1 (364, 156)	54 (46 to 62)	97 (94 to 99)	True	positives	1	(1	to	1)	22	(18	to	25)	54	(46	to	62)	162	(138	to	186)
							False	negatives	1	(1	to	1)	18	(15	to	22)	46	(38	to	54)	138	(114	to	162)
							False	positives	30	(10	to	60)	29	(10	to	58)	27	(9	to	54)	21	(7	to	42)
							True negatives		968 (938 to 988)				931 (902 to 950)				873 (846 to 891)				679 (658 to 693)			
OVA1	5	pre	All	2 (,)	94 (80 to 99)	62 (57 to 67)	True	positives	2	(2	to	2)	38	(32	to	40)	94	(80	to	99)	282	(240	to	297)
							False	negatives	0	(0	to	0)	2	(0	to	8)	6	(1	to	20)	18	(3	to	60)
							False	positives	379	(329	to	429)	365	(317	to	413)	342	(297	to	387)	266	(231	to	301)
							True negatives		619 (569 to 668)				595 (547 to 643)				558 (513 to 603)				434 (399 to 469)			
OVA1	4.4	Post	All	2 (,)	93 (85 to 97)	41 (35 to 48)	True	positives	2	(2	to	2)	37	(34	to	39)	93	(85	to	97)	279	(255	to	291)
							False	negatives	0	(0	to	0)	3	(1	to	6)	7	(3	to	15)	21	(9	to	45)
							False	positives	589	(519	to	649)	566	(499	to	624)	531	(468	to	585)	413	(364	to	455)
							True negatives		409 (349 to 479)				394 (336 to 461)				369 (315 to 432)				287 (245 to 336)			

5.4.3 Effect of borderline test results on test performance

For CA125 between thresholds of 30 U/ml to 40 U/ml in premenopausal women, there are 17 studies, 7 of which grouped borderline results appropriately for analysis (Figure 5.11) and for postmenopausal women 21 studies, 9 of which grouped borderline results appropriately for (Figure 5.12). From visual inspection of (Figure 5.11 compared to Figure 5.12) and a sensitivity analysis restricted to studies grouping borderline tumours appropriately, studies grouping borderline tumours appropriately have a lower sensitivity but this was statistically significant only for postmenopausal women. Formal heterogeneity tests of these two subgroups found a statistically significant lower average sensitivity in the low ROB group (-10.2%; 95% CI -15.3 to -5.1 (Appendix 15), but there was no statistical difference in specificities.

Investigation of the effects of borderline tumour categorisation could not be carried out for CA125 at thresholds between 60 and 70 or for HE4 due to lack of studies see Figure 5.8 and Figure 5.9.

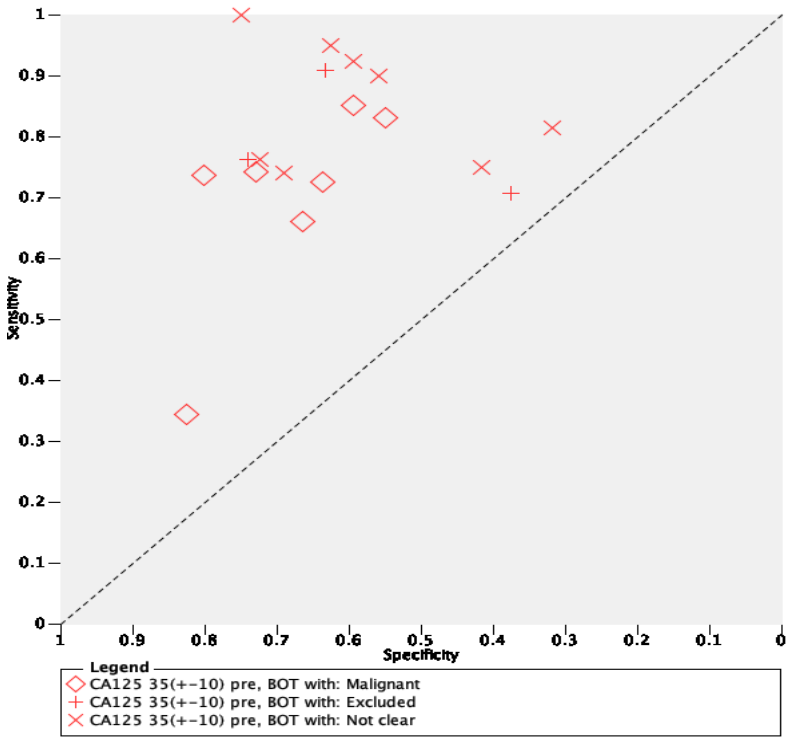


Figure 5. 11: Summary ROC Plot of 29 CA125 35(+-10) BOT for premenopausal women.

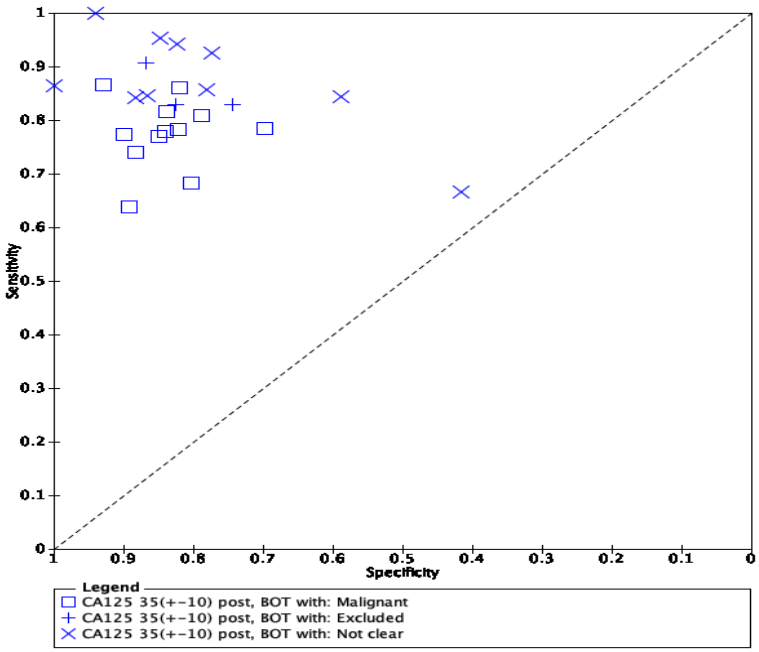


Figure 5. 12: Summary ROC Plot of CA125 35(+-10) BOT for postmenopausal women.

5.4.4 Other planned sensitivity analysis

Planned sensitivity analysis including only studies with low concern about applicability in the patient selection domain of QUADAS-2 compared to studies with high or unclear applicability concern was not conducted, as all studies had a rating of unclear applicability in this domain for both HE4 and CA125.

5.5 CA125 and HE4: comparison of test performance

For the comparison of CA125 and HE4 at different thresholds results from direct comparisons and those studies using an appropriate grouping of borderline tumours were preferred. Comparisons are based on all available studies where this was not available.

For interpretation of comparisons, one can conclude that there is a clear advantage of one test over another if one of the following is observed:

(i) both the difference in sensitivity and specificity are statistically higher for one test, so the test with the higher sensitivity and specificity would be preferred

(ii) the difference in sensitivity is not statistically different between tests but specificity is statistically higher for one test, so the test with the higher specificity is preferred

(iii) difference in specificity is not statistically different between tests but sensitivity is statistically higher for one test, so the test with the higher sensitivity is preferred.

Advantage of one test over another is not supported by the evidence if the following is observed:

(iv)There is no statistically significant difference between the sensitivity and specificity of the two tests

If there is a statistically significant difference in both the sensitivity and specificity between the two tests, with an increase in sensitivity combined with a decrease in specificity, or vice versa then interpretation of the advantages and disadvantages of tests depend on the consequences of false positive and false negative test results and the prevalence of disease. Prevalence of disease determines the absolute number of true positive and false positive test results.

Table 5.10 demonstrates differences in test performance between HE4 and CA125 at different thresholds. Table 5.11 provides a summary of those tests and thresholds where a test comparison results in opposite directions for the difference in sensitivity and specificity and illustrates the average difference in the absolute numbers of women who would have a true positive and true negative test result between the two tests, at four hypothetical prevalences of ovarian cancer representing different points in the clinical pathway. Prevalence increases along a clinical pathway as prior tests filter out causes of false positive test results.

Table 5. 10: Test performance: Comparison HE4 and CA125.

Menopausal status	Test 1 (Threshold)	Test 2 (Threshold)	Number studies (N total women, N OC)[a]	Difference in Sensitivity (95% CI)	Difference in Specificity (95% CI)	Notes
Pre **(CA125- HE4)**	HE4 70ish	CA125 35ish	18 (17 CA125, 11 HE4, 10 of both) (3621, 522 OC)	4 (-4.8 to 13.0)	-25.2 (-31.6 to -18.8)	*note Radosa 2011 outlier not omitted as otherwise single test accuracy varies from these estimates
Pre	HE4 70ish	CA125 35ish	10 each (1139, 248)	9.9 (1.0 to 18.7)	-26.8 (-33.4 to -20.1)	Direct only
Pre	HE4 70ish	CA125 35ish	6	12.8 (0.3 to 25.3)	-23.7 (-37.5 to -10.0)	Direct with BOT=1 only
Post	HE4 70ish	CA125 35ish	22 (21 CA125, 4 HE4, 4 of both) (3442,1372)	-2.2 (-8.1 to 3,8)	4.4 (-2.8 to 11.6)	Univariate HE4 same estimates, slightly different 95% CI to single CA125 - 2% less sens estimate, spec same
Post	HE4 70ish	CA125 35ish	4 each (462, 215)	-1.5 (-11.2 to 8.2)	12.2 (0.28 to 24.0)	Direct only
Post	HE4 70ish	CA125 35ish	2	-3.3 (-16.8 to 10.2)	20.5 (7.3 to 33.7)	Direct with BOT=1 only
Pre	HE4 70ish	CA125 65ish	15 (2481, 462)	-14.0 (-27.7 to 0.3)	-6.8 (-13.0 to -0.3)	All CA125 v.similar-1% differnce on 95% CI HE4 – same

Table 5.10 Continued...

Pre	HE4 70ish	CA125 65ish	0	NA	NA	Direct only
Pre	HE4 70ish	CA125 65ish	5	-11.4 (-25.8 to 3.1)	-3.6 (-8.1 to 9.1)	All BOT1 studies
Post	HE4 70ish	CA125 65ish	9 (1477, 696)	-11.3 (-19.4 to -3.3)	14.1 (1.6 to 26.7)	All CA125 same as other estimates HE4 only 1% difference in 95% CI
Post	HE4 70ish	CA125 65ish	0	NA	NA	Direct only
Post	HE4 70ish	CA125 65ish	5	-5.2 (-13.1 to 2.7)	27.1 (15.3 to 39.0)	All BOT1 studies[SM1]
Pre G3	HE4 140ish	CA125 35ish	17 (3042, 473)	23.0 (9.4 to 36.7)	-33.7 (-39.5 to -27.8)	All CA125 similar,2% diff in 95% CI HE4 same
Pre	HE4 140ish	CA125 35ish	4 (549, 99)	25.5 (9.1 to 41.9)	-30.6 (-39.4 to -21.7)	Direct Estimates slightly different from all study estimates - 3%ish different
Pre	HE4 140ish	CA125 35ish	8 (only 1 HE4 study)	28.1 (9.3 to 47.0)	-28.7 (-36.3 to -21.2)	BOT1 studies
Post	HE4 140ish	CA125 35ish	22 (3785, 1462)	12.6 (4.7 to 20.5)	-11.0 (-13.8 to -8.2)	All
Post	HE4 140ish	CA125 35ish	6 (793, 392)	14.9 (2.9 to 27.0)	-12.4 (-16.7 to -8.1)	Direct

Table 5.10 continued...

Post	HE4 140ish	CA125 35ish	10 (only 1 HE4 study)	25.0 (15.2 to 34.8)	-9.9 (-16.1 to -3.7)	BOT1 studies
Pre	HE4 140ish	CA125 65ish	8 (1312, 264)	3.7 (-13.9 to 21.4)	-15.2 (-20.3 to -10.0)	All
Pre	HE4 140ish	CA125 65ish	0	NA	NA	Direct
Pre	HE4 140ish	CA125 65ish	3 (2 CA125, 1 HE4)	12.7 (-5.7 to 31.1)	-12.1 (-15.5 to -8.7)	BOT1 studies
Post	HE4 140ish	CA125 65ish	11 (2151, 963)	-1.4 (-8.1 to 5.4|)	-1.3 (-4.3 to 1.7)	All
Post	HE4 140ish	CA125 65ish	0	NA	NA	Direct
Post	HE4 140ish	CA125 65ish	4 (1 HE4 205, 119; 3 CA125 842, 355)	20.6 (10.5 to 30.7)	0.1 (-6.1 to 6.4)	BOT1 studies

5.5.1 Test performance in premenopausal women

In premenopausal women comparison of test performance was possible between CA125 and HE4. There were no direct comparisons of OVA1 with CA125 or HE4. A statistical comparison of OVA1 with CA125 or HE4 using meta-regression was not possible with only 2 OVA1 studies.

A) One comparison demonstrates a clear advantage of one test over another: When HE4 at a threshold of 130-150pMol/L is compared to CA125 at a threshold of 60-70U/ml there is no significant difference in sensitivity but CA125 has a statistically significant lower specificity (-12.1, 95% CI -15.5 to -8.7). HE4 at a threshold of 130-150pMol/L is preferred over CA125 at a threshold of 60-70U/ml.

B) One comparison demonstrates no significant difference between tests: When CA125 at a threshold of 60-70U/ml is compared to HE4 at a threshold of 60-80pmol/L no significant difference in sensitivity or specificity is observed.

C) For three comparisons there is a statistically significant difference in both the sensitivity and specificity between the two tests, with an increase in sensitivity combined with a decrease in specificity, or vice versa. The choice between tests will therefore depend on a trade off in the absolute numbers of test errors which in turn will be determined by the prevalence of ovarian cancer in the tested population:

1) When CA125 at a threshold of 30-40 U/ml is compared to CA125 at a threshold of 60-70U/ml. CA125 30-40U/ml has a sensitivity of 71% (95% CI 58 to 82) which is higher than the sensitivity of CA125 60-70U/ml by 18.1% (95% CI 4.0 to 32.2). However, CA125 30-40U/ml has a specificity of 70% (95% CI 62 to 77) which is lower than CA125 60–70U/ml by 18.5% (95% CI -25.9 to -11.2). At a prevalence of 0.23% that is closer to represent a primary care population CA125 at a threshold of 60-70U/ml would be the preferred test. The higher sensitivity of CA125 at 30-40U/ml would on average identify no additional women with ovarian cancer, but result in 185 more women with a false positive result. However, at a point in the clinical pathway with a prevalence of 10%, an average additional 18 women with true positive results would be identified by CA125 at 30-40U/ml which might be considered to outweigh the disadvantages of an average of 167 more women with false positive results compared to CA125 60–70U/ml.

2) When CA125 at a threshold of 30-40U/ml is compared to HE4 at a threshold of 60-80p/Mol/L CA125 30-40U/ml has a sensitivity which is significantly higher, 12.8% more (95% CI 0.3 to 25.3) than HE4 at a threshold of 60-80p/Mol/L but a specificity of 70% (95% CI 62 to 77) which is 23.7% (95% CI -37.5 to -10.0) lower than HE4 at a threshold of 60-80pMol/L. At a prevalence of 0.23% typical of a primary care population HE4 at a threshold of 60-80pMol/L would be the preferred test. HE4 60-80pMol/L has a sensitivity of 74% (95% CI 68 to 79) and a specificity of 91% (95% CI 87 to 94). The higher sensitivity of CA125 at a threshold of 30-40U/ml would on average identify no additional women with ovarian cancer but 236 more women would receive a false positive result. In contrast at a point in the clinical pathway with a prevalence of 10%, an average additional 13 women with ovarian cancer who would be identified (true positive results) might be considered to outweigh the disadvantages of an average of 213 more women with false positive results. Table 5.11 presents results for four prevalence scenarios.

3) When CA125 at a threshold of 30-40U/ml is compared to HE4 at a threshold of 130-150pMol/L, CA125 30- 40U/ml has a sensitivity (95% CI 58 to 82) which is 25.5% (95% CI 9.1 to 41.9) higher than HE4 130-150pMol/L but a specificity of 71% (95% CI 58 to 82) which is 30.6% (95% CI -21.7 to -39.4) lower than HE4 130-150pMol/L. At a prevalence of 0.23% typical of a primary care presentation HE4 at a threshold of 130-150pMol/L would be the preferred test. On average the higher sensitivity of CA125 at 30-40U/ml would identify one additional woman with ovarian cancer, but 308 more women would receive a false positive result. In contrast at a point in the clinical pathway with a prevalence of 10%, an average additional 26 women with ovarian cancer who would be identified (true positive results) might be considered to outweigh the disadvantages of an average of 278 more women with false positive results.

In premenopausal women, based on consideration of A), B) and C) above and Table 5.9, HE4 at a threshold of 60-80pMol/L is the preferred test in settings with prevalence typical of primary care populations. HE4 at a threshold of 130-150pMol/L is the preferred test in test settings with prevalence typical of specialist settings.

5.5.2 Test performance in postmenopausal women

In postmenopausal women comparison of test performance was possible between CA125 and HE4. There were no direct comparisons of OVA1 with CA125 or HE4. A statistical comparison of OVA1 with CA125 or HE4 using meta-regression was not possible with only 2 OVA1 studies.

A) Two comparisons demonstrate clear advantage of one test over another:

1) CA125 30-40U/ml has a higher specificity than HE4 60-80pMol/L.

When CA125 at a threshold of 30-40U/ml is compared to HE4 at a threshold of 60-80pMol/L there is no statistically significant difference in sensitivity. CA125 30-40U/ml has a specificity of 82% (95% CI 80 to 84) based on direct comparison studies which is statistically significantly 12.2% (95% CI 0.28 to 24.0) higher compared to HE4 60-80pMol/L.

2) CA125 at 60-70U/ml has a higher specificity than HE4 60-80pMol/L.

When CA125 at a threshold of 60-70U/ml is compared to HE4 at a threshold of 60-80pMol/L evidence for any difference in sensitivity is inconsistent. However, CA125 60-70U/ml has a specificity of 92% (95% CI 89 to 94 based on indirect comparison studies which is 14.1% (95 % CI 1.6 to 26.7) higher than HE4 60-80pMol/L.

B) One comparison demonstrates weak evidence for one test being preferred over another.

CA125 60-70U/ml may have a higher sensitivity compared to HE4 130-150pMol/L.

When CA125 at a threshold of 60-70U/ml is compared to HE4 at a threshold of 130-150pMol/L, there is no evidence for a difference in specificity but based on small number of high quality CA125 has a sensitivity of 73% (95% CI 68 to 77) which is 20.6% (95% CI 10.5 to 30.7) higher compared to HE4 130-150pMol/L.

C) For one comparison there is a statistically significant higher sensitivity for one test at a particular threshold but the specificity is lower and therefore the choice between tests depends on the trade off in the absolute numbers of test errors which in turn will be determined by the prevalence of ovarian cancer in the tested population.

CA125 at a threshold of 30-40U/ml compared to HE4 at a threshold of 130-150pMol/L sensitivity is 15% (95% CI 3 to 27.0) higher but a specificity 12% (95% CI -8 to -17) lower.

At a prevalence of 0.23% typical of a primary care presentation the higher sensitivity of CA125 at a threshold of 30-40U/ml would not result in any additional ovarian cancers detected. However, the lower specificity of CA125 at a threshold of 30-40U/ml would result in 124 (81-167) more postmenopausal women with false positive test results.

At a higher prevalence typical of specialist settings, the higher sensitivity of CA125 at a threshold of 30-40U/ml would identify an additional 6 (1 - 11) women with ovarian cancer. However, an additional 119 (78 - 160) would have a false positive result.

In postmenopausal women, based on consideration of A), B) and C) above and Table 5.11, HE4 at a threshold of 130-150pMol/L is the preferred test in testing settings with prevalence typical of primary care populations. CA125 at a threshold of 30-40U/ml is the preferred test in test settings with prevalence typical of specialist settings.

Table 5. 11: Summary of findings: Difference between HE4 and CA125 at common thresholds

						Prevalence 0.23%		Prevalence 3%		Prevalence 10%		Prevalence 30%	
Test 1	Test 2	Menopausal status	N Studies (N, DP)	Difference sensitivity (95% CI)	Difference specificity (95% CI)[SM1]	Mean diff in TP (95% CI)	Mean diff in TN (95% CI)	Mean diff in TP (95% CI)	Mean diff in TN (95% CI)	Mean diff in TP (95% CI)	Mean diff in TN (95% CI)	Mean diff in TP (95% CI)	Mean diff in TN (95% CI)
HE4 70ish	HE4 140ish	Pre	3 (296, 72)	23 (3 to 43)	-7 (-4 to -11)	1 (0 to 1) fewer TP with HE4 at 140ish	71 (36 to 107) more TN with HE4 at 140ish	7 (1 to 13) fewer TP with HE4 at 140ish	69 (35 to 104) more TN with HE4 at 140ish	23 (3 to 43) fewer TP with HE4 at 140ish	64 (32 to 96) more TN with HE4 at 140ish	69 (9 to 129) fewer TP with HE4 at 140ish	50 (25 to 75) more TN with HE4 at 140ish
HE4 70ish	CA125 35ish	Pre	6 (0, 0)	13 (0 to 25)	-24 (-10 to -38)	0 (0 to 1) more TP with CA125 at 35ish	236 (100 to 374) fewer TN with CA125 at 35ish	4 (0 to 8) more TP with CA125 at 35ish	230 (97 to 364) fewer TN with CA125 at 35ish	13 (0 to 25) more TP with CA125 at 35ish	213 (90 to 338) fewer TN with CA125 at 35ish	38 (1 to 76) more TP with CA125 at 35ish	166 (70 to 263) fewer TN with CA125 at 35ish
HE4 140ish	CA125 35ish	Pre	4 (549, 99)	26 (9 to 42)	-31 (-22 to -39)	1 (0 to 1) more TP with CA125 at 35ish	308 (217 to 393) fewer TN with CA125 at 35ish	8 (3 to 13) more TP with CA125 at 35ish	300 (210 to 382) fewer TN with CA125 at 35ish	26 (9 to 42) more TP with CA125 at 35ish	278 (195 to 355) fewer TN with CA125 at 35ish	77 (27 to 126) more TP with CA125 at 35ish	216 (152 to 276) fewer TN with CA125 at 35ish
CA125 35ish	CA125 65ish	Pre	22 (0, 0)	20 (7 to 32)	-19 (-12 to -26)	0 (0 to 1) fewer TP with CA125 at 65ish	190 (-124 to -255) more TN with CA125 at 65ish	6 (2 to 10) fewer TP with CA125 at 65ish	184 (120 to 248) more TN with CA125 at 65ish	20 (7 to 32) fewer TP with CA125 at 65ish	171 (112 to 230) more TN with CA125 at 65ish	59 (21 to 97) fewer TP with CA125 at 65ish	133 (87 to 179) more TN with CA125 at 65ish
HE4 70ish	HE4 140ish	Post	10 (0, 0)	9 (-2 to 21)	-17 (-5 to -30)	0 (0 to 0) fewer TP with HE4 at 140ish	172 (46 to 298) more TN with HE4 at 140ish	3 (1 to 6) fewer TP with HE4 at 140ish	167 (45 to 290) more TN with HE4 at 140ish	9 (2 to 21) fewer TP with HE4 at 140ish	155 (41 to 269) more TN with HE4 at 140ish	28 (7 to 63) fewer TP with HE4 at 140ish	120 (32 to 209) more TN with HE4 at 140ish

HE4 70ish	CA125 35ish	Post	4 (462, 215)	-2 (-11 to 8)	12 (24 to 0)	0 (0 to 0) more TP with CA125 at 35ish	122 (239 to 3) fewer TN with CA125 at 35ish	0 (3 to 2) more TP with CA125 at 35ish	118 (233 to 3) fewer TN with CA125 at 35ish	2 (-11 to 8) more TP with CA125 at 35ish	110 (216 to 3) fewer TN with CA125 at 35ish	5 (34 to 25) more TP with CA125 at 35ish	85 (168 to 2) fewer TN with CA125 at 35ish
HE4 70ish	CA125 65ish	Post	10 (1477, 696)	-10 (-18 to 2)	14 (26 to 1)	0 (0 to 0) fewer TP with CA125 at 65ish	135 (259 to 9) more TN with CA125 at 65ish	3 (5 to 1) fewer TP with CA125 at 65ish	131 (250 to 9) more TN with CA125 at 65ish	10 (18 to 2) fewer TP with CA125 at 65ish	122 (234 to 8) more TN with CA125 at 65ish	30 (53 to 7) fewer TP with CA125 at 65ish	95 (182 to 6) more TN with CA125 at 65ish
HE4 140ish	CA125 35ish	Post	6 (793, 392)	15 (3 to 27)	-12 (-8 to -17)	0 (0 to 1) more TP with CA125 at 35ish	124 (81 to 167) fewer TN with CA125 at 35ish	4 (1 to 8) more TP with CA125 at 35ish	120 (79 to 162) fewer TN with CA125 at 35ish	15 (3 to 27) more TP with CA125 at 35ish	112 (73 to 150) fewer TN with CA125 at 35ish	45 (9 to 81) more TP with CA125 at 35ish	87 (57 to 117) fewer TN with CA125 at 35ish
CA125 65ish	CA125 35ish	Post	3 (842, 355)	8 (2 to 14)	-11 (-7 to -16)	0 (0 to 0) more TP with CA125 at 35ish	113 (70 to 155) fewer TN with CA125 at 35ish	2 (1 to 4) more TP with CA125 at 35ish	110 (68 to 150) fewer TN with CA125 at 35ish	8 (2 to 14) more TP with CA125 at 35ish	102 (63 to 140) fewer TN with CA125 at 35ish	24 (5 to 42) more TP with CA125 at 35ish	79 (49 to 109) fewer TN with CA125 at 35ish

5.6 Summary of results

The robustness and applicability of the results are limited by the heterogeneity between studies especially the lack of representation of the entire spectrum of disease by including only EOCs or excluding BOTs in some included studies and these are identified and downgraded in quality assessment. Bearing these caveats in mind the following recommendations are suggested.

5.6.1 Thresholds CA125

For CA125, both sensitivity (higher) and specificity (lower) were significantly different when threshold of 30-40U/ml was compared to threshold cut off 60-70U/ml in both pre and postmenopausal women. Prevalence will therefore determine whether increase in sensitivity is worth decrease in specificity.

5.6.2 Thresholds HE4

HE4 at a threshold of around 130-150pMol/L is preferred in postmenopausal women because of increased specificity. In premenopausal women, however, there is again a trade off between sensitivity and specificity increased sensitivity at a threshold of 60-70pMol/L has to be balanced against a decrease in specificity when compared to a threshold of 130-150pMol/L with prevalence being the determining factor.

5.6.3 Test comparisons

In premenopausal women, HE4 at a threshold of 60-80pMol/L is the preferred test in settings with prevalence typical of primary care populations. HE4 at a threshold of 130-150pMol/L is the preferred test in test settings with prevalence typical of specialist settings.

In postmenopausal women, HE4 at a threshold of 130-150pMol/L is the preferred test in settings with prevalence typical of primary care populations. CA125 at a threshold of 30-40U/ml is the preferred test in test settings with prevalence typical of specialist settings.

CHAPTER 6: DISCUSSION OF BIOMARKERS AND COMBINATION REVIEW

I present below, an overarching discussion of results from both the biomarkers and the combination test systematic reviews. In this section, I aim to discuss the common methodological challenges, pitfalls, data gaps across both reviews. I will also summarise interpretation of results, applicability across different prevalence rates of the target condition, OC, clinical and research implications and discuss recommendations.

6.1 Biomarkers review

6.1.1 Comparison of accuracy of CA125, HE4 and OVA1

Women with symptoms suspicious of ovarian cancer will present to primary care ideally or to emergency service for investigation and management. When the women present to the primary care in UK and have symptoms suspicious of OC, Nice guidelines recommend CA125 and referral for investigation to secondary care or tertiary care.

Biomarkers are unlikely to be used on their own independently in secondary or tertiary care for making management decisions. Their most likely use is perhaps as a triage test in low prevalence in primary care to guide referral. However, the real risk of missing cancers has to be balanced against the risk of causing anxiety, effect on the quality of life for the family, the pressure on resources and patient wishes. Biomarkers are a more objective test in comparison to ultrasound or combination testing with ultrasound as a component. In addition to the ease of use, cost and the lack of the need for special training for implementation make them attractive as stand-alone tests in a low prevalence testing i.e. primary care for triaging women with an index of suspicion of OC to further testing and management. Hence, a decision based on diagnostic test performance in low prevalence settings seems most of value in assessing the utility of biomarkers.

For biomarkers, I compared the accuracy of CA125, HE4 and OVA1. OVA1 results could not be used in a meta-analysis due to limited number of available studies as this is a relatively new

biomarker. Therefore, I compared the accuracy of CA125 and HE4 at commonly used clinical thresholds for test positivity for biomarkers; in pre and postmenopausal women, considering appropriate BOT categorisation (BOT with malignant compared to BOT combined with benign or unclear categorisation for analyses) using either direct comparison of tests (both tests done in in the same patient cohort) or indirect comparisons (tests done in different patient cohort). The results are also presented in the context of prevalence and the effect of threshold and accuracy on test performance based on hypothetical prevalence at different points in the clinical pathway as the patient progresses through the triage at different clinical settings - primary care, emergency services, secondary and tertiary.

The comparison of HE4 (threshold cut off of 60-80pMol/L and 130-150pMol/L) against CA125 (30-40U/ml and 60-70U/ml) for detection of OC shows the test performance is significantly different for pre and postmenopausal women both in direct and indirect comparison studies.

Studies that report an appropriate grouping of results of borderline tumours with malignant tumours were few but showed statistically significant differences in premenopausal women for CA125 cut off of 30-40U/ml to studies where BOT categorisation was unclear or excluded. This may be partially influenced by the included studies with unclear BOT categorisation that may have included some or all studies that have grouped BOTs appropriately but unclear due to unclear reporting.

When the thresholds are compared for pre and postmenopausal women, CA125 as expected display an increase in specificity and drop in sensitivity with increasing threshold Table (see Table 5.10). When the sensitivity and specificity of the test at different test positivity threshold cut offs were compared in hypothetically constructed 4 different prevalence settings to represent the spectrum in clinical care setting the expected trade offs were seen and therefore the magnitude of test errors (prevalence) will determine whether increase in sensitivity is worth the trade off with decrease in specificity.

HE4 at a test positivity threshold of 130-150pMol/L is recommended in postmenopausal women as test positivity threshold cut off of 60-80pMol/L showed a significantly lower specificity by 17.2% (95% CI 4.6 to 29.9) with similar sensitivity. However, in premenopausal women, the evidence is less clear as it again shows trade offs between sensitivity and specificity at different threshold test positivity cut offs; an increase in sensitivity at a threshold of 60-

80pMol/L has to be balanced against a decrease in specificity compared to a threshold of 130-150pMol/L to determine if the increase in sensitivity is worth the drop in specificity with prevalence being the arbitrary factor.

When CA125 and HE4 are compared at different thresholds, there were tests that showed i) a clear advantage of one test over the other (HE4 at a threshold of 130-150pMol/L is compared to CA125 at a threshold of 60-70U/ml) ii) no significant difference between tests (CA125 at a threshold of 60-70U/ml is compared to HE4 at a threshold of 60-80pmol/L) iii) there is a statistically significant difference in both the sensitivity and specificity between the two tests, with an increase in sensitivity combined with a decrease in specificity, or vice versa (all the other comparisons except the two mentioned above). Based on these considerations, 1) In premenopausal women HE4 at a threshold of 60-80pMol/L and 130-150pMol/L is the preferred test in settings with prevalence that resembles primary care populations and specialist settings respectively 2) In postmenopausal women, HE4 at a threshold of 130-150pMol/L is the preferred test in testing settings with prevalence typical of primary care populations and CA125 at a threshold of 30-40U/ml is the preferred test in test settings with prevalence typical of specialist settings. (Table 5.11).

Studies that report an appropriate grouping of borderline results with malignant were few a trend was observed (in CA 125 biomarker studies) for higher average estimates of sensitivity from studies that either excluded borderline tumours from analysis or where the classification of borderline tumours for analysis was unclear compared to studies where borderline tumours were classified as positive (grouped with malignant). This difference was significant for post-menopausal women.

6.1.2 Applicability of findings to the review question

The above results have to be interpreted carefully as comparisons are based on a small number of comparison studies. In addition, all the tests were conducted in high prevalence settings and the mathematical construct of test behaviour in low prevalence settings may not hold up in real life testing. The comparison studies at different threshold cut offs do not show any particular test where the accuracy outperforms the other to make a hard recommendation. The comparisons show the expected trade-off between sensitivity and specificity and in the context of prevalence may aid the clinicians and patients to guide the decisions within the context.

Table 5.11 presents results for four prevalence scenarios. Magnitude of test errors (prevalence) and or the significance and implications of false positives or false negatives or will determine whether increase in sensitivity is worth decrease in specificity.

Therefore, bearing the limitations discussed, I believe, HE4 is recommended at threshold test positive cut off of 60-80pMol/L for premenopausal women and 130-150pMol/L for postmenopausal women in low prevalence settings.

6.2 Combination review

6.2.1 Comparison of the accuracy of RMI1, ROMA, LR2, ACOG v3 (CA125) and ACOG v3 (OVA1)

In the UK, women with a suspected adnexal mass and with either an abnormal CA125 or US examination are referred for investigation to secondary care where an RMI1 is performed. Therefore, the performance of ROMA, LR2, ACOG v3 (CA125) and ACOG v3 (OVA1) is investigated relative to RMI1. A statistical comparison between ADNEX and other tests was not possible because only one ADNEX study was included in this review.

Due to variation in test positivity thresholds, a comparison of accuracy with reference to a fixed specificity of 90% was undertaken; this was chosen in keeping with a clinically acceptable false positive rate reflected in previous research and RCOG Guidelines. In premenopausal women, at a fixed specificity of 90%, LR2 and ROMA were found to have a statistically significantly higher sensitivity compared to RMI1. ACOG v3 (CA125) and ACOG v3 (OVA1) were observed to have a lower sensitivity compared to RMI1 at a fixed specificity of 90% but this was not statistically significant.

6.2.2 Applicability of findings to the review question

Translating these findings to estimates from an on-going trial of the prevalence of ovarian cancer in postmenopausal women in secondary care of 10% [231].

For a hypothetical cohort of 1000 premenopausal women in secondary care, assuming a prevalence of ovarian cancer of 10% (100 women have ovarian cancer and 900 women do not) a fixed test specificity of 90% would result in 90 false positives.

The results of the analyses suggest that in premenopausal women 18 (14 to 21) would have their cancer missed (false negatives) if tested with LR2, leading to a delay in diagnosis, morbidity and possibly mortality, 25 (20 to 31) would have their ovarian cancer missed by ROMA, 41 (35 to 46) by RMI, 27 (19 to 36) by ACOG v3 (CA125) and 18 (6 to 43) by ACOG vs OVA1.

In postmenopausal women there was no evidence for statistically significant differences in sensitivity for ROMA, LR2, ACOG v3 (CA125) or ACOG v3 (OVA1) at a fixed specificity of 90%. The results of the analyses suggest that in postmenopausal women 18 (9 to 31) would have their cancer missed if tested with LR2, 18 (11 to 27) would have their ovarian cancer missed if tested by ROMA, 28 (21 to 36) by RMI, 29 (18 to 44) by ACOG v3 (CA125) and 39 (10 to 78) by ACOG v3 OVA1.

Corresponding estimates adopting a baseline prevalence of ovarian cancer of 30%, (representative of a tertiary care setting) with a fixed specificity of 90% are illustrated in the Summary of Findings (Table 4.8 and 4.9).

It is important to note that in practice women and clinicians may be willing to accept a reduction in specificity to 90% to achieve a higher sensitivity and that this is likely to be influenced by factors such as baseline risk (prevalence) of ovarian cancer which will be dependent on healthcare setting and menopausal status and the adverse consequences of unnecessary investigation and treatment, for example loss of fertility.

6.3 Strengths and weaknesses of the review

6.3.1 Strengths

The strengths of this review are based on the thorough approach to the search, data extraction and data analysis and presentation of results and summaries of findings, all of which aimed at providing the best estimates of test accuracy relevant for clinical practice.

This is a comprehensive review of test combinations for the diagnosis of ovarian cancer. Although literature searches were completed in 2015 this review remains the most up to date comprehensive review to my knowledge. Sensitive search strategies were used 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). A comprehensive systematic search including a search for studies performed in primary care (family practice). No studies were identified of the accuracy of biomarkers to diagnose ovarian cancer in a primary care setting and this is identified as a gap in knowledge. The studies were analysed and test performance compared both by separate patient group (postmenopausal and premenopausal women) and by clinically relevant test thresholds in each patient group. This sought to identify the accuracy of each test that would have not only the most relevance test but the most useful threshold in clinical practice.

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 test combinations relevant to clinical practice at the time of writing. An attempt was made to mitigate against heterogeneity by restricting the analysis to primary tumours of adnexal origin and where this was not possible or unclear in studies with mixed primary, recurrent and metastatic disease, this was reflected in downgrading of quality assessment.

Previous reviews have not recognised the importance of separate analysis of pre and postmenopausal women at the same time as ensuring meta-analysis is adequately analysed for the test thresholds with different test performance. The test accuracy results were investigated for a range of thresholds for each biomarker using both numerical values (+/-10) and +/-2 standard deviations; results of both approaches were similar demonstrating the robustness of the analysis. Finally, the effect of head to head comparison of tests and appropriate categorisation of borderline tumours on test accuracy was undertaken by analysing results using only studies that included direct comparison and studies including and excluding appropriate BOTs categorisation respectively.

6.3.2 Weaknesses

Due to time and resource constraints non English Language studies were not considered for data extraction and analyses. The impact of this omission on study findings is unknown. A

further weakness of the review is the fact that the search strategy is now 4 years old. I am aware of more recent primary publications that may have an impact on the conclusions of this review, in particular publication of several papers investigating the accuracy of ADNEX. In this review, only one ADNEX study was identified and therefore it was not possible to include this test in statistical comparisons of test accuracy. Factors to consider regarding the importance of this omission include the fact that variables in the ADNEX model overlap considerably with components in the test combination LR2, which outperformed RMI in premenopausal women in the comparative analysis. However, variables included in the ADNEX model include whether US examination is being performed in secondary or tertiary care settings which suggests estimates of accuracy will have limited applicability in primary care settings.

The majority of the limitations of this review originate from deficiencies in the primary evidence base. Lack of data and poor reporting precluded quality assessment and investigation of potential important sources of variation in test accuracy estimates. These potential sources included clinical setting (primary versus specialist), target condition (in particular primary versus recurrent and metastatic disease) and cancer histological subtype and stage; included studies were very mixed with respect to the range of ovarian pathology included with some for example restricting to epithelial ovarian cancer whilst others included metastatic disease to the ovaries in the definition of disease positive which could not be disaggregated for the purposes of analysis. Lack of disaggregation in primary studies was most marked when considering pre and postmenopausal subgroups separately.

6.4 Applicability of findings to the review question

This review question aimed to answer the question of the accuracy of test combinations (symptoms and signs, imaging and biomarkers) for women with symptoms suspicious for ovarian cancer. In the UK, the National Institute for Health and Care Excellence (NICE)[115] and the Royal College of Obstetrician and Gynaecologists (RCOG)[232] recommend women with suspicious symptoms presenting in primary care should receive additional investigations with CA125 and if raised, a transvaginal ultrasound (TVS) scan should also be performed prior to referral to secondary care. Once in secondary care either RMI1 or ultrasound examination using International Ovarian Tumour Analysis (IOTA) criteria are used to determine if onward referral to tertiary care (gynaecological oncologist) is required for further management.[232]

The American College of Obstetrics and Gynecology (ACOG)[24] recommends TVS as the initial test of choice if physical examination suggests the presence of an adnexal mass. Subsequent referral to tertiary care is recommended in the presence of any one of risk factors (family history), abnormal findings on imaging, a raised CA125, a raised OVA1 or a raised ROMA test.

In the UK therefore, the presence of suspicious symptoms is a trigger for further investigation which includes testing with biomarkers and US in primary care and the use of combination tests in secondary care. The majority of included studies were judged as at high or unclear applicability to the review question on the basis that women were either asymptomatic or it was not clear if they were symptomatic at the point of index test use. Further no studies of the accuracy of test combinations to diagnose ovarian cancer in a generalist setting were identified. The majority of included studies have a prevalence of ovarian cancer that is in keeping with tertiary hospitals. Prevalence estimates varied in different reviews based on the studies; in premenopausal women median prevalence was CA125 studies - 22% (range 3% -18%) HE4 studies - 25% (range 8% - 81%) combination studies - 19% (range 9%- 28%) and postmenopausal women CA125 studies - 46% (range 8% -71%) HE4 studies - 50% (range 18% - 81%) combination studies - 44% (range 40%- 57%) which is higher than would be expected for either premenopausal or post-menopausal women indicating advanced position in the patient pathway. Estimates from the literature estimate the prevalence of ovarian cancer in women with symptoms was 0.23% in primary care.[83, 233] Test accuracy estimates from this review are therefore unlikely to be applicable to generalist settings where the prevalence of ovarian cancer is lower and the spectrum of the tested population more heterogeneous. With the exception of one study[138] all included women had a confirmed adnexal mass at the point of testing. This single study was identified as an outlier, with the lowest estimated specificity (53%) and highest sensitivity (94%). Early in the ovarian cancer 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 (for example 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 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.

With the exception of ROMA, all combination test included in this review included an US imaging component, particularly the LR2 and ADNEX models. The applicability of the review findings to practice will therefore be influenced by the skill of the US operator. With the exception of one LR2 study[203] all other studies were judged as at high or unclear risk of bias on the basis that sonographers were specialists or their level of skill was not reported; therefore it cannot be assumed that the performance of RMI, LR2, ACOG v3 (CA125), ACOG v3 (OVA1) 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 the fact that in the majority of 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[234] and excluding them from estimation of test accuracy will artificially alter the spectrum of disease compared to that occurring in practice. No evidence was found in the combination review to demonstrate that management of borderline tumours significantly affected estimates of accuracy. However, this investigation is limited by the volume of data available, lack of clarity in reporting and the possible effects of confounding. In a similar investigation undertaken for borderline tumours in the biochemical markers for the diagnosis of ovarian cancer, a trend was observed (in CA125 biomarker studies- larger dataset,) for higher average estimates of sensitivity from studies that either excluded borderline tumours from analysis or where the classification of borderline tumours for analysis was unclear compared to studies where borderline tumours were classified as positive (grouped with malignant). This difference was significant for postmenopausal women. It is therefore possible that estimates of sensitivity may be overestimated in this review.

6.5 Conclusions

6.5.1 Implications for practice

This review has demonstrated that menopausal status causes consistent and significant 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 should be considered separately in premenopausal and postmenopausal women.

For biomarkers, in premenopausal women, HE4 at a threshold of 60-80pMol/L and 130-150pMol/L is the preferred test in test settings with prevalence representative of primary care and specialist settings respectively; In postmenopausal women, HE4 at a threshold of 130-150pMol/L and CA125 at a threshold of 30-40U/ml is the preferred test in test settings with prevalence representative of primary care and specialist settings respectively.

The other advantages the biomarkers offer as a triage test in the primary care setting are they are less expensive (HE4 and CA125) less invasive, easier to do, less uncomfortable for patients, technically less challenging, and more objectively interpreted.

The implications of the findings from both systematic reviews to practice for premenopausal women presenting with an adnexal mass suspicious for ovarian cancer, LR2 or ROMA are the test combinations of choice to determine future management. Current Guidelines recommending RMI as a diagnostic or triage test in premenopausal women in secondary care settings should be changed to recommend LR2 or ROMA. Similarly, CA125 threshold of 200U/ml as recommended by the RCOG guidance is based on a single study and has a performance of sensitivity of 29% and specificity of 98% in premenopausal women. Given the extremely limited data on which this recommendation is based, this review recommends that this threshold should not be used as a cut off as a triage tool for referral from secondary to tertiary care. Premenopausal patients, being investigated with adnexal masses in secondary care would be more accurately triaged into tertiary care with the use of combination tests such as LR2 or ROMA.

For postmenopausal women presenting with an adnexal mass in specialist settings, no test combination demonstrated statistically significant performance differently to RMI.

The single ADNEX study could not be included in the statistical comparison of tests and I, therefore have to acknowledge the uncertainty about the relative performance of ADNEX.

Choice of which combination test (LR2 or ROMA) should replace RMI in practice in premenopausal women in secondary care will require consideration of relative costs and the feasibility of introducing the test. The adoption of ROMA does not rely on availability of

ultrasound expertise but would require investment in laboratory facilities for processing HE4 tests. No evidence was identified to support recommendation of a specific ROMA threshold for use in clinical practice beyond the ranges used in included studies: between 7.4 to 13.1 in premenopausal women and 14.4 to 27.7 in postmenopausal women. The adoption of LR2 into routine practice would require investment in sonographer training.

The implications of this for a test used in primary/Secondary care would be that 90% (900) women would be referred to specialist care appropriately. Inappropriate referral cause anxiety for the woman concerned as well as further inappropriate testing and possible unnecessary investigations. In addition, false positive referrals to secondary care divert resources from those women who actually have malignancy. The added implications of this for a test used in specialist care is that unnecessary false positive results may receive extensive surgery with consequent risk to health and to preservation of fertility for premenopausal women.

The implications of these findings for women presenting in generalist settings, early on in the diagnostic pathway, 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. The median prevalence of OC in the different reviews was upwards of 19% and as high as 50%. Included participants are therefore likely to represent a 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.

6.5.2 Implications for research

The priority for research at this time is to update this review to consider more recent publications (after 2015), in particular, to include ADNEX in a comparative analysis of accuracy. However, it should be noted that the ADNEX model includes a variable for the setting in which the model is applied (oncology versus other hospitals) and so whilst its comparison may be robust to the setting in which included studies have been conducted, the model would not be applicable to non-specialist settings in its current form.

Due to variation in thresholds reported in included studies and in the absence of evidence for clinically significant threshold effects for individual tests the approach to comparative analysis for combination tests was to fix specificity at a clinically applicable threshold of 90%. However, the difference in test performance in premenopausal compared to postmenopausal

women demonstrated in this review raises the possibility that using different tests in pre and postmenopausal women may be necessary to optimise diagnostic utility. In premenopausal women avoiding unnecessary surgery and preserving fertility by maximising specificity and minimising false positives may be considered relatively more important than in postmenopausal women. Primary research investigating the performance of tests at clinically agreed thresholds, in a clinically suspected population and separately in pre and postmenopausal women would facilitate exploration of thresholds and tests specific to the menopausal status of women being tested. Indeed, the results from a longitudinal test accuracy study to validate risk scores in symptomatic women Refining Ovarian Cancer Test accuracy Scores[231] is awaited. 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.

This review has demonstrated the lack of research on the accuracy of tests early in the ovarian cancer testing pathway: the first clinical encounter of symptomatic women in whom ovarian cancer is being considered as a differential diagnosis. The review also identifies that most studies have a high prevalence of cancer, suggesting the impact of prior testing and specialist care settings. The diagnostic utility of tests in generalist settings could be evaluated on the basis of current NICE Guidelines (NG 12) ; taking into account the financial and clinical costs of broadening the recommendations, the guideline development group agreed to use a 3% PPV threshold value to underpin recommendations for suspected cancer pathway referrals and urgent direct access investigations. Future studies performed earlier in the ovarian cancer diagnostic pathway should also take care to report aspects of setting that will have a bearing on test performance such as healthcare setting (for example 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.

CHAPTER 7: SUMMARY OF THESIS

7.1 Background

This thesis aimed to identify the best diagnostic tests for ovarian cancer through narrative review of reviews and systematic reviews. An initial scoping search identified a large volume of literature including many systematic reviews. Thus, faced with the fundamental question as to whether the planned systematic reviews were justified, I conducted a review of reviews to check if the research question had been previously answered and to assess the quality of the evidence. I found that published systematic reviews were of variable quality and scope and the answer lacking in clarity, both for accuracy of tests and applicability to clinical practice. Therefore, I conducted a further 2 systematic reviews to evaluate the accuracy, applicability and currency of biomarkers alone and combination testing in the diagnosis of ovarian cancer in symptomatic women. The search and methodology employed for these reviews are common. This section summarises the main findings and discusses the interpretation of results for clinical practice and research including the Quantity and quality of evidence.

7.2 Review of existing reviews

The Review of reviews includes 20 publications reporting 23 separate systematic reviews. Studies that evaluated tests that are currently used clinically or FDA approved for use in the initial diagnosis and management of ovarian cancer were included. Two reviews investigated the accuracy of symptoms (both individual and in combination), 27 investigated the accuracy of a range of biomarkers (CA125, HE4, OVA1) and 35 investigated the accuracy of test combinations (RMI1-4, ROMA, ACOG, LR1 and LR2).

7.2.1 Quality assessment

The review of existing reviews found methodological blind spots in; Question formulation with missing data on many key variables including: presentation, prior testing, menopausal status, test positivity thresholds of index test, target condition and reference standard leading to a lack of objectivity and transparency in inclusion of studies. Heterogeneity of ovarian cancer is a key

consideration for DTA and can affect test accuracy and recommendations. For e.g. defining ovarian cancer or benign disease by pathology type, symptomatic or asymptomatic are essential while considering inclusion criteria to enable selection of right studies for meta-analysis, comparison of tests, sensitivity analysis and make apt recommendations. The quality assessment also did not provide clarity or details on items pertaining to risk of bias and applicability (ROB and ROA). The applicability of the reviews could not be assessed due to lack of information regarding clinical presentation and prior testing in primary studies and its relevance to accuracy, with only one included review commenting on this key aspect. None of the reviews presented all the relevant information on PppITR. Quality assessment was unclear in 75% of reviews and hierarchical models was not used in 75% of reviews; important sources of heterogeneity are not incorporated, all resulting in poor applicability to clinical practice.

7.2.2 Reporting as assessed by PRISMA DTA checklist

All the reviews did not meet the reporting standard. Key elements of methodology such as selection criteria, study selection, data extraction or planned syntheses were lacking and where available lacked clarity; for e.g. even though the index test and the reference standard were the best reported items in the majority of the reviews, threshold of tests was not clarified. The methods described for meta-analyses often did not clarify the criteria for studies being pooled and no detail on how differences in threshold, histology and reference standard would be handled in analyses. Similar deficiencies were identified in the reporting of results. The lack of details in reporting precludes any assessment or comment regarding any study specific tailoring of the QUADAS tool that may have been undertaken to appropriately assess all aspects of ROB/ROA. All but 1 included reviews reported on accuracy of tests but stratified results by menopausal status/ BOTs/ stage or additional analyses investigating heterogeneity were reported in less than 25% of the studies.

7.2.3 Conclusion

In conclusion, my assessment of the existing reviews i) did not identify any health care models that compared all the tests head to head as stratified by menopausal status ii) identified major limitations in the existing reviews where differentiated results for pre and postmenopausal results were available, iii) did not address prevalence and spectrum of ovarian cancer in the populations studied. iv) Showed lack of clarity on inclusion, prevalence and the accuracy of

tests for BOTS; potentially affecting the accuracy of the overall results, as well as limiting clinical applicability for use in management which is a significant limitation v) Needs improved reporting of results using PRISMA DTA checklist vi) In my opinion, cannot be used to inform clinical practice.

7.3 Biomarkers and combination review

The biomarker and combination reviews identified common methodological issues in included primary studies. For both reviews, clinical pathways of patients from presentation to decision for surgery intervention were not detailed in any included studies and the majority of the studies only included women undergoing surgery for adnexal mass.

7.3.1 Quality of included primary studies

The majority of included studies showed unclear or high ROB and ROA due to issues with i) patient selection as inadequate information on symptoms ii) reference standard when information regarding inclusion of BOTs were unclear iii) flow and timing domain with lack of clarity of interval between the conduct of the index test and the reference standard and verification bias as studies only included women undergoing surgery. All studies for biomarkers were considered low ROB due to inherent objectivity of interpretation unlike studies incorporating ultrasound (RMI, LR2, ACOG and variations, ADNEX) where there is potential for confirmation bias i.e. knowledge about disease status affecting index test interpretation. Appropriate categorisation of BOTs, menopausal status and direct vs indirect comparison of tests in studies were considered when analyses were conducted.

7.3.2 Results of the meta-analyses

Results of the meta-analysis of the combination review

The meta-analyses finds i) a difference in the accuracy of diagnostic tests by menopausal status. Broadly, tests demonstrate greater sensitivity and lower specificity in postmenopausal women across all versions of index tests at all thresholds identified in the review ii) LR2 or ROMA demonstrate the best sensitivity at fixed specificity of 90% in premenopausal women. No test combination performance showed statistically significant difference compared to RMI in

postmenopausal women iii) Analysis of accuracy for tests in BOTs was only possible for ROMA and showed no difference in test accuracy iv) No significant accuracy difference by threshold was seen for ROMA or RMI1 v) Limited reporting on expertise of the operator in studies incorporating USS.

Statistical comparison between ADNEX and any test was not possible because only one ADNEX study was included in this review. Prevalence estimates in premenopausal women varied between 9% and 28%, median 19% and in postmenopausal women 40% to 57%, median 44% which represents a very highly selected population which represents tertiary care setting. Applicability to clinical practice is limited because of lack of investigation of important sources of heterogeneity and lack of primary studies in settings of lower prevalence, i.e. unselected populations

Results of the meta-analyses of the biomarker review

Meta-analyses were performed comparing different thresholds within each test for pre and postmenopausal women with appropriate BOT categorisation when enough studies were available and also for direct and indirect comparison.

The meta-analyses found difference in the accuracy of diagnostic tests by menopausal status.

1) For CA125 prevalence will determine the preferred test as sensitivity (higher) and specificity (lower) were significantly different between threshold cut off of 30-40U/ml and 60-70U/ml in both pre and postmenopausal women. The prevalence of the test setting will ordain if the increase in sensitivity is worth the decrease in specificity. 2) HE4 at a threshold of around 130-150pMol/ is recommended in postmenopausal women because of increased specificity. 3) In premenopausal women, there is again a trade off between sensitivity and specificity between test positivity cut off threshold for 60-70 pMol/L and 130-150pMol/L. Therefore, increased sensitivity at a threshold of 60-70 pMol/L has to be balanced against a decrease in specificity at a threshold of 130-150pMol/L with prevalence at the point of testing guiding the choice of threshold used. 4) Between HE4 is and CA125, HE4 is recommended in premenopausal women; i) HE4 at a threshold of 60-80pMol/L and 130-150pMol/L is the recommended test when the prevalence at the point of testing represents a typical primary care populations and specialist settings respectively. 5) In postmenopausal women, i) HE4 at a threshold of 130-150pMol/L is the preferred test in settings with prevalence typical of primary care populations

ii) CA125 at a threshold of 30-40U/ml is the preferred test in test settings with prevalence typical of specialist settings. 6) Analysis of accuracy for tests in BOTs was only possible for CA125 threshold of 30-40U/ml and shows statistically significant difference in postmenopausal women.

7.4 Implications

7.4.1 Implications of review of existing reviews

The narrative review of reviews influenced the search date and strategy of the systematic reviews on biomarkers and combinations. The deficiencies and limitations identified during the review of reviews helped tailor the quality assessment tool to provide information that is transparent and provide context to testing. The lack of clinical applicability of the data in the existing reviews informed the analyses on menopausal status, test thresholds, BOT categorisation and the plan for heterogeneity analyses for the systematic reviews.

7.4.2 Clinical Implications of combination and biomarker review

The review identifies no studies investigating the accuracy of biomarkers to diagnose primary ovarian cancer in a primary care setting.

For Premenopausal women presenting with an adnexal mass suspicious for ovarian cancer, LR2 or ROMA are the test combinations of choice to determine future management. RMI as a diagnostic or triage test in premenopausal women in secondary care settings should be replaced by LR2 or ROMA. This analyses showed no statistical differences between diagnostic tests in comparison to RMI in postmenopausal women presenting with an adnexal mass in specialist settings. As all the included studies were conducted in tertiary or secondary settings, implications of these findings for women presenting in generalist settings, early on in the diagnostic pathway, is less clear.

In prevalence representative of primary care setting, HE4 is recommended at threshold test positive cut off of 60-80pMol/L for premenopausal women and 130-150pMol/L for postmenopausal women. In prevalence representative of secondary care setting, HE4 at test

positive cut off of 60- 80pMol/L is recommended for premenopausal women and CA125 at threshold test positive cut off 30-40U/ml for postmenopausal women.

However, one must be mindful while interpreting these recommendations, i) they are based on studies which were all conducted in secondary care or tertiary care setting with varying prevalence ii) they are based on limited number of studies and iii) recommendations are based on a mathematical construct that likely represent average prevalence in primary and secondary care settings

7.4.3 Recommendations for future research

Studies based in lower prevalence and in the primary care setting are required to enable recommendations for primary care to improve triaging and resource utilisation. Primary research investigating the performance of tests in a clinically suspected population, at clinically agreed thresholds and reporting results separately in pre and postmenopausal women would facilitate exploration of thresholds and tests specific to the menopausal status and tailor recommendations. This would also be a benchmark for new tests discovered in the future to determine utility and role. The clinical presentation and setting of recruitment be clearly reported to determine relevant clinical utility and applicability to define the role of the test as triage, replacement or add on. The Information on clinical pathway needs to be reported to enable judging the results in context and pool studies appropriately for meta-analyses and reduce the impact of heterogeneity (symptomatic women/ prior testing/ clinical setting of the test). The expertise of the sonographers where ultrasound is a component of the test combination should be reported to enable applicability and implications on training and resource. Primary research investigating the performance of tests in a clinically suspected population, at clinically agreed thresholds and reporting results separately in pre and postmenopausal women would facilitate exploration of thresholds and tests specific to the menopausal status and tailor recommendations. Given what we now know regarding the biological behaviour primary studies should in future clearly report the histological details of tumours especially borderline tumours and early stage to enable patient tailored decisions and drive future goals for research. Test accuracy studies should report on all women undergoing testing (test negative patients and patients managed conservatively) using follow-up as the reference standard to avoid verification bias and enable patient counselling for testing and risk of false negative in low risk patients.

7.5 Strength and weakness

7.5.1 Strength and weakness of review of existing reviews

To my knowledge, this is the only review of reviews evaluating the currency and applicability of the existing diagnostic test accuracy reviews in ovarian cancer. It is also the first and at the time of writing, the only review that has assessed the reporting quality of the existing DTAs on ovarian cancer using the newly published PRISMA DTA checklist. The search is comprehensive and has evaluated all the tests in routine clinical practice. Non-English reviews are not included and double data screening and evaluation was not undertaken due to lack of resources. Random samples of review of reviews were double data evaluated as a pilot sample for consistency of understanding and use.

7.5.2 Strengths and weakness of biomarker and combination review

Comparison of all test combinations relevant to routine practise was undertaken. Based on the knowledge and understanding gained from review of existing reviews, gaps were identified and efforts made to tailor the review to answer the gaps on knowledge and make the review clinically relevant and limit the impact of heterogeneity. Systematic evaluation of effect of menopausal status, threshold effects and systematic evaluation of impact of categorisation of BOTs were undertaken to tailor recommendations and improve clinical utility. In the summary of findings tables, likely numbers of women with test results for biomarkers HE4 and CA125 at different test thresholds for different prevalence of ovarian cancer representing different steps in the clinical pathway of diagnosis are provided to demonstrate the clinical applicability and accuracy of test in different settings.

There was a significant interval between the search date and review results and this will be addressed by publishing an update soon. Foreign and unpublished literatures were excluded. Some of the objectives could not be achieved due to limitations in the primary studies.

7.6 Conclusion

In conclusion, my thesis finds key methodological issues in literature that preclude clear recommendations for diagnostic testing in women with adnexal masses. With the caveat of these limitations, I recommend i) Threshold test positivity of 60-80pMol/L and 130-150pMol/L for HE4 in pre and postmenopausal women for low prevalence setting ii) ROMA or LR2 in premenopausal women to replace RMI in secondary/tertiary setting. For postmenopausal women, I am unable to make similar recommendations. Further research is needed to overcome the issues identified.

APPENDIX

Appendix 1 **: Biomarkers Search strategy.**

Database: Ovid MEDLINE(R) 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 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="2009-2015")

Appendix 2 **: Data extraction form**

Please enter your initials (NR, RC)

STUDY IDENTIFICATION AND STUDY TYPE

Study ID (number allocated in included studies PDFs and 'notes' field in reference manager)	Comments
Study authors (surname and year) Country in which study conducted Study design Please choose from: - 'Prospective' cross-sectional test accuracy study **(P CS)** - 'Retrospective' cross-sectional test accuracy study **(R CS)** - Case-control test accuracy study **(CC)** - Comparison of the accuracy of tests or testing strategies in 2 different populations (e.g. a randomised trial of tests sor testing strategies) **(Between-person comparison - BPC)** - Within-person comparison of test accuracy **(WPC)** - Unclear **(U)**	

PATIENT SELECTION DETAIL

For studies comparing two index tests or testing strategies in different patient populations complete details for each patient population (copy and paste table if necessary)

FOR NON-COMPARATIVE STUDIES OR WITHIN-PERSON TEST COMPARISONS (WPC) assessing the accuracy of one index test or one index testing strategy **describe methods of participant selection as reported** (cut and paste from paper if possible)		**Comments**
Include total number of study participants		
FOR BETWEEN-PERSON COMPARATIVE STUDIES (BPC) of two index tests or testing strategies in different patient populations **describe methods of participant selection receiving each index test or testing strategy as reported**		
(cut and paste from paper if possible)		
Include total number of study participants receiving each test or testing strategy		
Clinical setting mentioned	**Y/N/U**	
If yes	**Primary/community**	
	Secondary/hospital/cancer unit/cancer centre	
	NR	
Risk factors such as age, menopause, family history, BRCA status, other cancers mentioned in the study		

INCLUDED PATIENT CHARACTERISTICS DETAIL

For studies comparing two index tests or testing strategies in different patient populations complete details for each patient population (copy and paste table if necessary)

FOR NON-COMPARATIVE STUDIES OR WITHIN-PERSON TEST COMPARISONS (WPC) describe characteristics of included patients **as reported (cut and paste from paper if possible)**		Comments
FOR BETWEEN-PERSON COMPARATIVE STUDIES (BPC) **of two index tests or testing strategies in different patient populations** describe characteristics of participants receiving each index test or testing strategy as reported **(cut and paste from paper if possible)**		
Age as reported or not reported ('NR') *(delete options as necessary)*	**- Age range:** **- Age mean (SD):**	
Menopausal status (n/%)	pre	Post
Prior test(s)	**Symptoms** **Signs** **Biomarker/s**	
Histology	**Number (%)**	
Benign	**Number (%)**	
Endometriosis		
Others		
Tumours of low malignant potential (LMP/borderline)	**Number (%)**	
Malignant	**Number (%)**	
I	**Number (%)**	
II	**Number (%)**	

III	**Number (%)**
IV	**Number (%)**

PATIENT SELECTION RISK OF BIAS

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 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 co-morbidities such as infertility and endometriosis	
Could the selection of patients have introduced bias?	**RISK: LOW/HIGH/UNCLEAR**
If a) and b) and c) 'YES' = low risk of bias	
If a) or b) or c) 'No' = high risk of bias	
If a) or b) or c) 'Unclear' = unclear risk of bias	

B. Concerns regarding applicability	
Describe included patients (prior testing, presentation, intended use of index test and setting):	
Is there concern that the included patients do not match the review question? ***a) Patients all symptomatic OR symptomatic and asymptomatic can be disaggregated*** ***b) Prior tests primary care: self reported symptoms*** ***c) Prior tests secondary care: self reported symptoms OR self reported symptoms PLUS one or more of biochemical markers and ultrasound***	**CONCERN: LOW/HIGH/UNCLEAR** **Low - a) and b) and C) Yes** **High - a) or b) or C) No** **Unclear - a) or b) or C) Unclear**

INDEX TEST(S) DETAILS

For studies comparing two index tests or testing strategies in different patient populations complete details for each index test or testing strategy (copy and paste table if necessary)

INDEX TEST(S)	Test (note the type of symptom, biomarker, ultrasound variable)	Test threshold (what constitutes abnormal test symptoms, signs, biomarkers or USS: threshold value or fixed value of abnormality	Threshold value or fixed value of abnormality	Clinical setting in which index test performed	If operator was blinded to previous test (s) result	Describe what prior tests information was available to those interpreting index test	Detail about conduct of index test that might be a source of heterogeneity (e.g. experience of operator (ultrasound, symptoms), type of technology (biomarkers)) - For test combinations test order and rule for combining tests
	Symptoms (list)	Yes No NR	(Number of symptoms and time element)	Primary/community/family practice Secondary: hospital/cancer unit/cancer centre/gyn oncologist NR/U/Mixed	Yes No NR		
	Signs	Yes No NR		Primary/community/family practice Secondary: hospital/cancer unit/cancer centre/gyn oncologist NR/U/Mixed	Yes No NR		
	Biomarkers	Yes No NR		Primary/community/family practice Secondary: hospital/cancer unit/cancer centre/gyn oncologist NR/U/Mixed	Yes No NR		
	USS	Yes No NR		Primary/community/family practice Secondary: hospital/cancer unit/cancer centre/gyn oncologist NR/U/Mixed	Yes No NR		
	Combination	Yes No NR		Primary/community/family practice Secondary: hospital/cancer unit/cancer centre/gyn oncologist NR/U/Mixed	Yes No NR		

If a **combination of tests (a testing strategy) was used for each participant** please detail:	**Combination:**	**Comments**
- What combination of tests?	**Order:**	
- The order in which tests were performed?	**Rule for combining tests:**	
- The rule for combining test results:		
E.g. + and + = surgery		
E.g. + and - = surgery		
E.g. + and - = no surgery		
E.g. – and - = no surgery		
Etc		
or not reported **('NR')**		

INDEX TEST(S)

(If more than one index test was used, please complete for each test).

A1. Risk of bias (symptoms)	
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 prespecified?	Yes/No/Unclear
Could the conduct or interpretation of the index test have introduced bias?	**RISK: LOW/HIGH/UNCLEAR** **High - a) or b) No** **Low- a) and b) Yes** **Unclear - a) or b) Unclear**

A2. Risk of bias (ultrasound**)**	
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 prespecified?	Yes/No/Unclear
Could the conduct or interpretation of the index test have introduced bias?	**RISK: LOW/HIGH/UNCLEAR** **High - a) or b) No** **Low - a) and b) Yes** **Unclear - a) or b) Unclear**

A3. Risk of bias (Biomarkers) rule different because objective test in comparison to US and symptom elicitation	
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 prespecified?	Yes/No/Unclear
Could the conduct or interpretation of the index test have introduced bias?	**RISK: LOW/HIGH/UNCLEAR** **High - b) No or (a) and b)) No** **Low - a) and b) Yes** **Unclear - a) or b) Unclear**

A4. Risk of bias (within-study combination)	
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 prespecified?	Yes/No/Unclear
c) i) Were symptoms/signs interpreted without knowledge of ultrasound or biomarkers; ii) was ultrasound interpreted without knowledge of biomarkers	Yes/No/Unclear
Could the conduct or interpretation of the index test have introduced bias?	**RISK: LOW/HIGH/UNCLEAR** **High - a) or b) or c) No** **Low - a) and b) or c)Yes** **Unclear - a) or b) or c) Unclear**

B. Concerns regarding applicability	
a) Is the skill of person performing US and eliciting symptoms detailed? (level of training and/or experience)	**Yes/No/Unclear/NA**
b) Was US performed in all patients by non-specialised sonographers	**Yes/No/Unclear**
c) Was US 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?	**CONCERN: LOW/HIGH/UNCLEAR** **High - a) or b) or C) No** **Low - a) and b) or c)Yes** **Unclear - a) or b) Unclear**

REFERENCE STANDARD AND TARGET CONDITION DETAIL

REFERENCE STANDARD

Surgery (%) Follow-up (%) and length of follow-up

TARGET CONDITION

TARGET CONDITION	**Epithelial ovarian cancer (EOC)**	Number (%)	**Comments**
Target conditions are ovarian cancer (see list of different histology of ovarian cancer)	Serous		
	Mucinous		
	Endometrioid		
	Clear		
	Germ cell tumours		
	Stromal cell tumours		
	LMP		
	Others (metastasis)		

REFERENCE STANDARD RISK OF BIAS

A. Risk of bias

Describe the reference standard and how it was conducted and interpreted:	
a) Is the reference standard likely to correctly classify the target condition? Index test +ve: Histology following laparoscopy or laparotomy	**Yes/No/Unclear**
b) Is the reference standard likely to correctly classify the target condition? Index test –ve: Yes - if a minimum follow-up period of greater than 12 months is included as required to assess whether the target condition is present No - if a minimum 12-month follow-up period is absent Unclear - if no information on follow-up period is included	**Yes/No/Unclear**
Could the reference standard, its conduct or its interpretation have introduced bias	**RISK: LOW/HIGH/UNCLEAR** **High - a) or b) No** **Low - a) and b) Yes** **Unclear - a) or b) Unclear**

DOMAIN 3: REFERENCE STANDARD (continued)

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN: Yes/No/Unclear**
Yes - ovarian cancer, borderline and metastatic disease are not differentiated (and cannot be for analysis) No - ovarian cancer, borderline and metastatic disease can be differentiated for analysis Unclear - unclear if ovarian cancer, borderline and metastatic disease have been disaggregated	

DOMAIN 4: 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 x 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 conduct or interpretation of reference standard have introduced bias?	**RISK: LOW/HIGH/UNCLEAR** **LOW - a) and b) and c) and d) - Yes** **HIGH - a) and b) and c) and d) - No** **UNCLEAR - a) and b) and c) and d) – Unclear**

COMPARATIVE DOMAIN (if applicable)

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? LOW - a) OR (b) and c)) - Yes HIGH - a) OR (b) and c)) - No UNCLEAR - a) OR (b) or c)) – Unclear	**RISK: LOW/HIGH/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 *Low - No* *High - Yes* *Unclear – Unclear*	**CONCERN: LOW/HIGH/UNCLEAR**

RISK OF BIAS FOR MULTIVARIABLE DIAGNOSTIC MODELLING STUDIES (if applicable)

1. Participant selection	**DEV**	***Yes/No/Unclear***	**VAL**	***Yes/No/Unclear***
a) Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	**DEV**	*Yes/No/Unclear*		
b) Were participants enrolled at a similar state of health, or were predictors considered to account for differences?	**DEV**	*Yes/No/Unclear*	**VAL**	*Yes/No/Unclear*
Could the selection of patients have introduced bias? **HIGH**: a) OR a) and b) - YES **LOW**: a) OR a) and b) - NO **UNCLEAR**: a) OR a) and b) – UNCLEAR		**HIGH/LOW/ UNCLEAR**		**HIGH/LOW/ UNCLEAR**
3. Predictors	**DEV**	***Yes/No/Unclear***	**VAL**	***Yes/No/Unclear***
a) Were predictors defined and assessed in a similar way for all participants?	**DEV**	*Yes/No/Unclear*		
b) Are all predictors available at the time the model is intended to be used	**DEV**	*Yes/No/Unclear*	**VAL**	*Yes/No/Unclear*
c) Were all relevant predictors analysed?: ***No*** if symptoms only; ***No*** if US index test only; ***No*** if combination of index tests (symptoms, US and biomarkers) but miss out US OR Symptom OR FDA approved biomarkers	**DEV**	*Yes/No/Unclear*	**VAL**	*Yes/No/Unclear*
Could the definition, measurement or analysis of predictors introduced bias? **HIGH**: a) OR b) OR c) - YES **LOW**: a) OR b) OR c) - NO **UNCLEAR**: a) OR b) OR c) – UNCLEAR	**DEV**	**HIGH/LOW/ UNCLEAR**	**VAL**	**HIGH/LOW/ UNCLEAR**
3. ANALYSIS	**DEV**	***Yes/No/Unclear***	**VAL**	***Yes/No/Unclear***
a) Were there a reasonable number of outcome events?	**DEV**	*Yes/No/Unclear*		
b) Were there a reasonable number of outcome events?			**VAL**	*Yes/No/Unclear*
c) Were non-binary predictors handled appropriately?	**DEV**	*Yes/No/Unclear*	**VAL**	*Yes/No/Unclear*
d) Was selection of predictors based on univariable analysis avoided?	**DEV**	*Yes/No/Unclear*		
e) Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	**DEV**	*Yes/No/Unclear*		
f) For the model or any simplified score, were relevant performance measures evaluated, e.g. calibration, discrimination, (re)classification and net benefit?	**DEV**	*Yes/No/Unclear*	**VAL**	*Yes/No/Unclear*
g) Was the model recalibrated or was it likely (based on the evidence presented, e.g. calibration plot) that recalibration was not needed?	**DEV**	*Yes/No/Unclear*	**VAL**	*Yes/No/Unclear*
h) Was model validation undertaken in individuals other than those in the model development (external validation)?			**VAL**	*Yes/No/Unclear*
Could the analysis strategy have introduced bias? **HIGH**: a) OR b) OR c) OR d) OR e) OR f) OR g) OR h) - YES **LOW**: a) OR b) OR c) OR d) OR e) OR f) OR g) OR h) - NO **UNCLEAR**: a) OR b) OR c) OR d) OR e) OR f) OR g) OR h) - UNCLEAR	**DEV**	**HIGH/LOW/ UNCLEAR**	**VAL**	**HIGH/LOW/ UNCLEAR**

TEST ACCURACY DATA

If reported please complete the following 2 x 2 contingency table. For studies investigating the accuracy of more than one index test or testing strategy please complete a 2 x 2 table for each test/testing strategy (cut and paste table as necessary). Imaging test results will be dichotomous.

LOWEST LEVEL OF AGGREGATION:

Fill in data as available.

	REFERENCE STANDARD (ovarian cancer)	REFERENCE STANDARD (borderline)	REFERENCE STANDARD (benign)	
INDEX TEST/TESTING STRATEGY +ve for ovarian cancer				**TOTAL INDEX TEST +ve**
INDEX TEST/TESTING STRATEGY +ve for borderline				**TOTAL INDEX TEST +ve**
INDEX TEST/TESTING STRATEGY +ve for benign				**TOTAL INDEX TEST –ve**
	DISEASE +ve	**TOTAL borderline**	**DISEASE -ve**	**TOTAL 'N'**
Aggregation borderline +ve	**TOTAL DISEASE +ve**		**TOTAL DISEASE -ve**	**TOTAL 'N'**
Aggregation borderline –ve	**TOTAL DISEASE +ve**	**TOTAL DISEASE -ve**		**TOTAL 'N'**

INSERT ANOTHER MORE DETAILED TABLE WITH SUB-CATEGORIES OF OVARIAN CANCER FOR LOW GRADE AND HIGH GRADE, TYPE 1 AND TYPE 2, EARLY-STAGE AND LATE-STAGE

	REFERENCE STANDARD (early-stage)	REFERENCE STANDARD (advanced-stage)	
INDEX TEST/TESTING STRATEGY +ve (early-stage)			**TOTAL INDEX TEST +ve**
INDEX TEST/TESTING STRATEGY +ve (late-stage)			**TOTAL INDEX TEST –ve**
	DISEASE +ve	**DISEASE -ve**	**TOTAL 'N'**

	REFERENCE (Type 1)	STANDARD	REFERENCE (Type 2)	STANDARD	
INDEX TEST/TESTING STRATEGY +ve (Type 1)					**TOTAL INDEX TEST +ve**
INDEX TEST/TESTING STRATEGY +ve (Type 2)					**TOTAL INDEX TEST -ve**
	DISEASE +ve		**DISEASE -ve**		**TOTAL 'N'**

Appendix 3 **: QUADAS-2.**

PATIENT SELECTION RISK OF BIAS

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 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 co-morbidities such as infertility and endometriosis	
Could the selection of patients have introduced bias?	**RISK: LOW/HIGH/UNCLEAR**
If a) and b) and c) 'YES' = low risk of bias	
If a) or b) or c) 'No' = high risk of bias	
If a) or b) or c) 'Unclear' = unclear risk of bias	

B. Concerns regarding applicability	
Describe included patients (prior testing, presentation, intended use of index test and setting):	
Is there concern that the included patients do not match the review question?	**CONCERN: LOW/HIGH/UNCLEAR**
a) Patients all symptomatic OR symptomatic and asymptomatic can be disaggregated	**Low - a) and b) and C) Yes**
b) Prior tests primary care: self reported symptoms	**High - a) or b) or C) No**
c) Prior tests secondary care: self reported symptoms OR self reported symptoms PLUS one or more of biochemical markers and ultrasound	**Unclear - a) or b) or C) Unclear**

INDEX TEST(S)

(If more than one index test was used, please complete for each test).

A1. Risk of bias (symptoms)

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 prespecified?	Yes/No/Unclear
Could the conduct or interpretation of the index test have introduced bias?	**RISK: LOW/HIGH/UNCLEAR** **High - a) or b) No** **Low - a) and b) Yes** **Unclear - a) or b) Unclear**

A2. Risk of bias (ultrasound)

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 prespecified?

Yes/No/Unclear

Could the conduct or interpretation of the index test have introduced bias?

RISK: LOW/HIGH/UNCLEAR

High - a) or b) No

Low - a) and b) Yes

Unclear - a) or b) Unclear

A3. Risk of bias (biomarkers) rule different because objective test in comparison to US and symptom elicitation

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 prespecified?	Yes/No/Unclear
Could the conduct or interpretation of the index test have introduced bias?	**RISK: LOW/HIGH/UNCLEAR** **High - a) or b) No** **Low - a) and b) Yes** **Unclear - a) or b) Unclear**

B. Concerns regarding applicability	
a) Is the skill of person performing US and eliciting symptoms detailed? (level of training and/or experience)	**Yes/No/Unclear/NA**
b) Was US performed in all patients by non-specialised sonographers	**Yes/No/Unclear**
c) Was US 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?	**CONCERN: LOW/HIGH/UNCLEAR** **High - a) or b) or C)No** **Low - a) and b) or c)Yes** **Unclear - a) or b) Unclear**

REFERENCE STANDARD RISK OF BIAS

A. Risk of bias	
Describe the reference standard and how it was conducted and interpreted:	
a) Is the reference standard likely to correctly classify the target condition? Index test +ve: Histology following laparoscopy or laparotomy	**Yes/No/Unclear**
b) Is the reference standard likely to correctly classify the target condition? Index test –ve: Yes - if a minimum follow-up period of greater than 12 months is included as required to assess whether the target condition is present No - if a minimum 12-month follow-up period is absent Unclear - if no information on follow-up period is included	**Yes/No/Unclear**
Could the reference standard, its conduct or its interpretation have introduced bias	**RISK: LOW/HIGH/UNCLEAR** **High - a) or b) No** **Low - a) and b) Yes** **Unclear - a) or b) Unclear**

DOMAIN 3: REFERENCE STANDARD (continued)

B. Concerns regarding applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN: Yes/No/Unclear**
Yes - ovarian cancer, borderline and metastatic disease are not differentiated (and cannot be for analysis)	
No - ovarian cancer, borderline and metastatic disease can be differentiated for analysis	
Unclear - unclear if ovarian cancer, borderline and metastatic disease have been disaggregated	

DOMAIN 4: 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 x 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 conduct or interpretation of reference standard have introduced bias?	**RISK: LOW/HIGH/UNCLEAR** **LOW - a) and b) and c) and d) - Yes** **HIGH - a) and b) and c) and d) - No** **UNCLEAR - a) and b) and c) and d) – Unclear**

COMPARATIVE DOMAIN (if applicable)

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? LOW - a) OR (b) and c))- Yes HIGH - a) OR (b) and c)) - No UNCLEAR - a) OR (b) or c)) – Unclear	**RISK: LOW/HIGH/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 *Low - No* *High - Yes* *Unclear - Unclear*	**CONCERN: LOW/HIGH/UNCLEAR**

RISK OF BIAS FOR MULTIVARIABLE DIAGNOSTIC MODELLING STUDIES (if applicable)

1. Participant selection	DEV	*Yes/No/Unclear*	VAL	*Yes/No/Unclear*
a) Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	**DEV**	*Yes/No/Unclear*		
b) Were participants enrolled at a similar state of health, or were predictors considered to account for differences?	**DEV**	*Yes/No/Unclear*	**VAL**	*Yes/No/Unclear*
Could the selection of patients have introduced bias? **HIGH**: a) OR a) and b) - YES **LOW**: a) OR a) and b) - NO **UNCLEAR**: a) OR a) and b) – UNCLEAR		**HIGH/LOW/ UNCLEAR**		**HIGH/LOW/ UNCLEAR**
3. Predictors	**DEV**	***Yes/No/Unclear***	**VAL**	***Yes/No/Unclear***
a) Were predictors defined and assessed in a similar way for all participants?	**DEV**	*Yes/No/Unclear*		
b) Are all predictors available at the time the model is intended to be used	**DEV**	*Yes/No/Unclear*	**VAL**	*Yes/No/Unclear*
c) Were all relevant predictors analysed?: ***No*** if symptoms only; ***No*** if US index test only; ***No*** if combination of index tests (symptoms, US and biomarkers) but miss out US OR Symptom OR FDA approved biomarkers	**DEV**	*Yes/No/Unclear*	**VAL**	*Yes/No/Unclear*
Could the definition, measurement or analysis of predictors introduced bias? **HIGH**: a) OR b) OR c) - YES **LOW**: a) OR b) OR c) - NO **UNCLEAR**: a) OR b) OR c) – UNCLEAR	**DEV**	**HIGH/LOW/ UNCLEAR**	**VAL**	**HIGH/LOW/ UNCLEAR**
3. ANALYSIS	**DEV**	***Yes/No/Unclear***	**VAL**	***Yes/No/Unclear***
a) Were there a reasonable number of outcome events?	**DEV**	*Yes/No/Unclear*		
b) Were there a reasonable number of outcome events?			**VAL**	*Yes/No/Unclear*
c) Were non-binary predictors handled appropriately?	**DEV**	*Yes/No/Unclear*	**VAL**	*Yes/No/Unclear*
d) Was selection of predictors based on univariable analysis avoided?	**DEV**	*Yes/No/Unclear*		
e) Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	**DEV**	*Yes/No/Unclear*		
f) For the model or any simplified score, were relevant performance measures evaluated, e.g. calibration, discrimination, (re)classification and net benefit?	**DEV**	*Yes/No/Unclear*	**VAL**	*Yes/No/Unclear*
g) Was the model recalibrated or was it likely (based on the evidence presented, e.g. calibration plot) that recalibration was not needed?	**DEV**	*Yes/No/Unclear*	**VAL**	*Yes/No/Unclear*
h) Was model validation undertaken in individuals other than those in the model development (external validation)?			**VAL**	*Yes/No/Unclear*
Could the analysis strategy have introduced bias? **HIGH**: a) OR b) OR c) OR d) OR e) OR f) OR g) OR h) - YES **LOW**: a) OR b) OR c) OR d) OR e) OR f) OR g) OR h) - NO **UNCLEAR**: a) OR b) OR c) OR d) OR e) OR f) OR g) OR h) – UNCLEAR	**DEV**	**HIGH/LOW/ UNCLEAR**	**VAL**	**HIGH/LOW/ UNCLEAR**

Appendix 4 **: Review reporting: Title, Introduction and Methods.**

Author and publication date	Title	Introduction			Methods (Yes-Y, No-N, Limited-L)									
		Rationale	**Clinical Back**	**Obj**	**Eligibility criteria**	**Info sources**	**Search strategy**	**Study selection**	**Data collection**	**Data extraction definitions**	**RoB & Appl**	**Accuracy measures (all per patient)**	**Planned Synthesis**	**Meta-analysis**
E.R. Myers / Feb 2006	N	Y	Y	Y	L	L	Y	Y	L	L	Y	Y	L	Y
P. Geomini /Feb 2009	Y	Y	L	L	L	L	L	Y	N	L	Y	Y	Y	Y
L.R. Medeiros/ Feb 2009	Y	Y	N	Y	L	L	L	Y	L	L	Y	Y	L	Y
NICE April /2011 (a)sym	N	Y	L	Y	L	L	Y	Y	L	L	Y	Y	L	N
NICE April /2011 (b) bio	N	Y	L	Y	Y	L	Y	Y	L	L	Y	Y	N	Y
NICE April /2011 (c)combi	N	Y	L	Y	L	L	Y	N (meta analysis selected 1 SR by Geomini and 1 additional updated study	Summary of Geomini; No meta analyses	Summary Limited	N	N	N	N
S.Yu/ Feb 2012	N (only meta analysis in title)	Y	N	L	L	L	L	Y	L	L	Y (but not reported in results)	Y	N	Y
J.E. Dodge/ Apr 2012	Y	Y	N	L	L	L	L	Y	N	L	Y	L	N	L
F. Li / June 2012	Y	Y	N	L	L	Y	L	Y	L	L	Y	Y	N	Y
L. Wu/ Sept 2012	Y	Y	N	L	L	L	L	Y	N	L	Y	Y	Y	Y
J. Lin/ Nov 2012	Y	Y	N	L	L	L	L	Y	L	N	L	Y	N	L
J. Lin/ Dec 2012	Y	Y	N	L	L	Y	N	L	L	L	L	Y	N	L
S.Ferraro/ Feb 2013	Y	Y	N	L	L	L	L	Y	L	L	Y	Y	N	L
Z. Yang/ Jul 2013	N (only meta analysis)	Y	N	L	L	Y	L	Y	N	L	Y	Y	N	L
J. Lin/ Dec 2013	Y	Y	N	L	L	L	L	Y	L	N	L	Y	N	L
J. Kaijser / Dec 2013	Y	Y	Y	L	L	Y	L	Y	L	Y	Y	L	L	Y
J.Wang/ Mar 2014	N(only metanalysis	Y	N	L	L	L	L	Y	L	L	Y	Y	L	Y
S. Zhen /Apr 2014	No (only metanalysis	Y	N	L	L	L	L	L	N	L	Y	Y	L	L
X.Y. Liao / Nov 2014	No (only metanalysis)	Y	N	L	L	L	L	Y	N	L	N	Y	N	L
A.C.L. Macedo / Sept 2014	Y	Y	N	L	Ls(all but Study design reported)	L	L	Y	L	L	Y	Y	N	Y
M. Stukan / Feb 2015	N	Y	L	L	L	L	L	N	N	L	Y	N	N	N

M.Ebell / Mar 2016	Y	Y	N	L	L	L	L	L	L	L	Y	Y	N	L
E.M.J. Meys/ May 2016	Y	Y	N	L	L	Y	L	Y	L	Y	Y	L	L	Y

Notes:

Title: Title acknowledges report as a SR; Objectives: (Y, N, Limited) was based on all, none or some of the following criteria being reported: Population, Index tests, target condition; etc

Appendix 5 **: Review reporting: Results and Discussion.**

Author and publication date	Title	Results (Yes-Y, No-N, Limited- L)						Discussion			Review funding
		Study selection	**Study charac**	**RoB & Appl**	**Results individual studies**	**Synthesis across studies**	**Additional analysis**	**Summary of evidence**	**Limitations**	**Conclusions**	
E.R. Myers / Feb 2006	N	L	Y	L	L	Y	Y,	Y	Y	Y	Y
P. Geomini / Feb 2009	Y	Y	L	L	N	Y	Y	L	N	L	N
L.R.Medeiros/ Feb 2009	Y	Y	L	Y	Y	Y	Y	Y	Y	N	N
NICE April /2011 (a)Symptoms	N	L	L	L	Y (only forest plot)	Y (as range)	N	L	Y	L	Y
NICE April /2011 (b)Biomarkers	N	L	Y	Y	Y	Y	N	Y	Y	L	Y
NICE April /2011 Combi	N	Summary of Geomini		N							Y
S. Yu / 2012	N	L	L	N (results not reported)	Y	Y	L	L	N	L	Y
J. E. Dodge/ Apr 2012	Y	N	L	N	L	L	N	Y	Y	L	N
F. Li / June 2012	Y	Y	L	Y	L	Y	L	Y	Y	N	N
L.Wu/ Sept 2012	Y	L	L	Y	L	Y	L	Y	L	Y	N
J.Lin/ Nov 2012	Y	Y	L	N	L	Y	N	L	L	N	N
J.Lin/ Dec 2012 (details for Index test only reported for HE4)	Y	N	Yes for HE4, No for CA125 and ROMA	N	Yes only for HE4, No for CA125 and ROMA	Yes for all	Limited for HE4, No for CA125 and ROMA	L	L	L	Y
S.Ferraro/ Feb 2013	Y	L	L	Y	L	Y	N	Y	Y	Y	N
Z. Yang/ Jul 2013	N	Y	N	Y	L	Y	N	L	Y	L	Y

J.Lin/ Dec 2013 (details for Index test only reported for HE4)	Y	Y	Yes for HE4, No for CA125 and ROMA	N	Yes only for HE4, No for CA125 and ROMA	Yes for all	Limited for HE4, No for CA125 and ROMA	L	L	L	Y
J. Kaijser / Dec 2013	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y
J.Wang / Mar 2014	N	Y	L	N	L	Y	L	L	Y	L	N
S. Zhen /Apr 2014	N	L	L	L	L	Y	L	Y	L	L	Y
X. Y. Liao / Nov 2014	N	L	L	N	L	Y	N	L	N	N	N
A.C.L. Macedo / Sept 2014	Y	Y	L	Y	Y	Y	L	Y	L	L	N
M. Stukan / Feb 2015	N	L	L	L	L	NA	NA	L	N	L	N
M.Ebell / Mar 2016	Y	Y	L	L	L	Y	N	Y	L	L	N
E.M.J. Meys/ May 2016	Y	L	L	Y	L	Y	L	Y	Y	Y	Y

Appendix 6 : List of excluded studies with reasons for exclusions.

List of excluded studies with reasons for exclusions (Reasons in order of frequency of studies excluded) (737 studies excluded)

No 2x2 table (N=253 studies excluded)

1. Ovarian cancer: screening, treatment, and followup. NIH Consensus Statement. 1994;12(3):1-30.

2. Ovarian cancer. Journal of Practical Nursing. 2009;59(3):12-4.

3. CA-125 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 grey-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.

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.

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.

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. 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.

67. 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.

68. Dorigo O, Berek JS. Personalizing CA125 levels for ovarian cancer screening. Cancer Prevention Research. 2011;4(9):1356-9.

69. 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.

70. Drake J. Diagnosis and management of the adnexal mass. American Family Physician. 1998;57(10):2471-6, 9-80.

71. 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.

72. 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.

73. Dutta S, Wang F-q, Phalen A, Fishman DA. Biomarkers for ovarian cancer detection and therapy. Cancer Biology & Therapy. 2010;9(9):668-77.

74. 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.

75. Farghaly SA. Current diagnosis and management of ovarian cysts. Clinical & Experimental Obstetrics & Gynecology. 2014;41(6):609-12.

76. 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.

77. 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.

78. Fischerova D, Burgetova A. Imaging techniques for the evaluation of ovarian cancer. Best Practice & Research in Clinical Obstetrics & Gynaecology. 2014;28(5):697-720.

79. Fischerova D, Zikan M, Dundr P, Cibula D. Diagnosis, treatment, and follow-up of borderline ovarian tumors. Oncologist. 2012;17(12):1515-33.

80. 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.

81. 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.

82. 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.

83. 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.

84. 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.

85. 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.

86. Fuste P. HE-4 in the diagnosis of abdominal masses. Comparison with ultrasonography and RMI. Tumor Biology. 2014;35:S9.

87. 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.

88. Geetha P, Nair MK. Granulosa cell tumours of the ovary. Australian & New Zealand Journal of Obstetrics & Gynaecology. 2010;50(3):216-20.

89. 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.

90. 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.

91. 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.

92. Goff B. Symptoms Associated With Ovarian Cancer. Clinical Obstetrics and Gynecology. 2012;55(1):36-42.

93. 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.

94. 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.

95. 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.

96. 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.

97. Griffin N, Grant LA, Sala E. Adnexal masses: characterization and imaging strategies. Seminars in Ultrasound, CT & MR. 2010;31(5):330-46.

98. 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.

99. 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.

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. 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.

117. 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.

118. 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.

119. 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.

120. 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.

121. 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.

122. 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.

123. 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.

124. 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.

125. 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.

126. 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.

127. 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.

128. 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.

129. Kim PS, Djazayeri S, Zeineldin R. Novel nanotechnology approaches to diagnosis and therapy of ovarian cancer. Gynecologic Oncology. 2011;120(3):393-403.

130. 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.

131. 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.

132. 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.

133. Konforte D, Diamandis EP. Is early detection of cancer with circulating biomarkers feasible? Clinical Chemistry. 2013;59(1):35-7.

134. 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.

135. 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.

136. 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.

137. 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.

138. Lalwani N, Dubinsky TJ. Clinically suspected adnexal mass. Ultrasound Quarterly. 2013;29(1):87-8.

139. 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.

140. 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.

141. 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.

142. 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.

143. Leeners B, Schild RL, Funk A, Hauptmann S, Kemp B, Schroder W, et al. Colour Doppler sonography improves the preoperative diagnosis of ovarian tumours made using conventional transvaginal sonography. European Journal of Obstetrics Gynecology and Reproductive Biology. 1996;64(1):79-85.

144. 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.

145. 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.

146. 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.

147. 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.

148. 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.

149. 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.

150. Liu JH, Zanotti KM. Management of the adnexal mass. Obstetrics & Gynecology. 2011;117(6):1413-28.

151. 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.

152. 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.

153. 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).

154. 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.

155. Manolov V, Avramova V, Radenkov R. He 4 - A new marker for ovarian cancer. Clinical Chemistry and Laboratory Medicine. 2011;49:S254.

156. 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.

157. 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.

158. 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.

159. 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.

160. McGuire WP. HE4--another marker for gynecologic cancers: do we really need one? Oncology (Williston Park). 2013;27(6):563, 6.

161. 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.

162. 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.

163. 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.

164. Miller RW, Ueland FR. Risk of malignancy in sonographically confirmed ovarian tumors. Clinical Obstetrics & Gynecology. 2012;55(1):52-64.

165. Miller RW, van Nagell JR. Preoperative evaluation of adnexal masses. Women's health. 2011;7(1):37-9.

166. 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.

167. 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.

168. 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.

169. 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.

170. 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.

171. 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.

172. 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.

173. 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).

174. 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.

175. 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.

176. Moore RG, Maclaughlan S. Current clinical use of biomarkers for epithelial ovarian cancer. Current Opinion in Oncology. 2010;22(5):492-7.

177. 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.

178. 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.

179. 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.

180. 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.

181. 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.

182. 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.

183. Murta EFC, Nomelini RS. Early diagnosis and predictors of malignancy of adnexal masses. Current Opinion in Obstetrics & Gynecology. 2006;18(1):14-9.

184. 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.

185. 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.

186. 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.

187. 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.

188. Nogales FF, Dulcey I, Preda O. Germ cell tumors of the ovary: an update. Archives of Pathology & Laboratory Medicine. 2014;138(3):351-62.

189. Nolen BM, Lokshin AE. Multianalyte assay systems in the differential diagnosis of ovarian cancer. Expert Opinion on Medical Diagnostics. 2012;6(2):131-8.

190. Nolen BM, Lokshin AE. Biomarker testing for ovarian cancer: clinical utility of multiplex assays. Molecular Diagnosis & Therapy. 2013;17(3):139-46.

191. 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.

192. 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.

193. 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.

194. Osmers RGW, Kuhn W. Early diagnosis of ovarian cancer using ultrasonography. Archives of Gynecology and Obstetrics. 1997;260(1-4):351-5.

195. Ozan H, Ozsoy S, Baykara S, Celik N. Review of 28 ovarian granulosa cell tumor cases. International Journal of Gynecological Cancer. 2011;3):S657.

196. 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.

197. 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.

198. 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]

Moznosti poedoperaeni predikce biologickeho chovani gynekologickych panevnich tumoru. Aktualni Gynekologie a Porodnictvi. 2014;6:53-9.

199. Prompeler HJ. [Ultrasonographic clarification of adnexal findings]. Radiologe. 2011;51(7):568-80.

200. 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.

201. Redman C, Duffy S, Dobson C. Improving early detection of ovarian cancer. Practitioner. 2011;255(1741):27-30, 3.

202. 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.

203. 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.

204. 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.

205. 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.

206. 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.

207. 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.

208. Sasaroli D, Coukos G, Scholler N. Beyond CA125: The coming of age of ovarian cancer biomarkers. Biomarkers in Medicine. 2009;3(3):275-88.

209. 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.

210. 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.

211. 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.

212. 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.

213. 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.

214. 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.

215. 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.

216. 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.

217. 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.

218. 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.

219. 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.

220. 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).

221. 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.

222. 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.

223. 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.

224. Su Z, Graybill WS, Zhu Y. Detection and monitoring of ovarian cancer. Clinica Chimica Acta. 2013;415:341-5.

225. 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.

226. Sumpaico WW. Comparison of roma to rmi for ovarian carcinoma in asia. International Journal of Gynecology and Obstetrics. 2012;119:S248-S9.

227. 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.

228. 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.

229. 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.

230. 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.

231. Twickler DM, Moschos E. Ultrasound and assessment of ovarian cancer risk. AJR American Journal of Roentgenology. 2010;194(2):322-9.

232. 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.

233. 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.

234. 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.

235. 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.

236. 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.

237. van Nagell JR, Jr. Early detection of ovarian cancer in symptomatic women. Lancet Oncology. 2012;13(3):223-4.

238. 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.

239. 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.

240. Vranes HS, Klaric P, Sonicki Z, Gall V, Jukic M, Vukovic A. Prediction of ovarian tumor malignancy. Collegium Antropologicum. 2011;35(3):775-9.

241. Weinberger V, Minar L. Diagnostics of malign ovarian tumors by ultrosound and CA 125-our experience. International Journal of Gynecological Cancer. 2013;1:498.

242. 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.

243. 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.

244. 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.

245. 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.

246. 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.

247. 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.html

http://ac.els-cdn.com/S1470204508700056/1-s2.0-S1470204508700056-main.pdf?_tid=b21772ca-620a-11e5-8fbf-00000aacb362&acdnat=1443023558_910592ab86216e8621e934ba16a60306.

248. 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.

249. 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.

250. 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.

251. 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.

252. 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.

253. 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.

Population not applicable (N=115)

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.

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. Awwad Y. New screening strategy may catch ovarian cancer at early stages. Women's Health. 2013;9(6):509.

12. BA IBA. Symptoms attributed to ovarian cancer by women with the disease1997. 122, 7-8, 30 passim p.

13. 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.

14. 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.

15. 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.

16. 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.

17. 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.

18. 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.

19. 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.

20. 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.

21. 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.

22. 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.

23. 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.

24. 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.

25. 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.

26. 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.

27. 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.

28. 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.

29. 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.

30. 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.

31. 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.

32. 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.

33. 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.

34. 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.

35. 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.

36. 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.

37. 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.

38. 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.

39. 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.

40. Hellstrom I. The HE4 (WFDC2) Protein Is a Biomarker for Ovarian Carcinoma1. Cancer Research. 2003;63:3695-700.

41. 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.

42. 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.

43. 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.

44. 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.

45. 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.

46. 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.

47. 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.

48. 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.

49. Kalfa N, Sultan C. Juvenile ovarian granulosa cell tumor: a benign or malignant condition? Gynecological Endocrinology. 2009;25(5):299-302.

50. 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.

51. 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.

52. 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.

53. 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.

54. 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.

55. Koldjeski D, Kirkpatrick MK, Swanson M, Everett L, Brown S. Ovarian cancer: early symptom patterns. Oncology nursing forum. 2003;30(6):927-33.

56. 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.

57. Kurjak A, Zalud I, Alfirevic Z. EVALUATION OF ADNEXAL MASSES WITH TRANSVAGINAL COLOR ULTRASOUND. Journal of Ultrasound in Medicine. 1991;10(6):295-7.

58. 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.

59. 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.

60. 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.

61. 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.

62. 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.

63. Menon U, Jacobs I. Ovarian cancer screening in the general population. Ultrasound in Obstetrics and Gynecology. 2000;15(5):350-3.

64. 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.

65. 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.

66. 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.

67. 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.

68. 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.

69. 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.

70. 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.

71. 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.

72. 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).

73. 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.

74. 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.

75. 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.

76. 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.

77. 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.

78. 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.

79. 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.

80. 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.

81. 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.

82. 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.

83. 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.

84. 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.

85. 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.

86. 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.

87. 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.

88. 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.

89. 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.

90. Sorensen SS, Mosgaard BJ. Combination of cancer antigen 125 and carcinoembryonic antigen can improve ovarian cancer diagnosis. Danish Medical Bulletin. 2011;58(11):A4331.

91. 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.

92. 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.

93. 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.

94. 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.

95. 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.

96. 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.

97. 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.

98. 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.

99. 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.

100. 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.

101. 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.

102. 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.

103. 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.

104. 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.

105. 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.

106. 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.

107. 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.

108. 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.

109. 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.

110. Wang S, Johnson S. Prediction of benignity of solid adnexal masses. Archives of Gynecology & Obstetrics. 2012;285(3):721-6.

111. 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.

112. 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.

113. 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.

114. 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.

115. 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.

Index test not applicable (N=98)

1. 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.

2. 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.

3. 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.

4. 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.

5. 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.

6. 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.

7. 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.

8. 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.

9. 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.

10. 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.

11. 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.

12. 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.

13. 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.

14. Barroilhet L, Vitonis A, Shipp T, Muto M, Benacerraf B. Sonographic predictors of ovarian malignancy. Journal of Clinical Ultrasound. 2013;41(5):269-74.

15. 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.

16. 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.

17. 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.

18. 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.

19. 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.

20. 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.

21. 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.

22. 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.

23. 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.

24. 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.

25. 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.

26. 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.

27. 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.

28. 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.

29. 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.

30. 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.

31. 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.

32. Goff BA, Mandel L, Muntz HG, Melancon CH. Ovarian carcinoma diagnosis. Cancer. 2000;89(10):2068-75.

33. 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.

34. 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.

35. 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.

36. 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.

37. 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.

38. 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.

39. 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.

40. 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.

41. 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.

42. 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.

43. 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.

44. 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.

45. 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.

46. 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.

47. 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.

48. 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.

49. 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.

50. 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.

51. 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.

52. 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.

53. 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.

54. 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.

55. 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.

56. 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.

57. 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.

58. 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.

59. 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.

60. Prompeler HJ, Madjar H, Sauerbrei W. Classification of adnexal tumors by transvaginal color Doppler. Gynecologic Oncology. 1996;61(3):354-63.

61. 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.

62. 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.

63. 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.

64. 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.

65. 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.

66. 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.

67. 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.

68. 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.

69. 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.

70. 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.

71. 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.

72. 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.

73. 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.

74. 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.

75. 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.

76. 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.

77. 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.

78. 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.

79. 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.

80. 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.

81. 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.

82. 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.

83. 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.

84. 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.

85. 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.

86. 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.

87. 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.

88. 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.

89. 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.

90. 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.

91. 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.

92. 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.

93. 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.

94. 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.

95. 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.

96. 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.

97. 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.

98. 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.

Not a test accuracy study (N=95)

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.

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.

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.

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.

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.

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.

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 (N=72)

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.

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.

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.

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.

Conference abstracts/posters (N=47 studies excluded)

1. Acharya UR, Sree VS, Saba L, Molinari F, Guerriero S, Suri JS. Ovarian tumor characterization and classification: a class of GyneScanTM systems. Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine & Biology Society. 2012;2012:4446-9.

2. Adams M, Gore M, Sporn T, Horlick E, King L, Avissar PA, et al. The detection and correlation of HE4 mRNA and protein expression in all major histologic subtypes of ovarian cancer. Histopathology. 2008;53:176.

3. Alhudiri IM, Murray A, Williamson K, Nunns D, O'Brien TJ, Robertson JF, et al. Autoantibodies in ovarian cancer. Tumor Biology. 2011;32:S76-S7.

4. Allard J, Somers E, Theil R, Moore RG. Use of a novel biomarker HE4 for monitoring patients with epithelial ovarian cancer. Journal of Clinical Oncology. 2008;26(15).

5. Bahrani-Mostafavi Z, Yeganeh PN, Parrott ME, Richardson C, Tait DL, Mostafavi MT. PAX8 protein detection in serum of patients with serous ovarian cancer. Cancer Research. 2014;1).

6. Baker C, Pasipanodya T, Dwivedi R. The management of suspected ovarian masses in premenopausal women in a DGH setting. BJOG: An International Journal of Obstetrics and Gynaecology. 2013;120:372.

7. Bassiouny D, Naeim M, Ismiil N, Dube V, Han G, Cesari M, et al. Primary versus metastatic ovarian tumors: Challenging the validity of common diagnostic features. Laboratory Investigation. 2015;95:275A.

8. Bayani J, Kuzmanov U, Batruch I, Cho CK, Smith C, Marrano P, et al. Integrated Genomic, MicroRNA (miRNA) and Proteomic Profiling of Ovarian Carcinoma for Biomarker Discovery. Cancer Research. 2010;1).

9. Brown A, Miller C, Robison K, Somers E, Allard J, Granai CO, et al. Differential expression of CA125 and a novel serum tumor marker HE4 in epithelial ovarian cancer. Journal of Clinical Oncology. 2008;26(15).

10. Buda F, Lugano A, Cicchetti M, Buda C, Messina G. CLINICAL EVALUATION OF THE UTILITY OF COMBINED USE OF SEVERAL TUMOR MARKERS IN PATIENTS WITH OVARIAN EPITHELIAL CANCER AND BENIGN GYNECOLOGICAL DISEASE. Tumor Biology. 2010;31:S113-S.

11. Casova M, Novotny Z, Topolcan O, Vrzalova J, Holubec L, Finek J, et al. HE4: AN APPROPRIATE MARKER FOR THE PRIMARY DIAGNOSIS AND FOLLOW-UP OF OVARIAN CANCER. Anticancer Research. 2008;28(6B):4028-9.

12. Casova M, Novotny Z, Vrzalova J, Topolcan O, Holubec L, Prazakova M, et al. HE4-A NEW MARKER FOR OVARIAN CANCER IN ROUTINE? Anticancer Research. 2008;28(5C):3232-.

13. Dane C, Bilgili N. Preoperative evaluation of borderline ovarian tumors by risk of malignancy index, CA125 and ultrasound. International Journal of Gynecological Cancer. 2014;2).

14. Di Legge A, Fagotti A, De Blasis I, Virgilio BA, Mascilini F, Moruzzi MC, et al. The sonographic prediction of invasive carcinoma in unilocular-solid ovarian cysts in premenopausal patients: A pilot study. International Journal of Gynecology and Obstetrics. 2012;119:S741.

15. Dorffler P, Caffier H. CA 125 and tissue polypeptide special antigen (TPS) as tumor markers in ovarian cancer: Additional information by TPS?. [German] Ca 125 Und Tps Als Tumormarker Beim Ovarialkarzinom: Welche Zusatzinformationen Bietet Tps? Archives of Gynecology and Obstetrics. 1993;254(1-4):957-9.

16. Duffy MJ, Bonfrer JM, Kulpa J, Rustin GJS, Soletormos G, Torre GC, et al. CA 125 in ovarian cancer: European Group on Tumor Marker (EGTM) guidelines for clinical use. Tumor Biology. 2006;27.

17. Gashenko E, Lebedeva V, Tsykalenko E, Russkikh G, Loktev K, Brak I, et al. Procathepsin B and endogenous inhibitors of cysteine proteases as possible biomarkers of ovarian cancer. European Journal of Cancer. 2014;50:S61.

18. Granatova J, Boiecka K, Tma J, Bednaiova K, Vranova J, Donatova Z. He4: A novel tumor marker in differential diagnostics of ovarian cancer. Clinical Chemistry and Laboratory Medicine. 2011;49:S240.

19. Gungor T, Baser E, Ozer I, Togrul C, Uzunlar O, Sen Ozyer S. Clinical analysis of ovarian granulosa cell tumor cases over a five year period: A single center experience in nkara, turkey. International Journal of Gynecological Cancer. 2012;22:E347.

20. Kaijser J, Van Gorp T, Van Holsbeke C, Sayasneh A, Vergote I, Bourne T, et al. IOTA simple descriptors (SD) or simple rules (SR) as a triage test in patients with ovarian tumours: Subsequent value of CA125, HE4 or ROMA in clinical reality? BJOG: An International Journal of Obstetrics and Gynaecology. 2013;120:371.

21. Koleva V, Puncheva E, Karagyozov I. Can HE4 and roma score be more helpful than CA125 alone as diagnostic tools in patients with pelvic mass? Biochimica Clinica. 2013;37:S175.

22. Kong B, Zhang Q, Hu G, Yang Q, Dong R, Xie X, et al. A multiplex methylation-specific PCR assay for detection of early-stage ovarian cancer using cell-free serum DNA. Journal of Clinical Oncology. 2013;1).

23. Lawicki S, Szmitkowski M, Gacuta-Szumarska E, Bedkowska E. Plasma levels and diagnostic utility of metalloproteinase 9 (MMP-9) in ovarian cancer patients. Biochimica Clinica. 2013;37:S177.

24. Liro M, Sniadecki M, Ziemann B, Wycinka E, Wydra D. Comparison between scoring system and subjective evaluation of adnexal masses by two independent ultrasound examiners with a different level of expertise: Training aspects. International Journal of Gynecological Cancer. 2013;1):1168.

25. Luciano DE, Hoffman J, Zhou XC, Spiel M, Luciano AA, Exacoustos C. Does three-dimensional sonography improve the diagnostic accuracy of ovarian tumors with previous inconclusive imaging? Journal of Minimally Invasive Gynecology. 2011;1):S136.

26. Maggioli C, Vitale G, Li Bassi S, Bernardi M. An unusual case of ascites. Italian Journal of Medicine. 2014;8:76.

27. Manders DB, Purinton SC, Lea JS, Richardson DL, Miller DS, Kehoe SM. Predictive value of carcinoembryonic antigen and cancer antigen 125 in identifying mucinous ovarian tumors. Gynecologic Oncology. 2011;123 (2):445-6.

28. McLemore MR, Aouizerat B. CA125 assay performance across methodological types: a meta-analytic review. Communicating Nursing Research. 2005;38:366-.

29. Mohamadnejad M, Leblanc JK, Al-Haddad MA, Sherman S, McHenry L, Cote GA, et al. Utility of Endoscopic Ultrasound for Evaluation of Pelvic Masses. Gastrointestinal Endoscopy. 2010;71(5):AB313-AB.

30. Molina R, Escudero JM, Fuste P. HE-4 utility to increase efficiency in patients with abdominal masses. Tumor Biology. 2014;35:S6.

31. Moore EK, Iavazzo C, Argent V, Leung E, Pitkin S, Benton S, et al. Does the risk of malignancy algorithm have a role in triaging symptomatic women for further investigation? Results of a pilot 'real world' study. International Journal of Gynecological Cancer. 2013;1):64.

32. Moore R, Miller M, MacLaughlan S, Vucetic Z, Gajewski W, DiSilvestro P. The use of the Risk of Ovarian Malignancy Algorithm (ROMA) with clinical assessment improves ovarian cancer detection in women with a pelvic mass. Gynecologic Oncology. 2012;125:S38.

33. Narang L, Ofuasia E. A comparative study assessing diagnostic accuracy of pre-operative transvaginal ultrasound and laparoscopy in the management of patients with ovarian cysts. Gynecological Surgery. 2011;8:S40-S1.

34. Nilsson O, Allard WJ, Moore RG, Skates SJ. A BIOMARKER ALGORITHM OF HE4 AND CA125 FOR DIFFERENTIAL DIAGNOSIS IN WOMEN PRESENTING WITH ADNEXAL MASS. Anticancer Research. 2008;28(6B):4028-.

35. Oz M, Selcuk I, Ozdal B, Biberoglu E, Bas S, Meydanli M, et al. Clinicopathological analysis of adult granulosa cell tumors: A retrospective analysis of 40 patients. International Journal of Gynecological Cancer. 2014;4):313-4.

36. Pariyar J. The role of Risk of Malignancy Index (RMI) based on serum CA-125 level, ultrasound score and menopausal status in pre-operative evaluation of adnexal mass. Journal of Clinical Oncology. 2008;26(15).

37. Putri I, How JA, Marino J, Villegas R, McNally O, Grover S, et al. A 32 year review of clinical presentation and the use of risk of malignancy index (RMI2) in diagnosis of ovarian malignancies in children and adolescents. International Journal of Gynecological Cancer. 2014;4):211-2.

38. Qiu F, Tian Y. The Establishment of Serum Detection Model for Ovarian Cancer with MALDI-TOF MS. Clinical Chemistry. 2009;55(6):A112-A3.

39. Qu KZ, Li H, Whitmire R, Sferruzza A, Bender RA. Human epididymis secretory protein 4 in serum (HE4) as a marker of ovarian cancer. Journal of Clinical Oncology. 2008;26(15).

40. Rai AJ, Zhang Z, Rosenzweig J, Shih I, Sokoll LJ, Chan DW. SELDI (Surface Enhanced Laser Desorption/Ionization) based identification of tumor markers for ovarian cancer-combination of CA125 and multiple markers improves detection performance. Clinical Chemistry. 2002;48(6):A21-A2.

41. Ruiz Escalera JF, Martinez Diez M, Rodriguez Espinosa M, Casero Ariza C, Lillo Munoz JA, Espejo Martinez AB, et al. An evaluation of CA 125 and HE4 for the classification of ovarian masses. Tumor Biology. 2014;35:S38.

42. Santotoribio JD, Garcia-De La Torre A, Canavate-Solano C, Arce-Matute F, Perez-Ramos S. Ca19.9 in combination with ca125 for diagnosis of ovarian mucinous cancer. Tumor Biology. 2014;35:S38.

43. Spiel M, Luciano D, Hoffman J, Zhou X, Luciano A, Joyner A, et al. 3D ultrasound as a preoperative tool to characterize indeterminate adnexal masses. Gynecologic Oncology. 2013;131 (1):284.

44. Svobodova S, Kucera R, Topolcan O, Presl J, Fuchsova JR, Vrzalova, et al. Comparison of HE4 with other tumor markers in ovarian cancer diagnostics. Tumor Biology. 2012;33:S89.

45. Uyan Baba D, Ugur M, Gungor T, Halisdemir Gurlek B, Kale I, Var T, et al. Clinical importance of human epidydimis protein-4 (HE4) in predicting malignancy potentials of adnexial masses. International Journal of Gynecological Cancer. 2011;3):S845.

46. Yadav SK, D'Angelo A, Abdurazaq B, Amso NN. Morphological changes in conservatively managed benign ovarian cyst characterised by simple ultrasonography rules in asymptomatic postmenopausal women: A retrospective cohort study based on the United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) data. BJOG: An International Journal of Obstetrics and Gynaecology. 2014;121:5.

47. Yu X, Cohen J, Deshmukh H, Zhang R, Shin JY, Osann K, et al. The association of serial ultrasounds and CA125 prior to diagnosis of ovarian cancer-Do they improve early detection? Gynecologic Oncology. 2008;111(2):385-6.

No test positivity threshold reprted (N=16)

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. 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.

4. 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.

5. 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.

6. 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.

7. 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.

8. 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.

9. 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.

10. 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.

11. 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.

12. 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.

13. 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.

14. 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.

15. 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.

16. 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.

Publication pre 1991 (N=15)

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.

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.

Case control study with healthy controls (N=9)

1. 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.

2. 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.

3. 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.

4. 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.

5. 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.

6. 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.

7. 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.

8. 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.

9. 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.

Full text not available (N=7)

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.

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.

Duplicate data reporting (N=6)

1. 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.

2. 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.

3. 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.

4. 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.

5. 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.

6. 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.

Developmental model superseded by model validation (N=4)

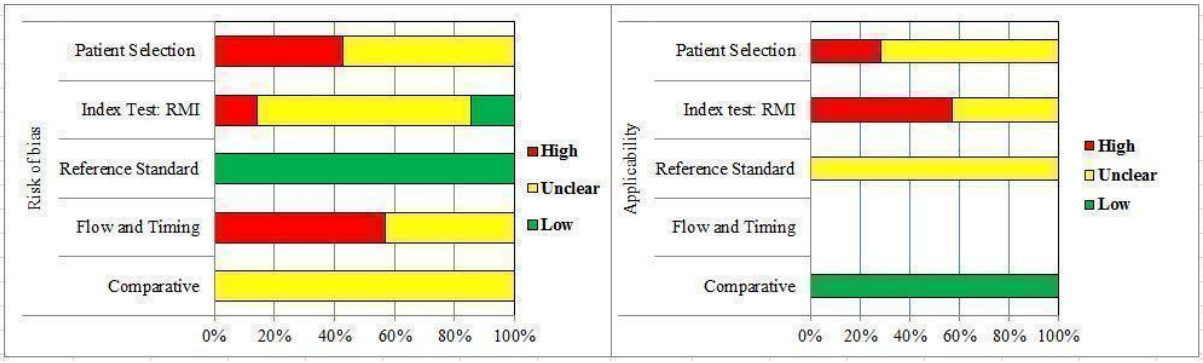
1. Hillaby K, Aslam N, Salim R, Lawrence A, Raju KS, Jurkovic D. The value of detection of normal ovarian tissue (the 'ovarian crescent sign') in the differential diagnosis of adnexal masses. Ultrasound in Obstetrics & Gynecology. 2004;23(1):63-7.

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.

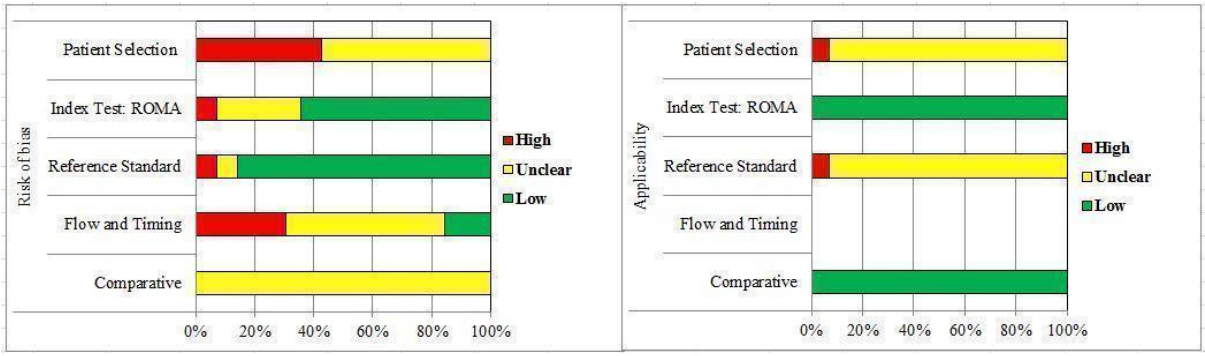
3. Timmerman D, Testa AC, Bourne T, Ameye L, Jurkovic D, Van Holsbeke C, et al. Simple ultrasound-based rules for the diagnosis of ovarian cancer. Ultrasound in Obstetrics & Gynecology. 2008;31(6):681-90.

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

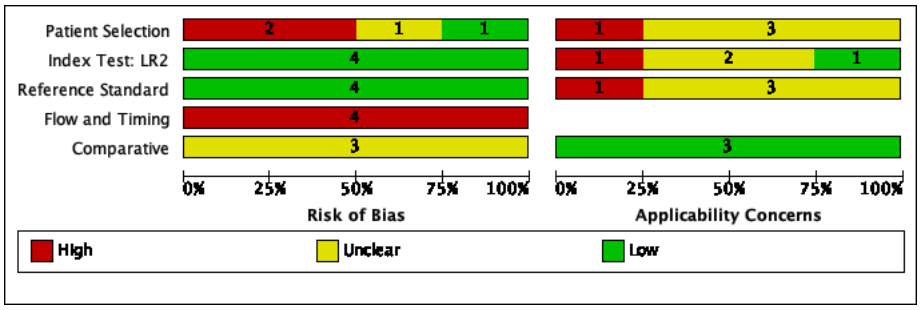
Appendix 7 **: Risk of bias and applicability concerns for RMI1 studies reporting results separately for pre and postmenopausal women.**



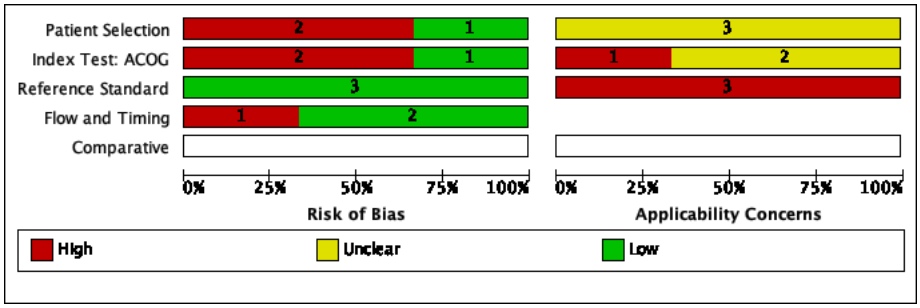
Appendix 8 **: Risk of bias and applicability concerns for ROMA studies reporting results separately for pre and postmenopausal women.**



Appendix 9 **: Risk of bias and applicability concerns graph for LR2 studies reporting results separately for pre and postmenopausal women.**



Appendix 10 **: Risk of bias and applicability concerns for ACOG v3 CA125 and ACOG v3 OVA1 studies reporting results separately for pre and postmenopausal women.**



Appendix 11 **: Comparison of test accuracy of LR2, RMI I, RMI III, ACOG V1 and ACOG V3 (CA125) at fixed thresholds between pre and postmenopausal women.**

Test	Studies N, D+	Analysis Method	Sensitivity	Specificity	Ratio of Sens odds (p value)	Ratio of Spec odds (p value)
LR2 pre	3, 2715, 588	Un	83.6 (78.6, 87.6)	90.4 (81.3, 95.3)	3.02 (2.04, 4.49) (P<0.0001)	0.21 (0.11, 0.43) (P<0.0001)
LR2 post	3, 1861, 1008	Un	93.9 (92.3, 95.3)	63.9 (60.6, 67.0)		
RMI1 200 pre	5, 2634, 502	Un	52.2 (45.9, 58.5)	95.4 (92.5, 97.3)	2.71 (1.92, 3.83) (P<0.0001)	0.42 (0.19, 0.94) (P=0.03)
RMI1 200 post	5, 1879, 862	Un	75.0 (69.5, 79.8)	90.1 (83.1, 94.4)		
RMI1 250 pre	1, 356, 31	Bp	54.8 (36.0, 72.7)	88.3 (84.3, 91.6)	3.88 (1.61, 9.34) (P=0.003)	0.52 (0.30, 0.90) (P=0.02)
RMI1 250 post	2, 220, 97	Un	82.5 (73.6, 88.8)	79.7 (71.6, 85.9)		
RMI3 200 pre	2, 606, 84	Un	73.6 (61.3, 83.1)	91.1 (87.3, 94.0)	2.15 (0.42, 10.9) (P=0.36)	0.15 (0.09, 0.23) (P<0.0001)
RMI3 200 post	2, 604, 314,	Un	85.8 (58.3, 96.3)	61.7 (56.0, 67.1)		
ACOG V1 pre	3, 716, 154	Un	69.4 (59.3, 78.0)	60.5 (36.1, 80.6)	4.64 (2.32, 9.27) (P=0.001)	0.73 (0.26, 2.10) (P=0.265)
ACOG V1 post	3, 1301, 585	Un	91.4 (86.3, 94.7)	53.0 (43.9, 61.9)		
ACOG V3 (CA125) pre	3, 899, 150	Un	78.0 (70.4, 84.1)	83.9 (65.1, 92.6)	1.54 (0.77, 3.12) (P=0.23)	0.76 (0.28, 2.04) (P=0.59)
ACOG V3 (CA125) post	3, 1224, 486	Un	84.6 (76.0, 90.5)	78.1 (71.0, 83.9)		

Notes to table: Analysis method: bp = single study estimate. uv=univariate analysis. bv=bivariate analysis. D+ = Primary ovarian cancer plus borderline ovarian tumours (where borderline tumours included in primary study analysis) minus metastatic disease (where this could be disaggregated from primary ovarian cancer in primary study analysis).

Appendix 12 **: Summary Estimates of Sensitivity and Specificity for all tests, all versions, all test positivity thresholds separately for premenopausal, postmenopausal and pre and postmenopausal mixed populations.**

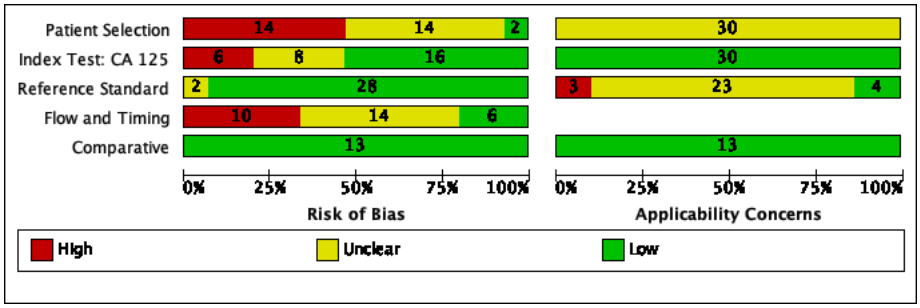
	PREMENOPAUSAL			POSTMENOPAUSAL			PRE AND POSTMENOPAUSAL COMB		
-Test Version -Threshold	**Studies** **N** **D+** **Analysis**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Studies** **N** **OC cases** **Analysis**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Studies** **N** **OC cases** **Analysis**	**Sensitivity (95% CI)**	**Specificity (95% CI)**
ADNEX 3% post test probability	1 1354 378 bp	97.9 (95.9-98.9	56.6 (53.4-59.6)	1 1049 602 bp	99.5 (98.6-99.8)	25.1 (21.3-29.3)	1 2403 980 bp	98.9 (97.8, 99.5)	46.6 (44.0, 49.2)
ADNEX 5% post test probability	1 1354 378 bp	97.6 (95.5-98.7)	69.5 (66.5-72.3)	1 1049 602 bp	98.8 (97.6-99.4)	37.4 (33.0-41.9)	1 2403 980 bp	98.4 (97.2, 9.2)	59.4 (56.8, 61.9)
ADNEX 10% post test probability	1 1354 378 bp	94.7 (92.0-96.6)	78.6 (75.9-81.1)	1 1049 602 bp	97.7 (96.1-98.6)	55.5 (50.9-60.0)	1 2403 980 bp	96.4 (94.8, 97.7)	71.3 (68.9, 73.7)
ADNEX 15% post test probability	1 1354 378 bp	90.5 (87.1-93.0)	83.4 (80.9-85.6)	1 1049 602 bp	96.5 (94.7-97.7)	63.5 (59.0-67.9)	1 2403 980 bp	94.2 (92.1, 95.8)	77.2 (95.0, 79.4)
LR2 10% post test probability	3 2715 558 un	83.6 (78.6, 87.6)	90.4 (81.3, 95.3)	3 1861 1008 un	93.9 (92.3, 95.3)	63.9 (60.6, 67.0)	4 4888 1728 Un	90.3 (87.5, 92.6)	81.5 (74.7, 86.8)
RMI1 200	5 2634 502	52.2 (45.9, 58.5)	95.4 (92.5, 97.3)	5 1879 862 un	75.0 (69.5, 79.8)	90.1 (83.1, 94.4)	31 12335 3688 bv	73.1 (68.6, 77.2)	91.9 (89.4, 93.8)
Appendix 12 Continued...									
250	356 31	(36.0, 72.7)	(84.3, 91.6)	2 220 97	82.5 (73.6, 88.8)	79.7 (71.6, 85.9)	14 4818 1452	71.5 (67.7, 75.1)	92.2 (88.4, 94.9)

	bp			un			bv		
RMI2 200	-	-	-	-	-	-	19 4989 1328 Bv	76.7 (69.9, 82.4)	85.2 (81.6, 88.2)
RMI2 250	-	-	-	-	-	-	6 911 256 Bv	71.5 (61.8, 79.6)	89.7 (86.1, 92.4)
RMI3 200	2 606 84 un	73.6 (61.3, 83.1)	91.1 (87.3, 94.0)	2 604 314 un	85.8 (58.3, 96.3)	61.7 (56.0, 67.1)	19 5945 1761 bv	75.9 (68.8, 81.9)	84.7 (80.0, 88.5)
RMI3 250	-	-	-	-	-	-	5 1683 436 Bv	78.4 (67.4, 86.5)	89.6 (86.3, 92.2)
RMI4 400	-	-	-	-	-	-	3 506 92 un	82.6 (73.5, 89.1)	89.1 (20.8, 73.2)
RMI4 450	-	-	-	-	-	-	6 1971 462 Bv	92.2 (90.2, 93.9)	
ROMA 12.5 pre 14.4 post	3 302 68 un	63.5 (51.0, 74.4)	89.3 (80.8, 94.3)	3 299 177 un	88.0 (80.6, 92.8)	68.3 (57.4, 77.5)	3 601 245 un	80.2 (72.0, 86.5)	82.6 (76.9, 87.2)
ROMA 13.1 pre 27.7 post	5 855 100 un	75.8 (66.1, 83.5)	80.8 (72.5, 87.0)	5 760 314 bv	87.8 (82.2, 94.3)	76.9 (72.7, 80.7)	5 1615 414 Bv	86.4 (81.1, 90.4)	79.5 (73.5, 84.5)
ROMA 7.4 pre 25.3 post	3 975 107 un	84.1 (63.4, 94.2)	75.6 (55.6, 88.5)	3 767 339 bn	95.3 (90.5, 97.7)	77.5 (59.4, 89.0)	5 2194 673 bv	92.1 (86.8, 95.4)	73.3 (56.9, 85.1)

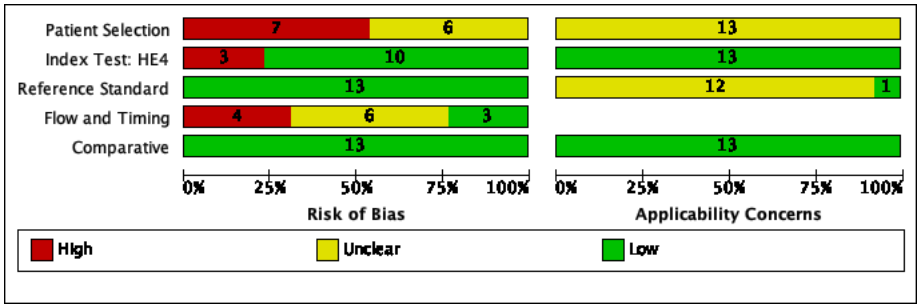
ROMA 7.4(+-2) pre 25.3(+-2) post	6 1269 173 bv	82.6 (71.6, 89.9)	80.0 (68.4, 88.1)	7 1396 544 bv	92.2 (87.4, 95.2)	81.8 (71.8, 88.8)	-	-	-
ROMA 13.1(+-2) pre 27.7 (+-2) post	13 1487 307 bv	78.3 (71.2, 84.0)	83.3 (78.2, 87.5)	9 1497 610 bv	90.7 (86.4, 93.7)	79.1 (74.6, 83.1)	-	-	-
ACOG V1	3 716 154 un	69.4 (59.3, 78.0)	60.5 (36.1, 80.6)	3 1301 585 un	91.4 (86.3, 94.7)	53.0 (43.9, 61.9)	2 1180 428 un	83.5 (73.1, 90.4)	52.5 (31.4, 72.7)
ACOG V2	1 240 48 bp	85.4 (72.2,93.9)	60.0 (52.6, 66.9)	1 597 263 bp	93.2 (89.4, 95.9)	59.9 (54.4, 65.2)	-	-	-
ACOG V3 CA125	3 899 150 un	78.0 (70.4, 84.1)	83.9 (65.1, 92.6)	3 1224 486 un	84.6 (76.0, 90.5)	78.1 (71.0, 83.9)	2 1286 325 un	79.4 (74.6, 83.4)	75.4 (69.1, 80.8)
ACOG V3 OVA1	1 235 45 bp	91.1 (78.8, 97.5)	43.2 (36.0, 50.5)	1 281 116 bp	94.8 (89.1, 98.1)	25.5 (19.0, 32.8)	1 516 161 bp	93.8 (88.9, 97.0)	34.9 (30.0, 40.1)

Notes to table: Analyses here have been undertaken using the bivariate model to estimate average sensitivity and specificity for each test at each specified threshold. Where possible estimates have been made separately for premenopausal, postmenopausal and pre and postmenopausal mixed groups. Analysis method: bp = single study estimate. uv=univariate analysis. bv=bivariate anlaysis. D+ = Primary ovarian cancer and borderline ovarian tumours (where borderline tumours included in primary study analysis). ACOG threshold: any one component test positive, overall test positive.

Appendix 13 **: CA125 Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies.**



Appendix 14 **: HE4 Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies.**



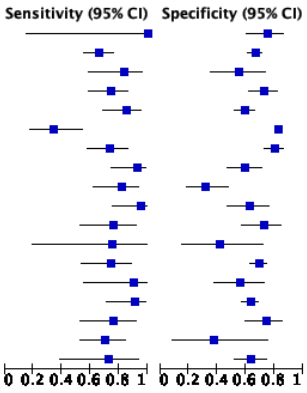
Appendix 15 : **Test performance: CA125 treatment of borderline results.**

Menopausal status	Test	Threshold	Number studies (N total, N OC)	Sensitivity (95% CI)	Specificity (95% CI)	Notes
Pre all studies	CA125	35ish	17	78 (71 to 84)	66 (61 to 71)	Bivariate, same metandi
Post all studies	CA125	35ish	21	84 (80 to 87)	83 (81 to 84)	Bivariate, same metandi
Pre BOT1 only	CA125	35ish	7	71 (58 to 82)	70 (62 to 77)	Bivariate agrees with metandi
Post BOT1 only	CA125	35ish	9	78 (74 to 82)	82 (80 to 84)	Bivariate agrees with metandi
BOT1-BOT2/3 Pre	CA125	35ish	17	-11.1 (-24.5 to 2.4)	7.3 (-3.1 to 17.8)	Bivariate
BOT1-BOT2/3 Post	CA125	35ish	21	-10.2 (-15.3 to -5.1)	-1.4 (-4.7 to 1.9)	univariable

Appendix 16 : **CA125 35 (+/- 10) pre post Forest plot.**

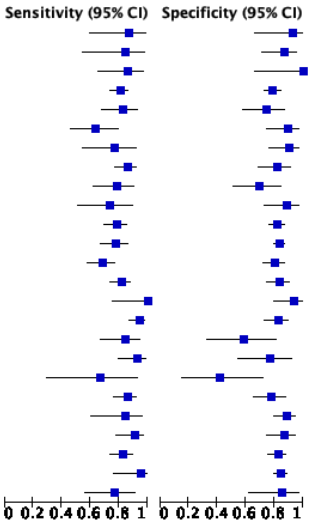
CA125 35(+-10) pre

Study	TP	FP	FN	TN	Threshold (pmol/L)	BOT with	Sensitivity (95% CI)	Specificity (95% CI)
Smikle 1995	2	12	0	36		Not clear	1.00 [0.16, 1.00]	0.75 [0.60, 0.86]
Van Calster 2007b	57	120	29	239	30.0	Malignant	0.66 [0.55, 0.76]	0.67 [0.61, 0.71]
Anton 2012	15	13	3	16	35.0	Malignant	0.83 [0.59, 0.96]	0.55 [0.36, 0.74]
Dowd 1993	32	21	11	57	35.0	Malignant	0.74 [0.59, 0.86]	0.73 [0.62, 0.82]
Holcomb 2011	29	79	5	116	35.0	Malignant	0.85 [0.69, 0.95]	0.59 [0.52, 0.66]
Radosa 2011	10	147	19	704	35.0	Malignant	0.34 [0.18, 0.54]	0.83 [0.80, 0.85]
Van Gorp 2011	31	28	11	114	35.0	Malignant	0.74 [0.58, 0.86]	0.80 [0.73, 0.86]
Bandiera 2011	24	28	2	41	35.0	Not clear	0.92 [0.75, 0.99]	0.59 [0.47, 0.71]
Benjapibal 2007	22	30	5	14	35.0	Not clear	0.81 [0.62, 0.94]	0.32 [0.19, 0.48]
Chen 2015	19	18	1	30	35.0	Not clear	0.95 [0.75, 1.00]	0.63 [0.47, 0.76]
Farzaneh 2014	16	13	5	34	35.0	Not clear	0.76 [0.53, 0.92]	0.72 [0.57, 0.84]
Kuesel 1992	3	7	1	5	35.0	Not clear	0.75 [0.19, 0.99]	0.42 [0.15, 0.72]
Molina 2011	20	70	7	156	35.0	Not clear	0.74 [0.54, 0.89]	0.69 [0.63, 0.75]
Ortiz-Munoz 2014	9	15	1	19	35.0	Not clear	0.90 [0.55, 1.00]	0.56 [0.38, 0.73]
Chan 2013	20	99	2	170	35.0	Excluded	0.91 [0.71, 0.99]	0.63 [0.57, 0.69]
Partheen 2011	16	13	5	37	35.0	Excluded	0.76 [0.53, 0.92]	0.74 [0.60, 0.85]
Stiekema 2014	24	5	10	3	35.0	Excluded	0.71 [0.53, 0.85]	0.38 [0.09, 0.76]
Franchi 1995	8	26	3	46	40.0	Malignant	0.73 [0.39, 0.94]	0.64 [0.52, 0.75]



CA125 35(+-10) post

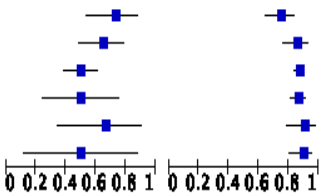
Study	TP	FP	FN	TN	Threshold (pmol/L)	BOT with	Sensitivity (95% CI)	Specificity (95% CI)
Antonic and Rakar 1995	13	1	2	13		Malignant	0.87 [0.60, 0.98]	0.93 [0.66, 1.00]
Smikle 1995	11	5	2	32		Not clear	0.85 [0.55, 0.98]	0.86 [0.71, 0.95]
Farzaneh 2014	19	0	3	9	25.0	Not clear	0.86 [0.65, 0.97]	1.00 [0.66, 1.00]
Van Calster 2007b	126	44	30	164	30.0	Malignant	0.81 [0.74, 0.87]	0.79 [0.73, 0.84]
Fujiwara 2015	34	10	7	29	31.5	Excluded	0.83 [0.68, 0.93]	0.74 [0.58, 0.87]
Anton 2012	23	4	13	33	35.0	Malignant	0.64 [0.46, 0.79]	0.89 [0.75, 0.97]
Botsis 1997	17	4	5	36	35.0	Malignant	0.77 [0.55, 0.92]	0.90 [0.76, 0.97]
Dowd 1993	80	9	13	41	35.0	Malignant	0.86 [0.77, 0.92]	0.82 [0.69, 0.91]
Engelen 2008	29	10	8	23	35.0	Malignant	0.78 [0.62, 0.90]	0.70 [0.51, 0.84]
Kusnetzoff 1998	17	4	6	30	35.0	Malignant	0.74 [0.52, 0.90]	0.88 [0.73, 0.97]
Maggino 1994	83	41	23	188	35.0	Malignant	0.78 [0.69, 0.86]	0.82 [0.77, 0.87]
Radosa 2011	60	65	17	339	35.0	Malignant	0.78 [0.67, 0.87]	0.84 [0.80, 0.87]
Schutter 1994	69	25	32	102	35.0	Malignant	0.68 [0.58, 0.77]	0.80 [0.72, 0.87]
Van Gorp 2011	97	14	22	72	35.0	Malignant	0.82 [0.73, 0.88]	0.84 [0.74, 0.91]
Avsar F.	13	2	0	31	35.0	Not clear	1.00 [0.75, 1.00]	0.94 [0.80, 0.99]
Bandiera 2011	82	17	5	79	35.0	Not clear	0.94 [0.87, 0.98]	0.82 [0.73, 0.89]
Benjapibal 2007	27	7	5	10	35.0	Not clear	0.84 [0.67, 0.95]	0.59 [0.33, 0.82]
Chen 2015	37	5	3	17	35.0	Not clear	0.93 [0.80, 0.98]	0.77 [0.55, 0.92]
Kuesel 1992	6	7	3	5	35.0	Not clear	0.67 [0.30, 0.93]	0.42 [0.15, 0.72]
Molina 2011	72	13	12	46	35.0	Not clear	0.86 [0.76, 0.92]	0.78 [0.65, 0.88]
Ortiz-Munoz 2014	16	10	3	75	35.0	Not clear	0.84 [0.60, 0.97]	0.88 [0.79, 0.94]
Chan 2013	39	7	4	46	35.0	Excluded	0.91 [0.78, 0.97]	0.87 [0.75, 0.95]
Partheen 2011	77	29	16	136	35.0	Excluded	0.83 [0.74, 0.90]	0.82 [0.76, 0.88]
Novotny 2012	20	36	1	199	36.0	Not clear	0.95 [0.76, 1.00]	0.85 [0.79, 0.89]
Franchi 1995	20	3	6	17	40.0	Malignant	0.77 [0.56, 0.91]	0.85 [0.62, 0.97]



Appendix 17 **: CA125 65 (+/- 10) pre post Forest plot.**

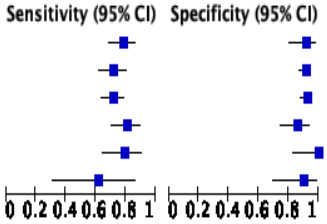
CA125 65(+-10) pre

Study	TP	FP	FN	TN	Threshold (pmol/L)	BOT with	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Fujiwara 2015	22	23	8	69	62.1	Excluded	0.73 [0.54, 0.88]	0.75 [0.65, 0.83]		
Dowd 1993	28	11	15	67	65.0	Malignant	0.65 [0.49, 0.79]	0.86 [0.76, 0.93]		
Van Calster 2007b	43	45	43	314	65.0	Malignant	0.50 [0.39, 0.61]	0.87 [0.84, 0.91]		
Gadducci 1992	8	26	8	171	65.0	Not clear	0.50 [0.25, 0.75]	0.87 [0.81, 0.91]		
Gadducci 1996	8	4	4	41	65.0	Not clear	0.67 [0.35, 0.90]	0.91 [0.79, 0.98]		
Strigini 1996	3	7	3	62	65.0	Not clear	0.50 [0.12, 0.88]	0.90 [0.80, 0.96]		



CA125 65(+-10) post

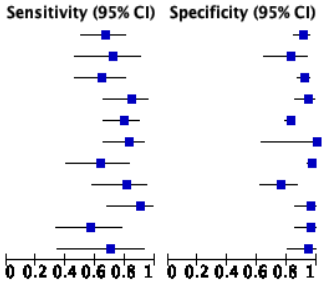
Study	TP	FP	FN	TN	Threshold (pmol/L)	BOT with	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Dowd 1993	73	4	20	46	65.0	Malignant	0.78 [0.69, 0.86]	0.92 [0.81, 0.98]		
Maggino 1994	76	19	30	210	65.0	Malignant	0.72 [0.62, 0.80]	0.92 [0.87, 0.95]		
Van Calster 2007b	112	16	44	192	65.0	Malignant	0.72 [0.64, 0.79]	0.92 [0.88, 0.96]		
Gadducci 1992	60	8	14	49	65.0	Not clear	0.81 [0.70, 0.89]	0.86 [0.74, 0.94]		
Gadducci 1996	35	0	9	20	65.0	Not clear	0.80 [0.65, 0.90]	1.00 [0.83, 1.00]		
Strigini 1996	8	2	5	19	65.0	Not clear	0.62 [0.32, 0.86]	0.90 [0.70, 0.99]		



Appendix 18 **: HE4 70 (+/- 10) pre/post Forest plot.**

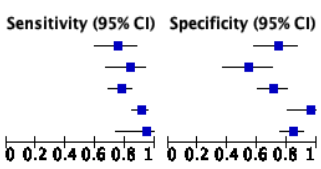
HE4 70(+-10) pre

Study	TP	FP	FN	TN	Threshold (pmol/L)	BOT with	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Van Gorp 2011	28	13	14	129	70.0	Malignant	0.67 [0.50, 0.80]	0.91 [0.85, 0.95]		
Anton 2012	13	5	5	24	70.0	Malignant	0.72 [0.47, 0.90]	0.83 [0.64, 0.94]		
Holcomb 2011	22	16	12	179	70.0	Malignant	0.65 [0.46, 0.80]	0.92 [0.87, 0.95]		
Bandiera 2011	22	4	4	65	70.0	Not clear	0.85 [0.65, 0.96]	0.94 [0.86, 0.98]		
Karlsen 2012	39	93	10	437	70.0	Not clear	0.80 [0.66, 0.90]	0.82 [0.79, 0.86]		
Stiekema 2014	28	0	6	8	70.0	Excluded	0.82 [0.65, 0.93]	1.00 [0.63, 1.00]		
Chan 2013	14	9	8	260	70.0	Excluded	0.64 [0.41, 0.83]	0.97 [0.94, 0.98]		
Partheen 2011	17	12	4	38	71.8	Excluded	0.81 [0.58, 0.95]	0.76 [0.62, 0.87]		
Chen 2015	18	2	2	46	72.3	Not clear	0.90 [0.68, 0.99]	0.96 [0.86, 0.99]		
Farzaneh 2014	12	2	9	45	75.0	Not clear	0.57 [0.34, 0.78]	0.96 [0.85, 0.99]		
Ortiz-Munoz 2014	7	2	3	32	77.0	Not clear	0.70 [0.35, 0.93]	0.94 [0.80, 0.99]		

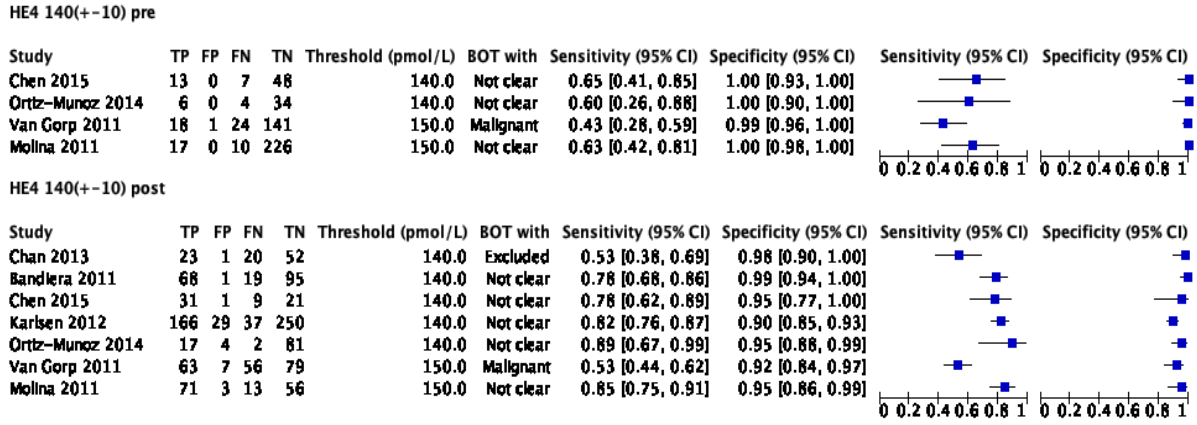


HE4 70(+-10) post

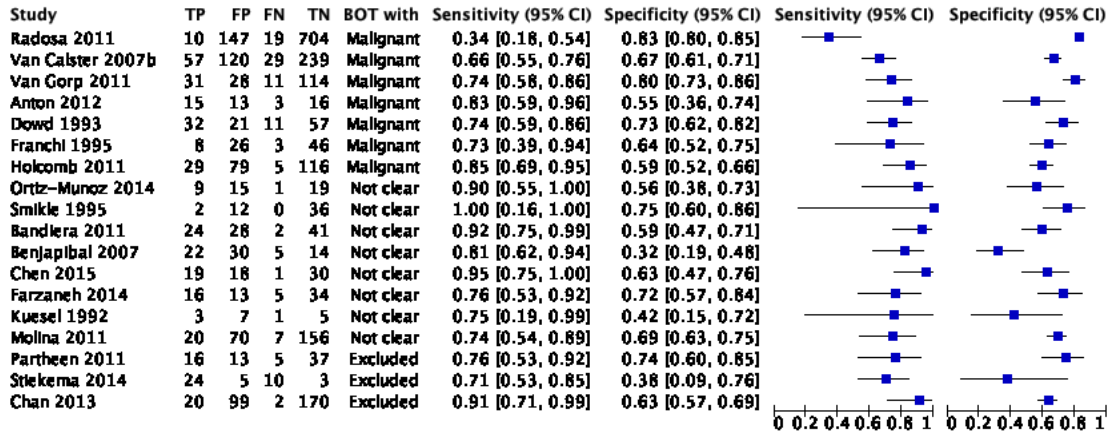
Study	TP	FP	FN	TN	Threshold (pmol/L)	BOT with	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Fujiwara 2015	31	10	10	29	63.3	Not clear	0.76 [0.60, 0.88]	0.74 [0.58, 0.87]		
Anton 2012	30	17	6	20	70.0	Malignant	0.83 [0.67, 0.94]	0.54 [0.37, 0.71]		
Van Gorp 2011	92	25	27	61	70.0	Malignant	0.77 [0.69, 0.84]	0.71 [0.60, 0.80]		
Stiekema 2014	103	1	10	25	70.0	Excluded	0.91 [0.84, 0.96]	0.96 [0.80, 1.00]		
Ortiz-Munoz 2014	18	13	1	72	77.0	Not clear	0.95 [0.74, 1.00]	0.85 [0.75, 0.92]		



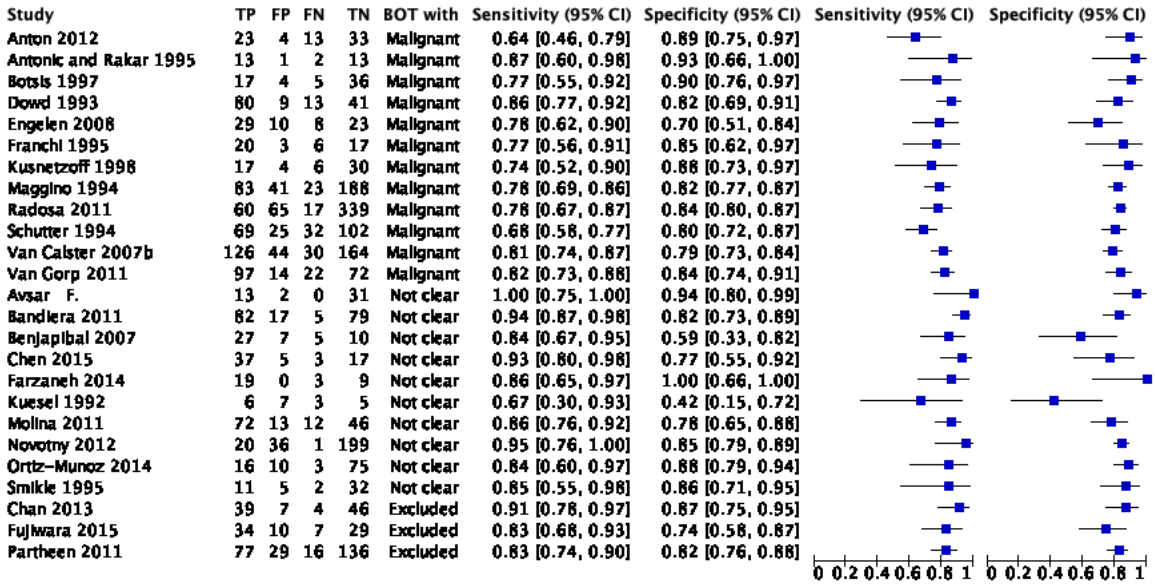
Appendix 19 **: HE4 140 (+/- 10) pre post Forest plot.**



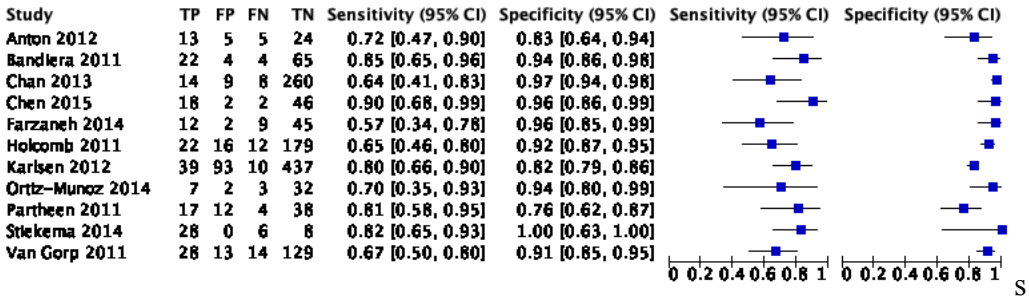
Appendix 20 **: CA125 35 (+/- 10) pre BOT Forest plot.**



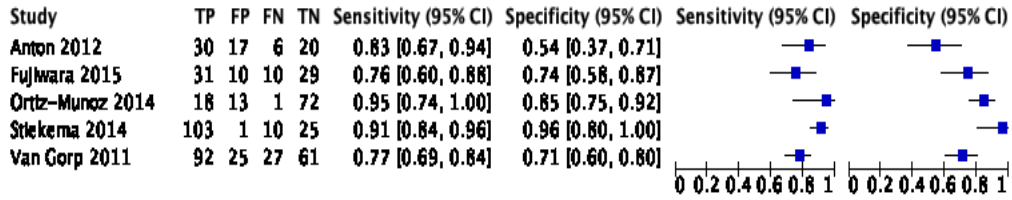
Appendix 21 **: CA125 65 (+/- 10) post BOT Forest plot.**



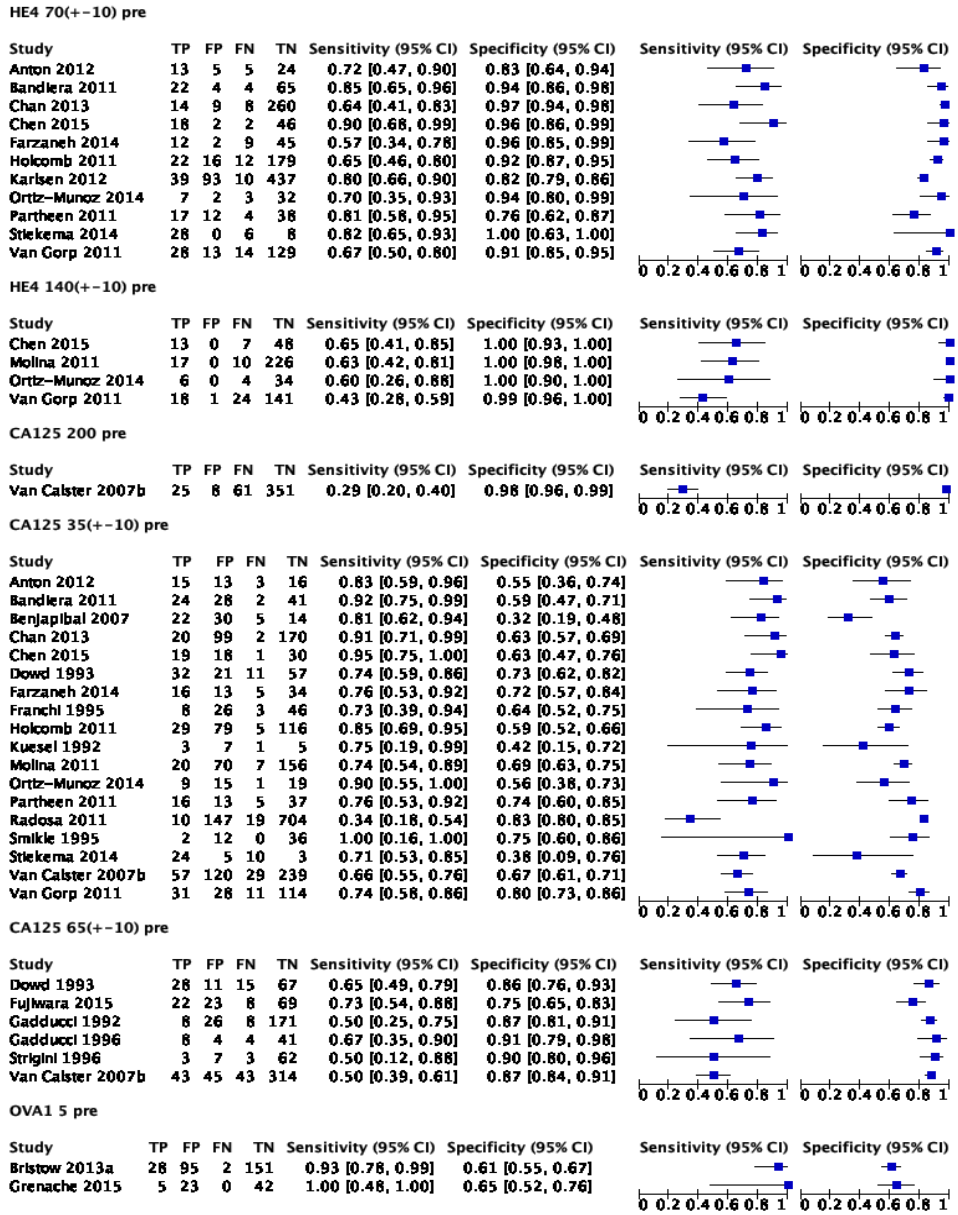
Appendix 22 **: HE4 70 (+/- 10) pre BOT Forest plot.**



Appendix 23 **: HE4 70 (+/- 10) post BOT Forest plot.**



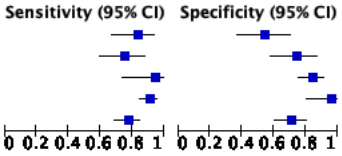
Appendix 24 : CA125, HE4 and OVA1- All tests Pre Forest plot.



Appendix 25 **: CA125, HE4 and OVA1- All tests post Forest plot.**

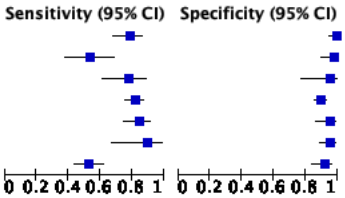
HE4 70(+-10) post

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Anton 2012	30	17	6	20	0.83 [0.67, 0.94]	0.54 [0.37, 0.71]
Fujiwara 2015	31	10	10	29	0.76 [0.60, 0.88]	0.74 [0.58, 0.87]
Ortiz-Munoz 2014	18	13	1	72	0.95 [0.74, 1.00]	0.85 [0.75, 0.92]
Stiekema 2014	103	1	10	25	0.91 [0.84, 0.96]	0.96 [0.80, 1.00]
Van Gorp 2011	92	25	27	61	0.77 [0.69, 0.84]	0.71 [0.60, 0.80]



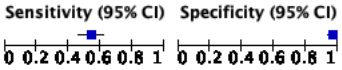
HE4 140(+-10) post

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bandiera 2011	68	1	19	95	0.78 [0.68, 0.86]	0.99 [0.94, 1.00]
Chan 2013	23	1	20	52	0.53 [0.38, 0.69]	0.98 [0.90, 1.00]
Chen 2015	31	1	9	21	0.78 [0.62, 0.89]	0.95 [0.77, 1.00]
Karlsen 2012	166	29	37	250	0.82 [0.76, 0.87]	0.90 [0.85, 0.93]
Molina 2011	71	3	13	56	0.85 [0.75, 0.91]	0.95 [0.86, 0.99]
Ortiz-Munoz 2014	17	4	2	81	0.89 [0.67, 0.99]	0.95 [0.88, 0.99]
Van Gorp 2011	63	7	56	79	0.53 [0.44, 0.62]	0.92 [0.84, 0.97]



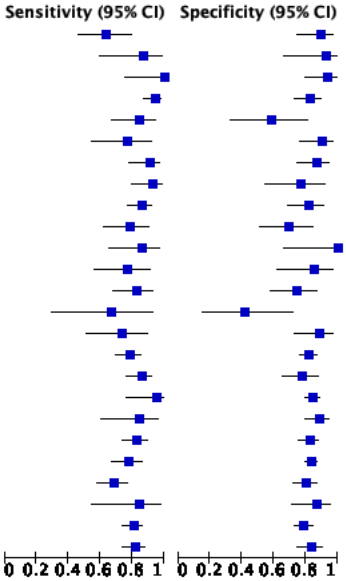
CA125 200 post

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Van Calster 2007b	85	6	71	202	0.54 [0.46, 0.62]	0.97 [0.94, 0.99]



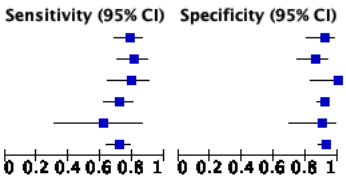
CA125 35(+-10) post

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Anton 2012	23	4	13	33	0.64 [0.46, 0.79]	0.89 [0.75, 0.97]
Antonic and Rakar 1995	13	1	2	13	0.87 [0.60, 0.98]	0.93 [0.66, 1.00]
Avsar F.	13	2	0	31	1.00 [0.75, 1.00]	0.94 [0.80, 0.99]
Bandiera 2011	82	17	5	79	0.94 [0.87, 0.98]	0.82 [0.73, 0.89]
Benjapibal 2007	27	7	5	10	0.84 [0.67, 0.95]	0.59 [0.33, 0.82]
Botsis 1997	17	4	5	36	0.77 [0.55, 0.92]	0.90 [0.76, 0.97]
Chan 2013	39	7	4	46	0.91 [0.78, 0.97]	0.87 [0.75, 0.95]
Chen 2015	37	5	3	17	0.93 [0.80, 0.98]	0.77 [0.55, 0.92]
Dowd 1993	80	9	13	41	0.86 [0.77, 0.92]	0.82 [0.69, 0.91]
Engelen 2008	29	10	8	23	0.78 [0.62, 0.90]	0.70 [0.51, 0.84]
Farzaneh 2014	19	0	3	9	0.86 [0.65, 0.97]	1.00 [0.66, 1.00]
Franchi 1995	20	3	6	17	0.77 [0.56, 0.91]	0.85 [0.62, 0.97]
Fujiwara 2015	34	10	7	29	0.83 [0.68, 0.93]	0.74 [0.58, 0.87]
Kuesel 1992	6	7	3	5	0.67 [0.30, 0.93]	0.42 [0.15, 0.72]
Kusnetzoff 1998	17	4	6	30	0.74 [0.52, 0.90]	0.88 [0.73, 0.97]
Maggino 1994	83	41	23	188	0.78 [0.69, 0.86]	0.82 [0.77, 0.87]
Molina 2011	72	13	12	46	0.86 [0.76, 0.92]	0.78 [0.65, 0.88]
Novotny 2012	20	36	1	199	0.95 [0.76, 1.00]	0.85 [0.79, 0.89]
Ortiz-Munoz 2014	16	10	3	75	0.84 [0.60, 0.97]	0.88 [0.79, 0.94]
Partheen 2011	77	29	16	136	0.83 [0.74, 0.90]	0.82 [0.76, 0.88]
Radosa 2011	60	65	17	339	0.78 [0.67, 0.87]	0.84 [0.80, 0.87]
Schutter 1994	69	25	32	102	0.68 [0.58, 0.77]	0.80 [0.72, 0.87]
Smikle 1995	11	5	2	32	0.85 [0.55, 0.98]	0.86 [0.71, 0.95]
Van Calster 2007b	126	44	30	164	0.81 [0.74, 0.87]	0.79 [0.73, 0.84]
Van Gorp 2011	97	14	22	72	0.82 [0.73, 0.88]	0.84 [0.74, 0.91]



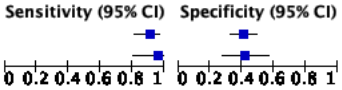
CA125 65(+-10) post

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Dowd 1993	73	4	20	46	0.78 [0.69, 0.86]	0.92 [0.81, 0.98]
Gadducci 1992	60	8	14	49	0.81 [0.70, 0.89]	0.86 [0.74, 0.94]
Gadducci 1996	35	0	9	20	0.80 [0.65, 0.90]	1.00 [0.83, 1.00]
Maggino 1994	76	19	30	210	0.72 [0.62, 0.80]	0.92 [0.87, 0.95]
Strigini 1996	8	2	5	19	0.62 [0.32, 0.86]	0.90 [0.70, 0.99]
Van Calster 2007b	112	16	44	192	0.72 [0.64, 0.79]	0.92 [0.88, 0.96]



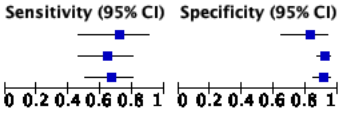
OVA1 4.4 post

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bristow 2013a	53	92	5	64	0.91 [0.81, 0.97]	0.41 [0.33, 0.49]
Grenache 2015	25	29	1	21	0.96 [0.80, 1.00]	0.42 [0.28, 0.57]



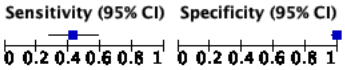
HE4 70(+−10) pre

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Anton 2012	13	5	5	24	0.72 [0.47, 0.90]	0.83 [0.64, 0.94]
Holcomb 2011	22	16	12	179	0.65 [0.46, 0.80]	0.92 [0.87, 0.95]
Van Gorp 2011	28	13	14	129	0.67 [0.50, 0.80]	0.91 [0.85, 0.95]



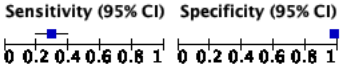
HE4 140(+−10) pre

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Van Gorp 2011	18	1	24	141	0.43 [0.28, 0.59]	0.99 [0.96, 1.00]



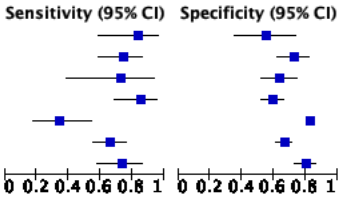
CA125 200 pre

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Van Calster 2007b	25	8	61	351	0.29 [0.20, 0.40]	0.98 [0.96, 0.99]



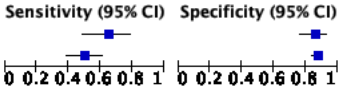
CA125 35(+−10) pre

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Anton 2012	15	13	3	16	0.83 [0.59, 0.96]	0.55 [0.36, 0.74]
Dowd 1993	32	21	11	57	0.74 [0.59, 0.86]	0.73 [0.62, 0.82]
Franchi 1995	8	26	3	46	0.73 [0.39, 0.94]	0.64 [0.52, 0.75]
Holcomb 2011	29	79	5	116	0.85 [0.69, 0.95]	0.59 [0.52, 0.66]
Radosa 2011	10	147	19	704	0.34 [0.18, 0.54]	0.83 [0.80, 0.85]
Van Calster 2007b	57	120	29	239	0.66 [0.55, 0.76]	0.67 [0.61, 0.71]
Van Gorp 2011	31	28	11	114	0.74 [0.58, 0.86]	0.80 [0.73, 0.86]



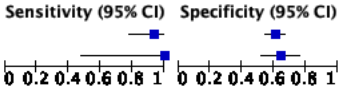
CA125 65(+−10) pre

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Dowd 1993	28	11	15	67	0.65 [0.49, 0.79]	0.86 [0.76, 0.93]
Van Calster 2007b	43	45	43	314	0.50 [0.39, 0.61]	0.87 [0.84, 0.91]



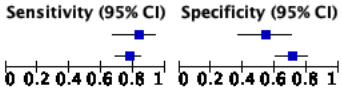
OVA1 5 pre

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bristow 2013a	28	95	2	151	0.93 [0.78, 0.99]	0.61 [0.55, 0.67]
Grenache 2015	5	23	0	42	1.00 [0.48, 1.00]	0.65 [0.52, 0.76]



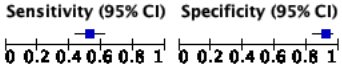
HE4 70(+-10) post

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Anton 2012	30	17	6	20	0.83 [0.67, 0.94]	0.54 [0.37, 0.71]
Van Gorp 2011	92	25	27	61	0.77 [0.69, 0.84]	0.71 [0.60, 0.80]



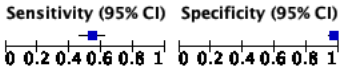
HE4 140(+-10) post

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Van Gorp 2011	63	7	56	79	0.53 [0.44, 0.62]	0.92 [0.84, 0.97]



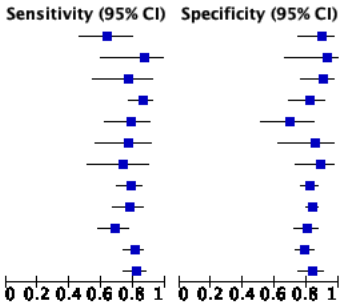
CA125 200 post

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Van Calster 2007b	85	6	71	202	0.54 [0.46, 0.62]	0.97 [0.94, 0.99]



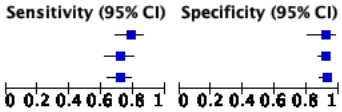
CA125 35(+-10) post

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Anton 2012	23	4	13	33	0.64 [0.46, 0.79]	0.89 [0.75, 0.97]
Antonic and Rakar 1995	13	1	2	13	0.87 [0.60, 0.98]	0.93 [0.66, 1.00]
Botsis 1997	17	4	5	36	0.77 [0.55, 0.92]	0.90 [0.76, 0.97]
Dowd 1993	80	9	13	41	0.86 [0.77, 0.92]	0.82 [0.69, 0.91]
Engelen 2008	29	10	8	23	0.78 [0.62, 0.90]	0.70 [0.51, 0.84]
Franchi 1995	20	3	6	17	0.77 [0.56, 0.91]	0.85 [0.62, 0.97]
Kusnetzoff 1998	17	4	6	30	0.74 [0.52, 0.90]	0.88 [0.73, 0.97]
Maggino 1994	83	41	23	188	0.78 [0.69, 0.86]	0.82 [0.77, 0.87]
Radosa 2011	60	65	17	339	0.78 [0.67, 0.87]	0.84 [0.80, 0.87]
Schutter 1994	69	25	32	102	0.68 [0.58, 0.77]	0.80 [0.72, 0.87]
Van Calster 2007b	126	44	30	164	0.81 [0.74, 0.87]	0.79 [0.73, 0.84]
Van Gorp 2011	97	14	22	72	0.82 [0.73, 0.88]	0.84 [0.74, 0.91]



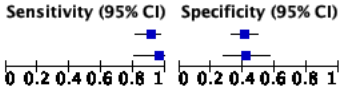
CA125 65(+-10) post

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Dowd 1993	73	4	20	46	0.78 [0.69, 0.86]	0.92 [0.81, 0.98]
Maggino 1994	76	19	30	210	0.72 [0.62, 0.80]	0.92 [0.87, 0.95]
Van Calster 2007b	112	16	44	192	0.72 [0.64, 0.79]	0.92 [0.88, 0.96]



OVA1 4.4 post

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bristow 2013a	53	92	5	64	0.91 [0.81, 0.97]	0.41 [0.33, 0.49]
Grenache 2015	25	29	1	21	0.96 [0.80, 1.00]	0.42 [0.28, 0.57]



LIST OF REFERENCES

1. Meinhold-Heerlein I, Fotopoulou C, Harter P, et al. Statement by the Kommission Ovar of the AGO: The New FIGO and WHO Classifications of Ovarian, Fallopian Tube and Primary Peritoneal Cancer. *Geburtshilfe Frauenheilkd* 2015;75(10):1021-27. doi: 10.1055/s-0035-1558079
2. Knudsen UB, Tabor A, Mosgaard B, et al. Management of ovarian cysts. *Acta Obstet Gynecol Scand* 2004;83(11):1012-21. doi: 10.1111/j.0001-6349.2004.00607.x
3. RCOG GJgB. Management of Suspected Ovarian Masses in Premenopausal Women: RCOG; 2011 [accessed 1.2.2019 2019.
4. Lee-Jones L. Ovarian tumours: an overview. *Atlas Genet Cytogenet Oncol Haematol* 2004;8(2):110-14.
5. SEER SEaERp. Cancer stat facts: ovarian cancer 2018 [Available from: https://seer.cancer.gov/statfacts/html/ovary.html accessed 13/9/2018 2018.
6. Gershenson DM. Clinical management potential tumours of low malignancy. *Best Pract Res Clin Obstet Gynaecol* 2002;16(4):513-27.
7. Swanton A, Bankhead CR, Kehoe S. Pregnancy rates after conservative treatment for borderline ovarian tumours: a systematic review. *Eur J Obstet Gynecol Reprod Biol* 2007;135(1):3-7. doi: 10.1016/j.ejogrb.2007.05.011
8. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;136(5):E359-86. doi: 10.1002/ijc.29210
9. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68(6):394-424. doi: 10.3322/caac.21492
10. CRUK crU. ovarian cancer statistics 2016 [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer accessed 13/9/2018 2018.
11. Torre LA, Islami F, Siegel RL, et al. Global Cancer in Women: Burden and Trends. *Cancer Epidemiol Biomarkers Prev* 2017;26(4):444-57. doi: 10.1158/1055-9965.EPI-16-0858
12. Torre LA, Trabert B, DeSantis CE, et al. Ovarian cancer statistics, 2018. *CA Cancer J Clin* 2018;68(4):284-96. doi: 10.3322/caac.21456
13. CRUK. Ovarain cancer statistics: Cancer research UK; 2011 [accessed 13/07/2015 2015.
14. Boyd J, Rubin SC. Hereditary ovarian cancer: molecular genetics and clinical implications. *Gynecol Oncol* 1997;64(2):196-206. doi: 10.1006/gyno.1996.4572

15. Ford D, Easton DF, Stratton M, et al. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am J Hum Genet* 1998;62(3):676-89.
16. Struewing JP, Hartge P, Wacholder S, et al. The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. *N Engl J Med* 1997;336(20):1401-8. doi: 10.1056/NEJM199705153362001 [published Online First: 1997/05/15]
17. Piek JM, Verheijen RH, Kenemans P, et al. BRCA1/2-related ovarian cancers are of tubal origin: a hypothesis. *Gynecol Oncol* 2003;90(2):491.
18. Kindelberger DW, Lee Y, Miron A, et al. Intraepithelial carcinoma of the fimbria and pelvic serous carcinoma: Evidence for a causal relationship. *Am J Surg Pathol* 2007;31(2):161-9. doi: 10.1097/01.pas.0000213335.40358.47
19. Marquez RT, Baggerly KA, Patterson AP, et al. Patterns of gene expression in different histotypes of epithelial ovarian cancer correlate with those in normal fallopian tube, endometrium, and colon. *Clin Cancer Res* 2005;11(17):6116-26. doi: 10.1158/1078-0432.CCR-04-2509
20. Kurman RJ, Shih Ie M. The origin and pathogenesis of epithelial ovarian cancer: a proposed unifying theory. *Am J Surg Pathol* 2010;34(3):433-43. doi: 10.1097/PAS.0b013e3181cf3d79
21. Ronnett BM, Zahn CM, Kurman RJ, et al. Disseminated peritoneal adenomucinosis and peritoneal mucinous carcinomatosis. A clinicopathologic analysis of 109 cases with emphasis on distinguishing pathologic features, site of origin, prognosis, and relationship to "pseudomyxoma peritonei". *Am J Surg Pathol* 1995;19(12):1390-408.
22. Veras E, Mao TL, Ayhan A, et al. Cystic and adenofibromatous clear cell carcinomas of the ovary: distinctive tumors that differ in their pathogenesis and behavior: a clinicopathologic analysis of 122 cases. *Am J Surg Pathol* 2009;33(6):844-53. doi: 10.1097/PAS.0b013e31819c4271
23. Seidman JD, Yemelyanova A, Zaino RJ, et al. The fallopian tube-peritoneal junction: a potential site of carcinogenesis. *Int J Gynecol Pathol* 2011;30(1):4-11. doi: 10.1097/PGP.0b013e3181f29d2a
24. American College of O, Gynecologists' Committee on Practice B-G. Practice Bulletin No. 174: Evaluation and Management of Adnexal Masses. *Obstet Gynecol* 2016;128(5):e210-e26. doi: 10.1097/AOG.0000000000001768
25. Smith-Bindman R, Poder L, Johnson E, et al. Risk of Malignant Ovarian Cancer Based on Ultrasonography Findings in a Large Unselected Population. *JAMA Intern Med* 2018 doi: 10.1001/jamainternmed.2018.5113
26. Katz VL LG, Lobo RA, et al. . Comprehensive gynecology. . Mosby; : St. Louis, MO: , 2007.

27. Hall GH, Turnbull LW, Richmond I, et al. Localisation of somatostatin and somatostatin receptors in benign and malignant ovarian tumours. *Br J Cancer* 2002;87(1):86-90. doi: 10.1038/sj.bjc.6600284
28. Stany MP, Hamilton CA. Benign disorders of the ovary. *Obstet Gynecol Clin North Am* 2008;35(2):271-84, ix. doi: 10.1016/j.ogc.2008.03.004
29. Padubidri V.G Sd. Shaws text of Gynecology. 16 ed. India Elsevier 2014:429.
30. Seidman JD RP, Kurman RJ: . Surface epithelial tumors of the ovary. 6th ed. ed. New York, Springer-Verlag2002.): 791–904
31. Kurjak A, Kupesic S, Jacobs I. Preoperative diagnosis of the primary fallopian tube carcinoma by three-dimensional static and power Doppler sonography. *Ultrasound Obstet Gynecol* 2000;15(3):246-51. doi: 10.1046/j.1469-0705.2000.00080.x
32. Limaiem F, Mlika M. Ovararian Cystadenoma. StatPearls. Treasure Island (FL)2018.
33. Koonings PP, Campbell K, Mishell DR, Jr., et al. Relative frequency of primary ovarian neoplasms: a 10-year review. *Obstet Gynecol* 1989;74(6):921-6.
34. Seidman. J KR. Benign, Proliferative Noninvasive (Borderline), and Invasive Epithelial Tumors of the Ovary: Global library of Womens medicine; 2008 [updated july 2008; cited 2019 8.2.19]. Available from: https://www.glowm.com/section_view/heading/Benign,%20Proliferative%20Noninvasive%20(Borderline),%20and%20Invasive%20Epithelial%20Tumors%20of%20the%20Ovary/item/248 accessed 7.2.19 2019.
35. Chene G, Ouellet V, Rahimi K, et al. The ARID1A pathway in ovarian clear cell and endometrioid carcinoma, contiguous endometriosis, and benign endometriosis. *Int J Gynaecol Obstet* 2015;130(1):27-30. doi: 10.1016/j.ijgo.2015.02.021
36. Brilhante AV, Augusto KL, Portela MC, et al. Endometriosis and Ovarian Cancer: an Integrative Review (Endometriosis and Ovarian Cancer). *Asian Pac J Cancer Prev* 2017;18(1):11-16. doi: 10.22034/APJCP.2017.18.1.11
37. Scully RE YR, Clement PB. , editor. *Tumours of the ovary, maldeveloped gonads, fallopian tube and broad ligament.* . 3rd ed. ed, 1998.
38. Scully RE, Barlow JF. "Mesonephroma" of ovary. Tumor of Mullerian nature related to the endometrioid carcinoma. *Cancer* 1967;20(9):1405-17.
39. Kao GF, Norris HJ. Unusual cystadenofibromas: endometrioid, mucinous, and clear cell types. *Obstet Gynecol* 1979;54(6):729-36.
40. Ziadi S, Trimeche M, Hammedi F, et al. Bilateral proliferating Brenner tumor of the ovary associated with recurrent urothelial carcinoma of the urinary bladder. *N Am J Med Sci* 2010;2(1):39-41.
41. Arey LB. The origin and form of the Brenner tumor. *Am J Obstet Gynecol* 1961;81:743-51.
42. Sagae S, Kudo R. Surgery for germ cell tumors. *Semin Surg Oncol* 2000;19(1):76-81.

43. Zalel Y, Piura B, Elchalal U, et al. Diagnosis and management of malignant germ cell ovarian tumors in young females. *Int J Gynaecol Obstet* 1996;55(1):1-10.
44. Outwater EK, Siegelman ES, Hunt JL. Ovarian teratomas: tumor types and imaging characteristics. *Radiographics* 2001;21(2):475-90. doi: 10.1148/radiographics.21.2.g01mr09475
45. Park SB, Kim JK, Kim KR, et al. Imaging findings of complications and unusual manifestations of ovarian teratomas. *Radiographics* 2008;28(4):969-83. doi: 10.1148/rg.284075069
46. Haroon S, Zia A, Idrees R, et al. Clinicopathological spectrum of ovarian sex cord-stromal tumors; 20 years' retrospective study in a developing country. *J Ovarian Res* 2013;6(1):87. doi: 10.1186/1757-2215-6-87
47. Wolsky RJMR, Joseph T. MD, MPH. Benign Sex Cord–Stromal Tumors of the Ovary That May Mimic Malignancies. *AJSP: Reviews & Reports* 2016;5(21):232-43. doi: 10.1097/PCR.0000000000000159
48. Roth LM. Recent advances in the pathology and classification of ovarian sex cord-stromal tumors. *Int J Gynecol Pathol* 2006;25(3):199-215. doi: 10.1097/01.pgp.0000192271.22289.e6
49. Silverberg SG, Bell DA, Kurman RJ, et al. Borderline ovarian tumors: key points and workshop summary. *Hum Pathol* 2004;35(8):910-7.
50. Serov SS SR, Sobin LH Histological typing of ovarian tumors. . Berlin Heidelberg Springer, 1973
51. Kurman RJ, Carcangiu, M.L., Herrington, C.S., Young, R.H. WHO Classification of Tumours of Female Reproductive Organs. 4 ed. Lyon: IARC 2014:307.
52. Seidman JD, Kurman RJ. Ovarian serous borderline tumors: a critical review of the literature with emphasis on prognostic indicators. *Hum Pathol* 2000;31(5):539-57.
53. Shih Ie M, Kurman RJ. Ovarian tumorigenesis: a proposed model based on morphological and molecular genetic analysis. *Am J Pathol* 2004;164(5):1511-8.
54. Malpica A, Deavers MT, Lu K, et al. Grading ovarian serous carcinoma using a two-tier system. *Am J Surg Pathol* 2004;28(4):496-504.
55. Sherman ME, Mink PJ, Curtis R, et al. Survival among women with borderline ovarian tumors and ovarian carcinoma: a population-based analysis. *Cancer* 2004;100(5):1045-52. doi: 10.1002/cncr.20080
56. Hannibal CG, Vang R, Junge J, et al. A nationwide study of serous "borderline" ovarian tumors in Denmark 1978-2002: centralized pathology review and overall survival compared with the general population. *Gynecol Oncol* 2014;134(2):267-73. doi: 10.1016/j.ygyno.2014.06.002

57. Seidman JD, Soslow RA, Vang R, et al. Borderline ovarian tumors: diverse contemporary viewpoints on terminology and diagnostic criteria with illustrative images. *Hum Pathol* 2004;35(8):918-33.
58. CRUK crU. Ovarian cancer [9.5.19]. Available from: https://www.cancerresearchuk.org/about-cancer/ovarian-cancer/risks-causes accessed 9.5.19 2019.
59. Talerman A. Germ cell tumors of the ovary. *Curr Opin Obstet Gynecol* 1997;9(1):44-7.
60. Tewari K, Cappuccini F, Disaia PJ, et al. Malignant germ cell tumors of the ovary. *Obstet Gynecol* 2000;95(1):128-33.
61. Goldberg JM, Piver MS, Jishi MF, et al. Age at onset of ovarian cancer in women with a strong family history of ovarian cancer. *Gynecol Oncol* 1997;66(1):3-9. doi: 10.1006/gyno.1997.4737
62. Lynch HT, Casey MJ, Shaw TG, et al. Hereditary Factors in Gynecologic Cancer. *Oncologist* 1998;3(5):319-38.
63. Arts-de Jong M, de Bock GH, van Asperen CJ, et al. Germline BRCA1/2 mutation testing is indicated in every patient with epithelial ovarian cancer: A systematic review. *Eur J Cancer* 2016;61:137-45. doi: 10.1016/j.ejca.2016.03.009
64. Trainer AH, Meiser B, Watts K, et al. Moving toward personalized medicine: treatment-focused genetic testing of women newly diagnosed with ovarian cancer. *Int J Gynecol Cancer* 2010;20(5):704-16.
65. Alsop K, Fereday S, Meldrum C, et al. BRCA mutation frequency and patterns of treatment response in BRCA mutation-positive women with ovarian cancer: a report from the Australian Ovarian Cancer Study Group. *J Clin Oncol* 2012;30(21):2654-63. doi: 10.1200/JCO.2011.39.8545
66. Stratton JF, Pharoah P, Smith SK, et al. A systematic review and meta-analysis of family history and risk of ovarian cancer. *Br J Obstet Gynaecol* 1998;105(5):493-9.
67. Easton DF, Ford D, Bishop DT. Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. *Am J Hum Genet* 1995;56(1):265-71.
68. Antoniou A, Pharoah PD, Narod S, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet* 2003;72(5):1117-30. doi: 10.1086/375033
69. Chen S, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. *J Clin Oncol* 2007;25(11):1329-33. doi: 10.1200/JCO.2006.09.1066
70. Pal T, Permuth-Wey J, Betts JA, et al. BRCA1 and BRCA2 mutations account for a large proportion of ovarian carcinoma cases. *Cancer* 2005;104(12):2807-16. doi: 10.1002/cncr.21536

71. Finch A, Beiner M, Lubinski J, et al. Salpingo-oophorectomy and the risk of ovarian, fallopian tube, and peritoneal cancers in women with a BRCA1 or BRCA2 Mutation. *JAMA* 2006;296(2):185-92. doi: 10.1001/jama.296.2.185
72. Marra G, Boland CR. Hereditary nonpolyposis colorectal cancer: the syndrome, the genes, and historical perspectives. *J Natl Cancer Inst* 1995;87(15):1114-25.
73. Iaccarino I, Marra G, Palombo F, et al. hMSH2 and hMSH6 play distinct roles in mismatch binding and contribute differently to the ATPase activity of hMutSalpha. *EMBO J* 1998;17(9):2677-86. doi: 10.1093/emboj/17.9.2677
74. Ramus SJ, Song H, Dicks E, et al. Germline Mutations in the BRIP1, BARD1, PALB2, and NBN Genes in Women With Ovarian Cancer. *J Natl Cancer Inst* 2015;107(11) doi: 10.1093/jnci/djv214
75. Song H, Dicks E, Ramus SJ, et al. Contribution of Germline Mutations in the RAD51B, RAD51C, and RAD51D Genes to Ovarian Cancer in the Population. *J Clin Oncol* 2015;33(26):2901-7. doi: 10.1200/JCO.2015.61.2408
76. Taylor A, Brady AF, Frayling IM, et al. Consensus for genes to be included on cancer panel tests offered by UK genetics services: guidelines of the UK Cancer Genetics Group. *J Med Genet* 2018;55(6):372-77. doi: 10.1136/jmedgenet-2017-105188
77. Richards MA. The size of the prize for earlier diagnosis of cancer in England. *Br J Cancer* 2009;101 Suppl 2:S125-9. doi: 10.1038/sj.bjc.6605402
78. Thomson CS, Forman D. Cancer survival in England and the influence of early diagnosis: what can we learn from recent EUROCARE results? *Br J Cancer* 2009;101 Suppl 2:S102-9. doi: 10.1038/sj.bjc.6605399
79. NCRAS Ncraas. Ovarian cancer: Public Health England; 2016 [cited 2019 3/3]. Available from: http://www.ncin.org.uk/cancer_type_and_topic_specific_work/cancer_type_specific_work/gynaecological_cancer/gynaecological_cancer_hub/resources/ovarian_cancer accessed 3/3/19 2019.
80. Barclay M, Gildea C, Poole J, et al. Factors Affecting Short-term Mortality in Women With Ovarian, Tubal, or Primary Peritoneal Cancer: Population-Based Cohort Analysis of English National Cancer Registration Data. *Int J Gynecol Cancer* 2016;26(1):56-65. doi: 10.1097/IGC.0000000000000562
81. Fitch M, Deane K, Howell D, et al. Women's experiences with ovarian cancer: reflections on being diagnosed. *Can Oncol Nurs J* 2002;12(3):152-68.
82. Myers ER, Bastian LA, Havrilesky LJ, et al. Management of adnexal mass. *Evid Rep Technol Assess (Full Rep)* 2006(130):1-145. [published Online First: 2007/09/15]

83. Bankhead CR, Kehoe ST, Austoker J. Symptoms associated with diagnosis of ovarian cancer: a systematic review. *BJOG* 2005;112(7):857-65. doi: 10.1111/j.1471-0528.2005.00572.x
84. Goff BA, Mandel LS, Drescher CW, et al. Development of an ovarian cancer symptom index: possibilities for earlier detection. *Cancer* 2007;109(2):221-7. doi: 10.1002/cncr.22371
85. Goff BA, Mandel L, Muntz HG, et al. Ovarian carcinoma diagnosis. *Cancer* 2000;89(10):2068-75.
86. Vine MF, Ness RB, Calingaert B, et al. Types and duration of symptoms prior to diagnosis of invasive or borderline ovarian tumor. *Gynecol Oncol* 2001;83(3):466-71. doi: 10.1006/gyno.2001.6411
87. Flam F, Einhorn N, Sjovall K. Symptomatology of ovarian cancer. *Eur J Obstet Gynecol Reprod Biol* 1988;27(1):53-7.
88. Ebell MH, Culp MB, Radke TJ. A Systematic Review of Symptoms for the Diagnosis of Ovarian Cancer. *Am J Prev Med* 2016;50(3):384-94. doi: 10.1016/j.amepre.2015.09.023 [published Online First: 2015/11/07]
89. Lim AW, Mesher D, Gentry-Maharaj A, et al. Predictive value of symptoms for ovarian cancer: comparison of symptoms reported by questionnaire, interview, and general practitioner notes. *J Natl Cancer Inst* 2012;104(2):114-24. doi: 10.1093/jnci/djr486
90. Rossing MA, Wicklund KG, Cushing-Haugen KL, et al. Predictive value of symptoms for early detection of ovarian cancer. *J Natl Cancer Inst* 2010;102(4):222-9. doi: 10.1093/jnci/djp500
91. Smith LH, Morris CR, Yasmeen S, et al. Ovarian cancer: can we make the clinical diagnosis earlier? *Cancer* 2005;104(7):1398-407. doi: 10.1002/cncr.21310
92. Gilbert L, Basso O, Sampalis J, et al. Assessment of symptomatic women for early diagnosis of ovarian cancer: results from the prospective DOvE pilot project. *Lancet Oncol* 2012;13(3):285-91. doi: 10.1016/S1470-2045(11)70333-3 [published Online First: 2012/01/20]
93. Bristow RE, Tomacruz RS, Armstrong DK, et al. Survival effect of maximal cytoreductive surgery for advanced ovarian carcinoma during the platinum era: a meta-analysis. *J Clin Oncol* 2002;20(5):1248-59. doi: 10.1200/JCO.2002.20.5.1248
94. Eisenkop SM, Spirtos NM, Friedman RL, et al. Relative influences of tumor volume before surgery and the cytoreductive outcome on survival for patients with advanced ovarian cancer: a prospective study. *Gynecol Oncol* 2003;90(2):390-6.
95. Winter WE, 3rd, Maxwell GL, Tian C, et al. Prognostic factors for stage III epithelial ovarian cancer: a Gynecologic Oncology Group Study. *J Clin Oncol* 2007;25(24):3621-7. doi: 10.1200/JCO.2006.10.2517

96. Jacobs I, Bast RC, Jr. The CA 125 tumour-associated antigen: a review of the literature. *Hum Reprod* 1989;4(1):1-12.
97. Moss EL, Hollingworth J, Reynolds TM. The role of CA125 in clinical practice. *J Clin Pathol* 2005;58(3):308-12. doi: 10.1136/jcp.2004.018077
98. Rai N, Nevin J, Downey G, et al. Outcomes following implementation of symptom triggered diagnostic testing for ovarian cancer. *Eur J Obstet Gynecol Reprod Biol* 2015;187:64-9. doi: 10.1016/j.ejogrb.2015.02.011
99. du Bois A, Quinn M, Thigpen T, et al. 2004 consensus statements on the management of ovarian cancer: final document of the 3rd International Gynecologic Cancer Intergroup Ovarian Cancer Consensus Conference (GCIG OCCC 2004). *Ann Oncol* 2005;16 Suppl 8:viii7-viii12. doi: 10.1093/annonc/mdi961
100. Eisenkop SM, Spirtos NM, Montag TW, et al. The impact of subspecialty training on the management of advanced ovarian cancer. *Gynecol Oncol* 1992;47(2):203-9.
101. Nirmala Rai RC, Kym Snell, Susan Mallett, Susan E Bayliss, Richard D Neal, Moji Balogun, Sean Kehoe, Jonathan J Deeks, Sudha Sundar, Clare Davenport. Symptoms, ultrasound imaging and biochemical markers alone or in combination for the diagnosis of ovarian cancer in women with symptoms suspicious of ovarian cancer. 7/12/2015 ed. Cochrane: Cochrane, 2015.
102. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med* 2019;170(1):51-58. doi: 10.7326/M18-1376
103. StataCorp. Statistical software: Scientific Research; 2015 [Available from: https://www.scirp.org/(S(i43dyn45teexjx455qlt3d2q))/reference/ReferencesPapers.aspx?ReferenceID=1736202 accessed 11.5.19 2019.
104. Alberg AJ, Park JW, Hager BW, et al. The use of "overall accuracy" to evaluate the validity of screening or diagnostic tests. *J Gen Intern Med* 2004;19(5 Pt 1):460-5. doi: 10.1111/j.1525-1497.2004.30091.x [published Online First: 2004/04/28]
105. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol* 2006;59(12):1331-2; author reply 32-3. doi: 10.1016/j.jclinepi.2006.06.011
106. Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58(10):982-90. doi: 10.1016/j.jclinepi.2005.02.022
107. Takwoingi Y, Leeflang MM, Deeks JJ. Empirical evidence of the importance of comparative studies of diagnostic test accuracy. *Ann Intern Med* 2013;158(7):544-54. doi: 10.7326/0003-4819-158-7-201304020-00006 [published Online First: 2013/04/03]

108. SAS software version 9.2 [Computer program]. Copyright, SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA, 2009. [program], 2009.
109. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 2005;58(9):882-93. doi: 10.1016/j.jclinepi.2005.01.016
110. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282(11):1061-6.
111. Willis BH, Quigley M. The assessment of the quality of reporting of meta-analyses in diagnostic research: a systematic review. *BMC Med Res Methodol* 2011;11:163. doi: 10.1186/1471-2288-11-163
112. Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25. doi: 10.1186/1471-2288-3-25
113. Takwoingi Y, Guo B, Riley RD, et al. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Stat Methods Med Res* 2017;26(4):1896-911. doi: 10.1177/0962280215592269
114. Meys EMJ, Jeelof LS, Achten NMJ, et al. Estimating risk of malignancy in adnexal masses: external validation of the ADNEX model and comparison with other frequently used ultrasound methods. *Ultrasound Obstet Gynecol* 2017;49(6):784-92. doi: 10.1002/uog.17225 [published Online First: 2016/08/16]
115. NICE CG. Ovarian cancer: recognition and initial management UK: NICE; 2011 [updated November 2017. Available from: https://www.nice.org.uk/Guidance/cg122 accessed 24/2/19 2019.
116. Ferraro S, Braga F, Lanzoni M, et al. Serum human epididymis protein 4 vs carbohydrate antigen 125 for ovarian cancer diagnosis: a systematic review. *J Clin Pathol* 2013;66(4):273-81. doi: 10.1136/jclinpath-2012-201031 [published Online First: 2013/02/22]
117. Zhen S, Bian LH, Chang LL, et al. Comparison of serum human epididymis protein 4 and carbohydrate antigen 125 as markers in ovarian cancer: A meta-analysis. *Mol Clin Oncol* 2014;2(4):559-66. doi: 10.3892/mco.2014.279 [published Online First: 2014/06/19]
118. Li F, Tie R, Chang K, 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)

119. Kaijser J, Sayasneh A, Van Hoorde K, et al. Presurgical diagnosis of adnexal tumours using mathematical models and scoring systems: a systematic review and meta-analysis. *Hum Reprod Update* 2014;20(3):449-62. doi: 10.1093/humupd/dmt059
120. Wang J, Gao J, Yao H, et al. Diagnostic accuracy of serum HE4, CA125 and ROMA in patients with ovarian cancer: a meta-analysis. *Tumour Biol* 2014;35(6):6127-38. doi: 10.1007/s13277-014-1811-6 [published Online First: 2014/03/15]
121. Macedo AC, da Rosa MI, Lumertz S, et al. Accuracy of serum human epididymis protein 4 in ovarian cancer diagnosis: a systematic review and meta-analysis. *Int J Gynecol Cancer* 2014;24(7):1222-31. doi: 10.1097/IGC.0000000000000192 [published Online First: 2014/08/01]
122. Stukan M, Dudziak M, Ratajczak K, et al. Usefulness of diagnostic indices comprising clinical, sonographic, and biomarker data for discriminating benign from malignant ovarian masses. *J Ultrasound Med* 2015;34(2):207-17. doi: 10.7863/ultra.34.2.207 [published Online First: 2015/01/24]
123. Yu S, Yang HJ, Xie SQ, et al. Diagnostic value of HE4 for ovarian cancer: a meta-analysis. *Clin Chem Lab Med* 2012;50(8):1439-46. doi: 10.1515/cclm-2011-0477 [published Online First: 2012/08/08]
124. 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. *Eur J Obstet Gynecol Reprod Biol* 2013;167(1):81-5. doi: 10.1016/j.ejogrb.2012.10.036 [published Online First: 2012/12/12]
125. Geomini P, Kruitwagen R, Bremer GL, et al. The accuracy of risk scores in predicting ovarian malignancy: a systematic review. *Obstet Gynecol* 2009;113(2 Pt 1):384-94. doi: 10.1097/AOG.0b013e318195ad17
126. Medeiros LR, Rosa DD, da Rosa MI, et al. Accuracy of CA 125 in the diagnosis of ovarian tumors: A quantitative systematic review. *European Journal of Obstetrics Gynecology and Reproductive Biology* 2009;142(2):99-105.
127. Dodge JE, Covens AL, Lacchetti C, et al. Preoperative identification of a suspicious adnexal mass: a systematic review and meta-analysis. *Gynecol Oncol* 2012;126(1):157-66. doi: 10.1016/j.ygyno.2012.03.048
128. Wu L, Dai ZY, Qian YH, et al. Diagnostic value of serum human epididymis protein 4 (HE4) in ovarian carcinoma: a systematic review and meta-analysis. *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society* 2012;22(7):1106-12.
129. Liao XY, Huang GJ, Gao C, et al. A meta-analysis of serum cancer antigen 125 array for diagnosis of ovarian cancer in Chinese (Provisional abstract). *Database of Abstracts of Reviews of Effects* 2014; (2).

http://onlinelibrary.wiley.com/o/cochrane/cldare/articles/DARE-12014071537/frame.html.

130. Lin JY, Qin JB, Li XY, et al. Diagnostic value of human epididymis protein 4 compared with mesothelin for ovarian cancer: a systematic review and meta-analysis. *Asian Pac J Cancer Prev* 2012;13(11):5427-32. doi: 10.7314/apjcp.2012.13.11.5427 [published Online First: 2013/01/16]
131. McInnes MDF, Moher D, Thombs BD, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA* 2018;319(4):388-96. doi: 10.1001/jama.2017.19163 [published Online First: 2018/01/25]
132. Drapkin R, von Horsten HH, Lin Y, et al. Human epididymis protein 4 (HE4) is a secreted glycoprotein that is overexpressed by serous and endometrioid ovarian carcinomas. *Cancer Res* 2005;65(6):2162-9. doi: 10.1158/0008-5472.CAN-04-3924
133. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155(8):529-36. doi: 10.7326/0003-4819-155-8-201110180-00009
134. Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med* 2006;25(20):3443-57. doi: 10.1002/sim.2380 [published Online First: 2005/12/14]
135. Cochrane. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. : Cochrane 2013.
136. Anton C, Carvalho FM, Oliveira EI, et al. A comparison of CA125, HE4, risk ovarian malignancy algorithm (ROMA), and risk malignancy index (RMI) for the classification of ovarian masses. *Clinics (Sao Paulo)* 2012;67(5):437-41. doi: 10.6061/clinics/2012(05)06 [published Online First: 2012/06/06]
137. Stiekema A, Lok CA, Kenter GG, et al. A predictive model combining human epididymal protein 4 and radiologic features for the diagnosis of ovarian cancer. *Gynecol Oncol* 2014;132(3):573-7. doi: 10.1016/j.ygyno.2014.01.005
138. Karlsen MA, Sandhu N, Hogdall C, 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. *Gynecol Oncol* 2012;127(2):379-83. doi: 10.1016/j.ygyno.2012.07.106 [published Online First: 2012/07/28]
139. Van Gorp T, Cadron I, Despierre E, et al. HE4 and CA125 as a diagnostic test in ovarian cancer: prospective validation of the Risk of Ovarian Malignancy Algorithm. *Br J Cancer* 2011;104(5):863-70. doi: 10.1038/sj.bjc.6606092 [published Online First: 2011/02/10]

140. Van Gorp T, Veldman J, Van Calster B, 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. *Eur J Cancer* 2012;48(11):1649-56. doi: 10.1016/j.ejca.2011.12.003 [published Online First: 2012/01/10]

141. 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 Pac J Cancer Prev* 2014;15(5):1949-53. doi: 10.7314/apjcp.2014.15.5.1949 [published Online First: 2014/04/11]

142. Sayasneh A, Kaijser J, Preisler J, et al. A multicenter prospective external validation of the diagnostic performance of IOTA simple descriptors and rules to characterize ovarian masses. *Gynecol Oncol* 2013;130(1):140-6. doi: 10.1016/j.ygyno.2013.04.003 [published Online First: 2013/04/13]

143. Testa A, Kaijser J, Wynants L, et al. Strategies to diagnose ovarian cancer: new evidence from phase 3 of the multicentre international IOTA study. *Br J Cancer* 2014;111(4):680-8. doi: 10.1038/bjc.2014.333 [published Online First: 2014/06/18]

144. Di Legge A, Testa AC, Ameye L, 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 Obstet Gynecol* 2012;40(3):345-54. doi: 10.1002/uog.11167 [published Online First: 2012/05/23]

145. Sayasneh A, Wynants L, Preisler J, et al. Multicentre external validation of IOTA prediction models and RMI by operators with varied training. *Br J Cancer* 2013;108(12):2448-54. doi: 10.1038/bjc.2013.224 [published Online First: 2013/05/16]

146. Van Calster B, Timmerman D, Valentin L, 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* 2012;119(6):662-71. doi: 10.1111/j.1471-0528.2012.03297.x [published Online First: 2012/03/07]

147. Van Holsbeke C, Van Calster B, Bourne T, et al. External validation of diagnostic models to estimate the risk of malignancy in adnexal masses. *Clin Cancer Res* 2012;18(3):815-25. doi: 10.1158/1078-0432.CCR-11-0879 [published Online First: 2011/11/25]

148. Van Holsbeke C, Van Calster B, Valentin L, 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. *Clin Cancer Res* 2007;13(15 Pt 1):4440-7. doi: 10.1158/1078-0432.CCR-06-2958 [published Online First: 2007/08/03]

149. Yavuzcan A, Caglar M, Ozgu E, et al. Should cut-off values of the risk of malignancy index be changed for evaluation of adnexal masses in Asian and Pacific populations? *Asian Pac J Cancer Prev* 2013;14(9):5455-9. doi: 10.7314/apjcp.2013.14.9.5455 [published Online First: 2013/11/02]
150. Antovska SV, Bashevska N, Aleksioska N. Predictive values of the ultrasound parameters, CA-125 and risk of malignancy index in patients with ovarian cancer. *Klin Onkol* 2011;24(6):435-42. [published Online First: 2012/01/20]
151. Mol BW, Boll D, De Kanter M, et al. Distinguishing the benign and malignant adnexal mass: an external validation of prognostic models. *Gynecol Oncol* 2001;80(2):162-7. doi: 10.1006/gyno.2000.6052 [published Online First: 2001/02/13]
152. Ma S, Shen K, Lang J. A risk of malignancy index in preoperative diagnosis of ovarian cancer. *Chin Med J (Engl)* 2003;116(3):396-9. [published Online First: 2003/06/05]
153. Akturk E, Karaca RE, Alanbay I, et al. Comparison of four malignancy risk indices in the detection of malignant ovarian masses. *J Gynecol Oncol* 2011;22(3):177-82. doi: 10.3802/jgo.2011.22.3.177 [published Online First: 2011/10/15]
154. Andersen ES, Knudsen A, Rix P, et al. Risk of malignancy index in the preoperative evaluation of patients with adnexal masses. *Gynecol Oncol* 2003;90(1):109-12. doi: 10.1016/s0090-8258(03)00192-6 [published Online First: 2003/06/25]
155. Aslam N, Tailor A, Lawton F, et al. Prospective evaluation of three different models for the pre-operative diagnosis of ovarian cancer. *BJOG* 2000;107(11):1347-53. doi: 10.1111/j.1471-0528.2000.tb11646.x [published Online First: 2000/12/16]
156. Bailey J, Tailor A, Naik R, et al. Risk of malignancy index for referral of ovarian cancer cases to a tertiary center: does it identify the correct cases? *Int J Gynecol Cancer* 2006;16 Suppl 1:30-4. doi: 10.1111/j.1525-1438.2006.00468.x [published Online First: 2006/03/07]
157. Davies AP, Jacobs I, Woolas R, et al. The adnexal mass: benign or malignant? Evaluation of a risk of malignancy index. *Br J Obstet Gynaecol* 1993;100(10):927-31. doi: 10.1111/j.1471-0528.1993.tb15109.x [published Online First: 1993/10/01]
158. Dotlic J, Terzic M, Likic I, et al. Evaluation of adnexal masses: correlation between clinical, ultrasound and histopathological findings. *Vojnosanit Pregl* 2011;68(10):861-6. [published Online First: 2011/12/15]
159. Enakpene CA, Omigbodun AO, Goecke TW, et al. Preoperative evaluation and triage of women with suspicious adnexal masses using risk of malignancy index. *J Obstet Gynaecol Res* 2009;35(1):131-8. doi: 10.1111/j.1447-0756.2008.00869.x [published Online First: 2009/02/14]
160. Hakansson F, Hogdall EV, Nedergaard L, et al. Risk of malignancy index used as a diagnostic tool in a tertiary centre for patients with a pelvic mass. *Acta Obstet Gynecol*

Scand 2012;91(4):496-502. doi: 10.1111/j.1600-0412.2012.01359.x [published Online First: 2012/01/11]

161. Huchon C, Metzger U, Bats AS, et al. Value of three-dimensional contrast-enhanced power Doppler ultrasound for characterizing adnexal masses. *J Obstet Gynaecol Res* 2012;38(5):832-40. doi: 10.1111/j.1447-0756.2011.01785.x [published Online First: 2012/03/15]
162. Jacobs IJ, Rivera H, Oram DH, et al. Differential diagnosis of ovarian cancer with tumour markers CA 125, CA 15-3 and TAG 72.3. *Br J Obstet Gynaecol* 1993;100(12):1120-4. doi: 10.1111/j.1471-0528.1993.tb15177.x [published Online First: 1993/12/01]
163. Meray O, Turkcuoglu I, Meydanli MM, et al. Risk of malignancy index is not sensitive in detecting non-epithelial ovarian cancer and borderline ovarian tumor. *J Turk Ger Gynecol Assoc* 2010;11(1):22-6. [published Online First: 2010/01/01]
164. Morgante G, la Marca A, Ditto A, et al. Comparison of two malignancy risk indices based on serum CA125, ultrasound score and menopausal status in the diagnosis of ovarian masses. *Br J Obstet Gynaecol* 1999;106(6):524-7. doi: 10.1111/j.1471-0528.1999.tb08318.x [published Online First: 1999/07/30]
165. Radosa MP, Camara O, Vorwergk J, et al. Preoperative multimodal strategies for risk assessment of adnexal masses: analysis of 1362 cases in a gynecologic cancer center. *Int J Gynecol Cancer* 2011;21(6):1056-62. doi: 10.1097/IGC.0b013e3182187eb0 [published Online First: 2011/06/03]
166. Rossi A, Braghin C, Soldano F, et al. A proposal for a new scoring system to evaluate pelvic masses: Pelvic Masses Score (PMS). *Eur J Obstet Gynecol Reprod Biol* 2011;157(1):84-8. doi: 10.1016/j.ejogrb.2011.02.009 [published Online First: 2011/03/29]
167. Simsek HS, Tokmak A, Ozgu E, et al. Role of a risk of malignancy index in clinical approaches to adnexal masses. *Asian Pac J Cancer Prev* 2014;15(18):7793-7. doi: 10.7314/apjcp.2014.15.18.7793 [published Online First: 2014/10/09]
168. Terzic M, Dotlic J, Likic I, et al. Risk of malignancy index validity assessment in premenopausal and postmenopausal women with adnexal tumors. *Taiwan J Obstet Gynecol* 2013;52(2):253-7. doi: 10.1016/j.tjog.2013.04.017 [published Online First: 2013/08/07]
169. Tingulstad S, Hagen B, Skjeldestad FE, 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. *Br J Obstet Gynaecol* 1996;103(8):826-31. doi: 10.1111/j.1471-0528.1996.tb09882.x [published Online First: 1996/08/01]

170. Ulusoy S, Akbayir O, Numanoglu C, et al. The risk of malignancy index in discrimination of adnexal masses. *Int J Gynaecol Obstet* 2007;96(3):186-91. doi: 10.1016/j.ijgo.2006.10.006 [published Online First: 2007/02/07]
171. Arun-Muthuvel V, Jaya V. Pre-operative evaluation of ovarian tumors by risk of malignancy index, CA125 and ultrasound. *Asian Pac J Cancer Prev* 2014;15(6):2929-32. doi: 10.7314/apjcp.2014.15.6.2929 [published Online First: 2014/04/26]
172. Asif N, Sattar A, Dawood MM, et al. Pre-operative evaluation of ovarian mass: risk of malignancy index. *J Coll Physicians Surg Pak* 2004;14(3):128-31. doi: 03.2004/JCPSP.128131 [published Online First: 2004/07/02]
173. Dasari P CLPP, Sridhar MG. . Ultrasonographic scoring and risk of malignancy index in preoperative prediction of ovarian malignancy. . *Journal of Gynecologic Surgery* 2013;;29(2):61-4. doi: https://doi.org/10.1089/gyn.2012.0086
174. Manjunath AP, Pratapkumar, Sujatha K, et al. Comparison of three risk of malignancy indices in evaluation of pelvic masses. *Gynecol Oncol* 2001;81(2):225-9. doi: 10.1006/gyno.2001.6122 [published Online First: 2001/05/02]
175. Bouzari Z, Yazdani S, Ahmadi MH, et al. Comparison of three malignancy risk indices and CA-125 in the preoperative evaluation of patients with pelvic masses. *BMC Res Notes* 2011;4:206. doi: 10.1186/1756-0500-4-206 [published Online First: 2011/06/22]
176. Irshad F IM, 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.
177. Jabeen R KS, Naveed S. . Risk of Malignancy index in the preoperative evaluation of patients with ovarian masses. *Rawal Medical Journal* 2015 40(1):78-80.
178. Mohammed ABF AV, 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.
179. Obeidat BR, Amarin ZO, Latimer JA, et al. Risk of malignancy index in the preoperative evaluation of pelvic masses. *Int J Gynaecol Obstet* 2004;85(3):255-8. doi: 10.1016/j.ijgo.2003.10.009 [published Online First: 2004/05/18]
180. Moolthiya W, Yuenyao P. The risk of malignancy index (RMI) in diagnosis of ovarian malignancy. *Asian Pac J Cancer Prev* 2009;10(5):865-8. [published Online First: 2010/02/19]
181. Ong C, Biswas A, Choolani M, et al. Comparison of risk of malignancy indices in evaluating ovarian masses in a Southeast Asian population. *Singapore Med J* 2013;54(3):136-9. doi: 10.11622/smedj.2013046 [published Online First: 2013/04/03]
182. Yamamoto Y, Yamada R, Oguri H, et al. Comparison of four malignancy risk indices in the preoperative evaluation of patients with pelvic masses. *Eur J Obstet Gynecol*

Reprod Biol 2009;144(2):163-7. doi: 10.1016/j.ejogrb.2009.02.048 [published Online First: 2009/03/31]

183. Yamamoto Y TA, 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. doi: https://doi.org/10.1016/j.cogc.2014.11.001
184. Clarke SE, Grimshaw R, Rittenberg P, et al. Risk of malignancy index in the evaluation of patients with adnexal masses. *J Obstet Gynaecol Can* 2009;31(5):440-5. doi: 10.1016/s1701-2163(16)34175-5 [published Online First: 2009/07/17]
185. Kader Ali Mohan GR, Jaaback K, Proietto A, et al. Risk Malignancy Index (RMI) in patients with abnormal pelvic mass: Comparing RMI 1, 2 and 3 in an Australian population. *Aust N Z J Obstet Gynaecol* 2010;50(1):77-80. doi: 10.1111/j.1479-828X.2009.01105.x [published Online First: 2010/03/12]
186. Chan KK, Chen CA, Nam JH, et al. The use of HE4 in the prediction of ovarian cancer in Asian women with a pelvic mass. *Gynecol Oncol* 2013;128(2):239-44. doi: 10.1016/j.ygyno.2012.09.034 [published Online First: 2012/10/16]
187. Fujiwara H, Suzuki M, Takeshima N, 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. *Tumour Biol* 2015;36(2):1045-53. doi: 10.1007/s13277-014-2738-7 [published Online First: 2014/10/20]
188. Grenache DG, Heichman KA, Werner TL, et al. Clinical performance of two multi-marker blood tests for predicting malignancy in women with an adnexal mass. *Clin Chim Acta* 2015;438:358-63. doi: 10.1016/j.cca.2014.09.028 [published Online First: 2014/10/07]
189. Moore RG, McMeekin DS, Brown AK, et al. A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecol Oncol* 2009;112(1):40-6. doi: 10.1016/j.ygyno.2008.08.031 [published Online First: 2008/10/15]
190. Moore RG, Miller MC, Disilvestro P, et al. Evaluation of the diagnostic accuracy of the risk of ovarian malignancy algorithm in women with a pelvic mass. *Obstet Gynecol* 2011;118(2 Pt 1):280-8. doi: 10.1097/AOG.0b013e318224fce2 [published Online First: 2011/07/22]
191. Sandri MT, Bottari F, Franchi D, et al. Comparison of HE4, CA125 and ROMA algorithm in women with a pelvic mass: correlation with pathological outcome. *Gynecol Oncol* 2013;128(2):233-8. doi: 10.1016/j.ygyno.2012.11.026 [published Online First: 2012/12/04]

192. Partheen K, Kristjansdottir B, Sundfeldt K. Evaluation of ovarian cancer biomarkers HE4 and CA-125 in women presenting with a suspicious cystic ovarian mass. *J Gynecol Oncol* 2011;22(4):244-52. doi: 10.3802/jgo.2011.22.4.244 [published Online First: 2012/01/17]

193. Molina R, Escudero JM, Auge JM, et al. HE4 a novel tumour marker for ovarian cancer: comparison with CA 125 and ROMA algorithm in patients with gynaecological diseases. *Tumour Biol* 2011;32(6):1087-95. doi: 10.1007/s13277-011-0204-3 [published Online First: 2011/08/25]

194. Ortiz-Munoz B, Aznar-Oroval E, Garcia Garcia A, et al. HE4, Ca125 and ROMA algorithm for differential diagnosis between benign gynaecological diseases and ovarian cancer. *Tumour Biol* 2014;35(7):7249-58. doi: 10.1007/s13277-014-1945-6 [published Online First: 2014/04/29]

195. Montagnana M, Danese E, Ruzzenente O, 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? *Clin Chem Lab Med* 2011;49(3):521-5. doi: 10.1515/CCLM.2011.075 [published Online First: 2011/02/04]

196. Novotny Z, Presl J, Kucera R, et al. HE4 and ROMA index in Czech postmenopausal women. *Anticancer Res* 2012;32(9):4137-40. [published Online First: 2012/09/21]

197. Kadija S, Stefanovic A, Jeremic K, et al. The utility of human epididymal protein 4, cancer antigen 125, and risk for malignancy algorithm in ovarian cancer and endometriosis. *Int J Gynecol Cancer* 2012;22(2):238-44. doi: 10.1097/IGC.0b013e318234f852 [published Online First: 2012/01/05]

198. Kim YM, Whang DH, Park J, 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. *Clin Chem Lab Med* 2011;49(3):527-34. doi: 10.1515/CCLM.2011.085 [published Online First: 2011/02/16]

199. Chen WT, Gao X, Han XD, et al. HE4 as a serum biomarker for ROMA prediction and prognosis of epithelial ovarian cancer. *Asian Pac J Cancer Prev* 2014;15(1):101-5. doi: 10.7314/apjcp.2014.15.1.101 [published Online First: 2014/02/18]

200. Farzaneh F, Honarvar Z, Yaraghi M, et al. Preoperative evaluation of risk of ovarian malignancy algorithm index in prediction of malignancy of adnexal masses. *Iran Red Crescent Med J* 2014;16(6):e17185. doi: 10.5812/ircmj.17185 [published Online First: 2014/07/30]

201. Bandiera E, Romani C, Specchia C, 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 Epidemiol Biomarkers Prev* 2011;20(12):2496-506. doi: 10.1158/1055-9965.EPI-11-0635 [published Online First: 2011/10/27]

202. Chen X, Zhou H, Chen R, et al. Development of a multimarker assay for differential diagnosis of benign and malignant pelvic masses. *Clin Chim Acta* 2015;440:57-63. doi: 10.1016/j.cca.2014.11.013 [published Online First: 2014/12/03]
203. Nunes N, Ambler G, Hoo WL, et al. A prospective validation of the IOTA logistic regression models (LR1 and LR2) in comparison to subjective pattern recognition for the diagnosis of ovarian cancer. *Int J Gynecol Cancer* 2013;23(9):1583-9. doi: 10.1097/IGC.0b013e3182a6171a [published Online First: 2013/10/04]
204. Timmerman D, Ameye L, Fischerova D, et al. Simple ultrasound rules to distinguish between benign and malignant adnexal masses before surgery: prospective validation by IOTA group. *BMJ* 2010;341:c6839. doi: 10.1136/bmj.c6839 [published Online First: 2010/12/16]
205. Bristow RE, Smith A, Zhang Z, et al. Ovarian malignancy risk stratification of the adnexal mass using a multivariate index assay. *Gynecol Oncol* 2013;128(2):252-9. doi: 10.1016/j.ygyno.2012.11.022 [published Online First: 2012/11/28]
206. Dearking AC, Aletti GD, McGree ME, et al. How relevant are ACOG and SGO guidelines for referral of adnexal mass? *Obstet Gynecol* 2007;110(4):841-8. doi: 10.1097/01.AOG.0000267198.25223.bc [published Online First: 2007/10/02]
207. Im SS, Gordon AN, Buttin BM, et al. Validation of referral guidelines for women with pelvic masses. *Obstet Gynecol* 2005;105(1):35-41. doi: 10.1097/01.AOG.0000149159.69560.ef [published Online First: 2004/12/31]
208. Ware Miller R, Smith A, DeSimone CP, et al. Performance of the American College of Obstetricians and Gynecologists' ovarian tumor referral guidelines with a multivariate index assay. *Obstet Gynecol* 2011;117(6):1298-306. doi: 10.1097/AOG.0b013e31821b1d80 [published Online First: 2011/05/11]
209. Bristow RE, Hodeib M, Smith A, et al. Impact of a multivariate index assay on referral patterns for surgical management of an adnexal mass. *Am J Obstet Gynecol* 2013;209(6):581 e1-8. doi: 10.1016/j.ajog.2013.08.009 [published Online First: 2013/08/15]
210. Van Calster B, Van Hoorde K, Valentin L, 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. doi: 10.1136/bmj.g5920 [published Online First: 2014/10/17]
211. Ueland FR. A Perspective on Ovarian Cancer Biomarkers: Past, Present and Yet-To-Come. *Diagnostics (Basel)* 2017;7(1) doi: 10.3390/diagnostics7010014

212. Kobayashi E, Ueda Y, Matsuzaki S, et al. Biomarkers for screening, diagnosis, and monitoring of ovarian cancer. *Cancer Epidemiol Biomarkers Prev* 2012;21(11):1902-12. doi: 10.1158/1055-9965.EPI-12-0646

213. Andersen MR, Lowe KA, Goff BA. Value of symptom-triggered diagnostic evaluation for ovarian cancer. *Obstet Gynecol* 2014;123(1):73-9. doi: 10.1097/AOG.0000000000000051

214. Holcomb K, Vucetic Z, Miller MC, et al. Human epididymis protein 4 offers superior specificity in the differentiation of benign and malignant adnexal masses in premenopausal women. *Am J Obstet Gynecol* 2011;205(4):358 e1-6. doi: 10.1016/j.ajog.2011.05.017 [published Online First: 2011/07/05]

215. AvŞar FŞ, Emel; Isenlik, Sıtkı; Uzun, Mustafa; Özyiğit, Eda; ßinkaya, Ayşegül; Gümüş, Mehmet; ×elen, Emine; üstüner, Işık; Keskin, Levent; Aydoğmuş, Serpil. The Value of Ultrasonography and Tumor Markers in the Differential Diagnosis of Postmenopausal Ovarian Cysts. *Turkish Journal of Medical Sciences* 2009, Ip. ; 39 (2,):179-83.

216. Botsis D, Kassanos D, Kalogirou D, et al. Transvaginal color Doppler and CA125 as tools in the differential diagnosis of postmenopausal ovarian masses. *Maturitas* 1997;26(3):203-9. doi: 10.1016/s0378-5122(96)01097-3 [published Online First: 1997/04/01]

217. Engelen MJ, Bongaerts AH, Sluiter WJ, et al. Distinguishing benign and malignant pelvic masses: the value of different diagnostic methods in everyday clinical practice. *Eur J Obstet Gynecol Reprod Biol* 2008;136(1):94-101. doi: 10.1016/j.ejogrb.2006.10.004 [published Online First: 2006/11/23]

218. Kusnetzoff D GD, Damonte C, Sananes C, Giaroli A, di Paolo G, Sardi Differential diagnosis of pelvic masses: Usefulness of CA125, transvaginal sonography and echo-Doppler. *International Journal of Gynecological Cancer* 2002;8(4):315-21. doi: https://doi.org/10.1046/j.1525-1438.1998.97107.x

219. Maggino T, Gadducci A, D'Addario V, et al. Prospective multicenter study on CA 125 in postmenopausal pelvic masses. *Gynecol Oncol* 1994;54(2):117-23. doi: 10.1006/gyno.1994.1179 [published Online First: 1994/08/01]

220. Antonic J, Rakar S. Colour and pulsed Doppler US and tumour marker CA 125 in differentiation between benign and malignant ovarian masses. *Anticancer Res* 1995;15(4):1527-32. [published Online First: 1995/07/01]

221. Schutter EM, Kenemans P, Sohn C, et al. Diagnostic value of pelvic examination, ultrasound, and serum CA 125 in postmenopausal women with a pelvic mass. An international multicenter study. *Cancer* 1994;74(4):1398-406. doi: 10.1002/1097-0142(19940815)74:4<1398::aid-cncr2820740433>3.0.co;2-j [published Online First: 1994/08/15]

222. Franchi M, Beretta P, Ghezzi F, et al. Diagnosis of pelvic masses with transabdominal color Doppler, CA 125 and ultrasonography. *Acta Obstet Gynecol Scand* 1995;74(9):734-9. doi: 10.3109/00016349509021184 [published Online First: 1995/10/01]

223. Gadducci A, Ferdeghini M, Prontera C, et al. The concomitant determination of different tumor markers in patients with epithelial ovarian cancer and benign ovarian masses: relevance for differential diagnosis. *Gynecol Oncol* 1992;44(2):147-54. doi: 10.1016/0090-8258(92)90030-m [published Online First: 1992/02/01]

224. Gadducci A, Baicchi U, Marrai R, et al. Preoperative evaluation of D-dimer and CA 125 levels in differentiating benign from malignant ovarian masses. *Gynecol Oncol* 1996;60(2):197-202. doi: 10.1006/gyno.1996.0025 [published Online First: 1996/02/01]

225. Strigini FA, Gadducci A, Del Bravo B, et al. Differential diagnosis of adnexal masses with transvaginal sonography, color flow imaging, and serum CA 125 assay in pre- and postmenopausal women. *Gynecol Oncol* 1996;61(1):68-72. doi: 10.1006/gyno.1996.0098 [published Online First: 1996/04/01]

226. Van Calster B, Timmerman D, Bourne T, et al. Discrimination between benign and malignant adnexal masses by specialist ultrasound examination versus serum CA-125. *J Natl Cancer Inst* 2007;99(22):1706-14. doi: 10.1093/jnci/djm199 [published Online First: 2007/11/15]

227. Smikle CB, Lunt CC, Hankins GD. Clinical predictors in the evaluation of a pelvic mass. *Mil Med* 1995;160(5):233-5. [published Online First: 1995/05/01]

228. Doed JR, Quinn MA, Rome R, et al. Women with a pelvic mass--when to perform an ultrasound. *Aust N Z J Obstet Gynaecol* 1993;33(4):404-7. doi: 10.1111/j.1479-828x.1993.tb02122.x [published Online First: 1993/11/01]

229. Kuesel AC, Kroft T, Prefontaine M, et al. Lipoprotein(a) and CA125 levels in the plasma of patients with benign and malignant ovarian disease. *Int J Cancer* 1992;52(3):341-6. doi: 10.1002/ijc.2910520302 [published Online First: 1992/09/30]

230. Benjapibal M, Neungton C. Pre-operative prediction of serum CA125 level in women with ovarian masses. *J Med Assoc Thai* 2007;90(10):1986-91. [published Online First: 2007/11/29]

231. Sundar S, Rick C, Dowling F, et al. Refining Ovarian Cancer Test accuracy Scores (ROCkeTS): protocol for a prospective longitudinal test accuracy study to validate new risk scores in women with symptoms of suspected ovarian cancer. *BMJ Open* 2016;6(8):e010333. doi: 10.1136/bmjopen-2015-010333

232. RCOG RCoOaG. The Management of Ovarian Cysts in PostmenopausalWomen 2016 [cited 2020 2.2.20]. Available from:

https://www.rcog.org.uk/globalassets/documents/guidelines/green-top-guidelines/gtg_34.pdf accessed 2.2.20 2020.

233. Hamilton W, Peters TJ, Bankhead C, et al. Risk of ovarian cancer in women with symptoms in primary care: population based case-control study. *BMJ* 2009;339:b2998. doi: 10.1136/bmj.b2998 [published Online First: 2009/08/27]

234. Skirnisdottir I, Garmo H, Wilander E, et al. Borderline ovarian tumors in Sweden 1960-2005: trends in incidence and age at diagnosis compared to ovarian cancer. *Int J Cancer* 2008;123(8):1897-901. doi: 10.1002/ijc.23724 [published Online First: 2008/07/29]