Diagnostic biomarkers for ovarian cancer: a systematic review

Jannick Achtergaele, KU Leuven

Promotor: Jan Verbakel, KU Leuven

Master of Family Medicine

Masterproef Huisartsgeneeskunde
Academiejaar: 2024 – 2025

©Copyright by KU Leuven

Deze masterproef is een examendocument dat niet werd gecorrigeerd voor eventueel vastgestelde fouten. Met uitzondering van het door de wet toegestaan gebruik, is zonder voorafgaande schriftelijke toestemming van de auteur(s) het overnemen, kopiëren, gebruiken of realiseren van deze uitgave of gedeelten ervan verboden. Voor aanvragen tot of informatie in verband met het overnemen en/of gebruik van gedeelten uit deze publicatie, wendt u zich tot de KU Leuven, Faculteit Geneeskunde, Campus Gasthuisberg O&N2, Herestraat 49 – bus 400, 3000 Leuven.

Voorafgaande schriftelijke toestemming van de promotor(en) is eveneens vereist voor het aanwenden van de in dit afstudeerwerk beschreven (originele) methoden, producten en programma's voor commercieel nut en voor de inzending van deze publicatie ter deelname aan wetenschappelijke prijzen of wedstrijden

INDEX

MANUSCRIPT **4**

Abstract *4*

Article *6*

Introduction *6*

Methods *8*

Results *11*

Discussion *13*

Figures *17*

Authors *24*

Conflict of interest *24*

Financial disclosures *24*

Funding *24*

MANUSCRIPT

Abstract

Background: Ovarian cancer is the most lethal gynaecologic tumour, with mortality rates of 10-30% in stage I-II disease and up to 70% in stage III-IV. The lack of an effective screening strategy, combined with the observation that approximately 80% of cases are diagnosed at an advanced stage, underscores the importance and potential benefit of achieving timely and accurate diagnosis. Although diagnosis of ovarian cancer remains histological, biomarkers play an important role in the diagnostic work-up. Currently, most important biomarkers are Cancer Antigen 125 (CA125), Human Epididymis Protein 4 (HE4), multimodality tools Risk of Malignancy Index (RMI) and Risk of Ovarian Malignancy Algorithm (ROMA).

Objective: To conduct a systematic review of the current literature on major biomarkers involved in the diagnosis of ovarian cancer and identify which demonstrate the highest sensitivity and specificity. This could provide the groundwork for establishing generalized screening algorithms and thus reduce mortality through early diagnosis.

Methods: Medline and Embase databases were searched for publications that included MESH-terms “Ovarian cancer”, “CA-125”, “HE4”, “ROMA” and “RMI”. Of 923 articles identified, 9 high quality papers met the inclusion criteria, after QUADAS-2 quality assessment. Data on sensitivity and specificity for biomarkers were extracted and compared across the included articles.

Results: In screening trials in asymptomatic, postmenopausal women, CA125 achieved a pooled sensitivity and specificity of 74% and 97% respectively, with a PPV of only 6% due to low disease prevalence. No other screening tools could be compared, because of a lack of data and high heterogeneity between study designs. For preoperative evaluation of ovarian masses, CA125 and ROMA offered the highest sensitivity of 80% and 77% respectively, with a specificity of 80% and 83%. HE4 had a lower sensitivity of 61%, but a high specificity of 93%, whereas RMI showed a moderate sensitivity of 70%, with a specificity of 89%.

Conclusion: CA125 remains a cornerstone biomarker for ovarian cancer, but its limited sensitivity in women without underlying risk factors underscores the need of better screening modalities, as it results in a high number of false negative cases, potentially causing a delay in ovarian cancer diagnosis. In women with ovarian masses, HE4 and combined algorithms (ROMA, RMI) improve diagnostic accuracy, but are not sufficient as stand-alone screening or differentiation tools, as they still result in a high number of false negative cases. This data underscores the need for novel biomarkers or multimodal strategies to achieve effective early diagnosis and reduce overall ovarian cancer mortality.

Abstract

Achtergrond: Ovariumcarcinoom is de meest dodelijke gynaecologische maligniteit, met mortaliteitscijfers variërend van 10–30% in stadium I–II tot meer dan 70% in stadium III–IV. Het ontbreken van een effectieve screeningsstrategie en het feit dat ongeveer 80% van de initiële diagnosen in een gevorderd stadium wordt gesteld, maken vroege en efficiënte detectie bijzonder wenselijk. Hoewel de definitieve diagnose van ovariumkanker histopathologisch wordt gesteld, spelen biomerkers een belangrijke rol in het diagnostisch traject. Op dit moment zijn de meest gebruikte biomerkers Cancer Antigen 125 (CA125), Human Epididymis Protein 4 (HE4), en de gecombineerde risicobeoordelingsmodellen zoals de Risk of Malignancy Index (RMI) en het Risk of Ovarian Malignancy Algorithm (ROMA).

Doelstelling: Het uitvoeren van een systematisch literatuuronderzoek naar de voornaamste biomerkers die betrokken zijn bij de diagnose van ovariumkanker, met als doel de biomerkers met de hoogste sensitiviteit en specificiteit te identificeren. Dit kan een basis vormen voor de ontwikkeling van gestandaardiseerde screeningsalgoritmen, met als uiteindelijke doel het reduceren van mortaliteit door middel van vroegtijdige diagnose.

Methodologie: In de databanken Medline en Embase werd gezocht naar publicaties die de MeSH-termen “Ovarian cancer”, “CA-125”, “HE4”, “ROMA” en “RMI” bevatten. Van de 923 geïdentificeerde artikels voldeden er na kwaliteitsbeoordeling met het QUADAS-2 instrument in totaal 9 hoogkwalitatieve studies aan de inclusiecriteria. Gegevens over sensitiviteit en specificiteit van de onderzochte biomerkers werden geëxtraheerd en onderling vergeleken.

Resultaten: In screeningsonderzoeken bij asymptomatische, postmenopauzale vrouwen behaalde CA125 een gepoolde sensitiviteit van 74% en specificiteit van 97%, maar met een positieve voorspellende waarde van slechts 6% door de lage prevalentie van de ziekte. Andere screeningsmethoden konden niet vergeleken worden wegens een gebrek aan uniforme data en aanzienlijke heterogeniteit in studiedesigns. Voor de preoperatieve evaluatie van ovariummassa’s toonden CA125 en ROMA de hoogste sensitiviteit (respectievelijk 80% en 77%), met een specificiteit van 80% en 83% respectievelijk. HE4 vertoonde een lagere sensitiviteit van 61%, maar wel een hoge specificiteit van 93%. RMI behaalde een matige sensitiviteit van 70% en een specificiteit van 89%.

Conclusie: CA125 blijft een hoeksteen in de diagnostiek van ovariumkanker, maar de beperkte sensitiviteit bij vrouwen zonder verhoogd risico benadrukt de nood aan betere screeningsstrategieën. De hoge frequentie van vals negatieve resultaten kan leiden tot vertraging in de diagnose en behandeling. Bij vrouwen met een ovariummassa verhogen HE4 en gecombineerde algoritmen zoals ROMA en RMI de diagnostische accuraatheid, maar zijn ze op zichzelf onvoldoende betrouwbaar als screenings- of differentiatietool. Deze bevindingen benadrukken de dringende behoefte aan nieuwe biomerkers of multimodale benaderingen om een effectieve vroegtijdige detectie mogelijk te maken en zo de mortaliteit ten gevolge van ovariumkanker te reduceren.

Article

1. Introduction

Ovarian cancer (OC) is a major global health problem and is one of the deadliest gynaecologic malignancies in the world. Despite advances in treatment, overall mortality for ovarian cancer remains high with rates up to 10-30% in stage I-II and up to 70% in stage III-IV, underscoring the urgent need for effective screening strategies to detect the disease at an earlier stage, when treatment outcomes are more favourable(1). OC is notoriously difficult to recognize early in primary care. It often presents with vague, non-specific symptoms, such as persistent abdominal bloating, pelvic or abdominal pain, early satiety, or urinary urgency, which overlap with benign conditions(2). This nonspecific presentation contributes to diagnostic delays; for example, one review found that patients under 55 years old were more likely to have had three or more GP consultations before eventual referral and diagnosis of ovarian cancer(3). This makes OC often labelled the "silent killer" because it is asymptomatic in the early stages, leading to late diagnosis and poor prognosis.

Currently, diagnosis of OC remains multiphasic: initial evaluation, including complete physical examination, imaging (usually transvaginal ultrasound of the pelvis) and laboratory tests to determine if there is an adnexal mass or elevated tumour markers. Symptom-triggered testing with tumor biomarkers has become an established strategy in primary care. At present, the National Institute for Health and Care Excellence (NICE) advises that women with persisting or unexplained symptoms suggestive of possible OC should undergo a serum testing of biomarkers in primary care(4). The gold standard of OC diagnosis remains histological, either after surgery or biopsy.

OC represents a heterogeneous group of malignant neoplasms arising from the ovary. Histologically, ovarian cancers are classified into three major categories: epithelial tumors, germ cell tumors, and sex cord stromal tumors. Of these, epithelial ovarian cancers (EOCs) are by far the most prevalent, accounting for approximately 90% of all cases(5). EOCs are further subdivided into distinct histological subtypes, including high-grade serous carcinoma, low-grade serous carcinoma, endometrioid carcinoma, mucinous carcinoma, and clear cell carcinoma. These subtypes differ not only in their morphological and clinical features, but also in their molecular pathogenesis and response to treatment. In this systematic review we mainly focus on EOC.

Screening for OC could play a crucial role in its early detection, but unlike other types of cancer, such as breast or cervical cancer, there is currently no consensus on routine screening for OC. To lower OC mortality, a screening program must be able to detect the disease at an early stage, as treatment outcomes are strongly influenced by the stage at which it is diagnosed(6). The absence of a definitive screening method for OC stems from several factors, including the complex pathogenesis and heterogeneity of the disease, its relative rarity, and the difficulty in identifying reliable biomarkers for early detection. The main concern is that previous screening programs targeting the general population, i.e. asymptomatic women at average risk, have not clearly demonstrated a significant reduction in mortality. Another concern is that screening could be harmful due to large numbers of false positives, causing psychological stress and potential adverse events from biopsy or surgery(7,8).

OC screening is conducted with a combination of blood biomarkers, specifically cancer antigen 125 (CA125, also called mucin 16 [MUC16]) and human epididymis protein 4 (HE4) or its synonym WAP four-disulfide core domain protein 2 (WFDC2), which can be combined with pelvic ultrasound and pre- or postmenopausal status. CA-125 and HE4 are circulating tumour biomarkers that can be measured in peripheral blood. Although CA125 and HE4 are the best available diagnostic biomarkers for OC, they lack sensitivity and specificity, resulting in a high number of false positives and false negatives(9). These biomarkers can also be integrated with clinical parameters to assess the likelihood of malignancy after an ovarian mass was detected, for example using the Risk of Malignancy Index (RMI) and Risk of Ovarian Malignancy Algorithm (ROMA) (10,11). RMI

incorporates the serum CA125 level, ultrasound imaging findings and menopausal status into a single risk score, while ROMA combines the serum levels of CA125 and HE4 with menopausal status to stratify a patient's risk of OC. The Risk of Ovarian Cancer Algorithm or ROCA is a longitudinal screening strategy that incorporates serial CA125 measurements and detects change-points to improve early detection of ovarian cancer(12).

As mentioned, these biomarkers can be utilized in 2 different strategies. They are primarily used in a preoperative setting as a tool to further differentiate adnexal masses into benign of malignant categories but can also be used in the screening of asymptomatic women. This study will mainly focus on the latter, given the impact it could have on diagnosing OC in an early stage, potentially lowering mortality rates in the general population.

Given the major impact of ovarian cancer on women's health and the limited opportunities for early detection, a comprehensive evaluation of existing screening methods is essential. A systematic review provides an evidence-based synthesis of the available literature, allowing a thorough examination of efficacy, benefits, and potential drawbacks of various ovarian cancer screening methods.

2. Methods

Protocol and registration

This systematic review is part of an assignment necessary for the license of a general practitioner. The review protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) number CRD42024576173 before the study selection began, to guard against methodological bias and to promote transparency of the intended methods. The review followed the selection process proposed by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for transparent reporting.

Search Strategy

A comprehensive search strategy was performed between August 2024 and April 2025 on MEDLINE (via PubMed) and EMBASE using various synonyms and related terms, to encompass a broad scope of this literature search. The following MeSH terms and phrases were used: “Ovarian Neoplasms” AND “Diagnosis” AND “Sensitivity” AND “Specificity” AND “CA-125” (and its synonym cancer antigen 125) OR “WFDC2 protein, human” (and its synonym HE4) OR “ROMA” OR “RMI”.

Eligibility Criteria

Non-English articles and studies published before 2000 were excluded to reflect modern technology, clinical practice, patient populations, and higher methodological quality. Studies were eligible for inclusion if they were primary diagnostic accuracy studies (prospective or retrospective cohort or cross-sectional designs) evaluating the performance of at least one of the specified index tests for detecting ovarian cancer. Inclusion criteria included women above the age of 18 years old, with or without an adnexal mass or other clinical suspicion of ovarian cancer, who underwent both an index test assessment and a reference standard diagnostic confirmation. Each included study was required to report sufficient data to construct 2×2 contingency tables (i.e., true positives, false positives, true negatives, false negatives) for the index tests against the reference standard, enabling calculation of sensitivity and specificity. Minimal sample size required for subset groups was n > 50. There were no restrictions on menopausal status of patients. Exclusion criteria included borderline or non-EOC tumours, pregnant women, male or paediatric cases.

Study Selection

First, all studies were screened based on title and abstracts by one reviewer to identify articles that potentially met the inclusion criteria. At this stage, papers that did not met inclusion criteria were excluded. Second, the full text of all articles that passed the initial screening (or that remained uncertain) was examined in detail against the predefined eligibility criteria. Duplicate studies or studies using duplicate data were excluded (**Supplementary table 1**). The results of the study selection process are summarized in a PRISMA flow diagram (**Figure 1**), which depicts the number of records identified, screened, assessed for eligibility, and ultimately included in the review.

Index Tests

We aimed to review four diagnostic modalities commonly used for screening or for the preoperative assessment of ovarian masses: serum CA125, serum HE4, RMI, and ROMA. Cutoff values were fixed to increase comparability between studies and thus index tests were deemed positive if: CA125 level was above 35 U/mL and HE4 level for premenopausal women > 70 pmol/L or postmenopausal women > 140 pmol/L. Thresholds for ROMA scores were set at ≥ 13.1% for premenopausal women and at ≥ 27.7% for postmenopausal women. A RMI score threshold was set at > 200.

Reference Test

The reference or golden standard for diagnosing the presence or absence of ovarian malignancy was obtained by histopathological examination of surgical specimens. For patients who did not undergo immediate surgery, a follow-up strategy with imaging was used in some studies to confirm

that no malignancy became evident over time. When such follow-up was employed as the reference standard for test-negative cases, the study was included if follow-up duration was adequate (typically at least 12 months) to reasonably exclude a missed malignancy.

Data Extraction

The following data was collected from each study: study identifiers (first author, publication year, journal or source), study design, participant characteristics (number of participants, age range or mean age, menopausal status distribution, inclusion/exclusion criteria of the study, and any notable clinical features of the cohort), details of index test implementation and outcome data for diagnostic performance (**Supplementary table 1**). For each index test within each study, raw data were extracted to fill a 2x2 table. From these counts, we recorded or calculated sensitivity, specificity and other metrics as needed. If certain data was incomplete, a derivation or calculation was made from the available data. If an included study evaluated multiple index tests or multiple thresholds, data for each relevant test or threshold were extracted separately.

Assessment of Methodological Quality and Risk of Bias

The methodological quality of each included study was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2), assessing the quality of each study included for biases in four domains: patient selection, index test, reference test and flow and timing of the study. These bias and concerns regarding applicability were assessed and categorized as low, high, or unclear. This was performed to facilitate the exploration of potential sources of heterogeneity.

The evaluation was conducted in accordance with established quality criteria, which included detailed examination of the inclusion and exclusion criteria for patient selection, the methodology and execution of the index test, the nature and validity of the reference standard, whether blinding of the assessors to index test results was implemented, and the time interval between administration of the index test and application of the reference standard.

Outcomes

Primary outcomes were the key diagnostic accuracy measures of each index test for detecting ovarian cancer. Both sensitivity and specificity were considered crucial measures of performance and were the main focus of data synthesis. Secondary outcomes included additional diagnostic accuracy metrics derived from the 2×2 tables. These comprised the positive predictive value (PPV) and negative predictive value (NPV) of each test, which indicate the probability of malignancy given a positive test and the probability of benign disease given a negative test, respectively. If available, outcomes were stratified by subgroup (e.g., premenopausal vs postmenopausal women) in the analysis phase to explore differences in test performance.

Data Synthesis and Analysis

We performed a meta-analysis of the diagnostic performance of the index tests, provided that the included studies were sufficiently comparable in terms of design, patient population, and reported outcomes. For each index test (CA125, HE4, RMI, and ROMA), we first examined the extracted data by plotting the estimates of sensitivity and specificity from each study on forest plots and in Heterogeneity was assessed using the tau-squared (Tau^2) test, the chi-squared (Chi^2) test and the I square (I^2) test. Heterogeneity was considered low if I^2 was <25%, moderate if I^2 was between 25% and 75% and high if I^2 was > 75%. Statistical analyses were performed using GraphPad Prism 10.4.2 and MetaAnalysisOnline.com(13).

Subgroup and sensitivity analyses were pre-specified to explore potential sources of heterogeneity. We planned subgroup analyses by menopausal status (considering studies or subsets of participants that were premenopausal vs. postmenopausal) because the performance of CA125

and related indices is known to differ in these groups. If enough studies reported separate results by menopausal status, we analysed those strata separately.

3. Results

Between April 2024 and April 2025, we reviewed current literature and found 520 articles on PubMed and 403 Embase. Abstracts were screened on relevance, study design and quality. We matched 112 studies to the QUADAS-2 criteria and finally 9 studies that met our inclusion/exclusion criteria were included as shown in the PRISMA flow-diagram (**Figure 1**). These included 3 randomized controlled trials(14–16), 4 cohort studies(17–20), 1 systematic review(21) and 1 meta-analysis(22) (**Table 1**).

Figure 2 shows QUADAS-2 analysis of the included studies. Reference standard, flow and timing domains were robust and were low risk in most studies. Patient selection had the highest risk of bias overall due to the heterogeneity in study designs and those included in the systematic review and meta-analysis. While most studies were low concern for index test bias, one had selected optimal cut-offs from the study data itself(19). Overall, most studies were at low risk of bias in each domain, with a few notable exceptions as described above.

Characteristics of included studies

Each of the three randomized controlled trials assessed the utility of biomarkers within a screening framework. Given the low incidence of OC in women, large populations were screened. Follow-up ranged from 7 to 13 years to ensure no arising OC in the control group were missed. In the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)(14) postmenopausal women were recruited from 13 medical centres and allocated to annual multimodal screening (MMS) including serum CA125 with detection above baseline concentration using ROCA, annual transvaginal ultrasound screening (USS), or no screening. The primary outcome was mortality due to OC. In the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial(15) to determine the effects of screening of OC on mortality, postmenopausal women randomly received screening using CA125 and transvaginal ultrasound or no screening. In the Shizuoka Cohort Study of Ovarian Cancer Screening (SCSOCS)(16), postmenopausal women either received screening with physical examination, pelvic ultrasound and CA125 or no screening.

All cohort studies were retrospective, excluding Gentry-Maharaj(17) being a post-factum analysis of HE4 performance in UKCTOCS. Together with Funston et al(20), healthy women were used as a control group. However, Funston et al(20) did include borderline tumours. Sagi-Dain et al(18) and Kim et al(19) used women with benign ovarian tumours as a control group. Given their retrospective nature, selection and information bias are prudent.

Chacon et al(21) reviewed the diagnostic efficacy of ROMA versus RMI in differentiation of adnexal masses. High risk of bias was reported in some studies due to knowledge of reference test results before index testing and lack of reporting whether pathologists were blinded from index tests. An important note is that in one study the threshold for RMI was 250, contrary to our preset threshold of 200, and for ROMA multiple studies had different cutoffs.

Suri et al(22) analysed the diagnostic performance of ROMA and thus CA125 and HE4 in pre- or postmenopausal women with data from 32 studies. Women with benign tumours served as the control group. Unfortunately, there was no sub-group analysis for CA125 and HE4 in pre- or postmenopausal women due to inadequate data.

Diagnostic performance of biomarkers used in a screening strategy

Figures 3 and **4** show separate and pooled results for CA125 sensitivity and specificity using a random effects model. Overall, 3 studies were analysed with a total of 120.569 postmenopausal women(14–16). A significant heterogeneity was detected ($p<0.01$), suggesting differences in study design. Analysis showed a pooled sensitivity, specificity, PPV and NPV for detecting OC in postmenopausal women of 73.89% [CI 95%, 56.96-90.82%], 96.95% [CI 95%, 93.86-100%], 6.1% [CI 95%, 0.7-32.8%] and 99.98% [CI 95%, 99.95-99.99%] respectively. PPV and NPV were

calculated at a pooled study prevalence of 0.092%. Unfortunately, since none of the included studies used HE4, ROMA or RMI as a screening strategy, a comparison could not be made.

Diagnostic performance of biomarkers used in pre-operative assessment of adnexal mass

Figure 5 illustrates the sensitivity and specificity of different diagnostic methods studied for the preoperative differentiation of ovarian masses(17–22). Mean sensitivity for CA125 was 79.8% [CI 95%, 68.7–90.9%] and the mean specificity was 80.3% [CI 95%, 65.3–95.2%]. For HE4, mean sensitivity and specificity were 61.3% [CI 95%, 35.8–86.8%] and 92.7% [CI 95%, 81.0–104.5%] respectively. Mean sensitivity for ROMA was 77.4% [CI 95%, 63.0–91.8%] and specificity 83.6% [CI 95%, 72.9–94.2%]. RMI demonstrated a mean sensitivity of 69.8% [CI 95%, -2.6–142.2%] and mean specificity of 88.7% [CI 95%, 71.5–105.8%].

To account for the heterogeneity caused by menopausal status of women with adnexal masses, we performed subgroup analyses for the sensitivity and specificity of CA125 and ROMA (**Figure 6A-D**). For the other screening methods, HE4 and RMI, this was not feasible due to the limited data available in the studies included in this systematic review. The pooled sensitivity of ROMA screening in premenopausal women was 77.8% [CI 95%, 76.95–78.67%], compared to 81.9% [CI 95%, 80.31–83.44%] in postmenopausal women. This difference was statistically significant between subgroups, indicating that part of the observed heterogeneity may be attributed to differences in age or menopausal status. The pooled sensitivity of CA125 did not differ significantly between premenopausal and postmenopausal women. However, this may be attributed to the limited number of studies available for comparison. Similarly, subgroup analyses did not reveal any notable differences in the specificity of CA125 or ROMA between pre- and postmenopausal women.

4. Discussion

We investigated the diagnostic accuracy of different biomarkers as a screening tool for OC in two settings, either as a screening strategy in asymptomatic postmenopausal women, or as predictive tool in women with adnexal masses. If used as a screening strategy in detection of ovarian cancer in postmenopausal women, CA125 achieved 74% sensitivity and 97% specificity, but a positive predictive value of only 6% due to low disease prevalence in the included articles. In women with an adnexal mass, CA125 and the combined ROMA algorithm had the highest sensitivity of 80% and 74% respectively, while HE4 had the highest specificity of 93%. RMI showed moderate sensitivity of 70% and similarly high specificity of 89%. No single test is ideal and even the combination of multiple biomarkers or the application of risk algorithms such as ROMA and RMI does not result in significantly improved overall diagnostic accuracy.

HE4's higher specificity but lower sensitivity relative to CA125 could be explained by their biological differences. CA125 often rises in several benign conditions (e.g. endometriosis, peritonitis and heart failure) whereas HE4 rarely does, yielding fewer false positives. Conversely, some ovarian cancers, especially early-stage or certain subtypes, do not secrete HE4, so HE4 alone can miss cases that could be detected using CA125. This complementarity is why combining the two markers could improve performance. However, this could not clearly be demonstrated in our study analysing ROMA performance.

When comparing the two risk indices RMI and ROMA, we found they perform similarly overall. RMI, first introduced by Jacobs et al. in 1990(11), combines CA125, ultrasound, and menopausal status and is valued for high specificity but has modest sensitivity, partly due to reliance on ultrasound and the limitations of CA125 in early disease. The newer ROMA algorithm, developed by Moore et al. in 2011(10), was designed to improve sensitivity by adding HE4. As our results confirm, ROMA achieves higher sensitivity than RMI, while retaining comparable specificity. ROMA can thus serve as a useful alternative to RMI, in centres without specialized ultrasound. In our analysis, ROMA's overall accuracy was similar to RMI.

Given the substantial heterogeneity among the included studies, the results should be interpreted with caution. This could be partially explained due to differences in age, ethnicity, subtypes of OC and differences in biomarker testing. For example, CA125 and HE4 testing can be performed using different methods: Enzyme-linked immunosorbent assay (ELISA), Radioimmunoassay (RIA) and Chemiluminescence immunoassay (CLIA), potentially leading to heterogeneous results.

From a general practice perspective, the early recognition of ovarian cancer and appropriate use of diagnostic biomarkers are pivotal in reducing delays to diagnosis. The role of the GP is central: it is often the GP who first evaluates a woman's vague abdominal or pelvic symptoms and decides whether to initiate investigations for a possible malignancy. This thesis highlights the importance of GPs maintaining a high index of suspicion for ovarian cancer in patients with persistent, unexplained symptoms. Timely action at the primary care level, ordering a CA125 test and, if indicated, arranging a prompt ultrasound can expedite the diagnostic pathway and potentially lead to detection at an earlier stage, when treatment outcomes are significantly better. Simultaneously, there is a nuanced balance GPs must navigate in, between facilitating early cancer detection and minimizing unwarranted diagnostic procedures. Considering CA125's limitations, clinicians should exercise caution to avoid deriving false reassurance from a normal test result as shown in this review. Implicating that a normal test result should not delay further investigation when clinical suspicion remains.

Strengths

This review has several strengths. We conducted a comprehensive literature search and applied strict quality criteria (including QUADAS-2) to ensure only high-quality studies were included.

Moreover, by covering both screening trials and clinical diagnostic studies, we provide a broad view of biomarker performance across different scenarios.

Limitations

Nevertheless, this review has limitations. Only nine studies met our inclusion criteria, and their considerable heterogeneity in study design, populations, and test protocols, limits the generalizability of the findings and complicates comparisons across studies. Baseline CA125 concentrations may be influenced by population-specific factors, including age, ethnicity and prior medical history or non-systematic application of a uniform testing methodology. These variables were not incorporated into our analysis, except for pre- and postmenopausal status in women with adnexal masses. While the findings of this study might have considerable applicability, careful consideration is required when generalizing the results to healthcare contexts.

Additionally, data on certain subgroups such as early-stage cancers or pre- and postmenopausal women, were sparse. This prevents firm conclusions for those categories. Finally, the possibility of publication bias cannot be excluded, as studies reporting negative or inconclusive results may be underrepresented in the published literature. Similarly, there is a potential for selection bias arising from the inclusion and exclusion of studies being performed by a single reviewer.

Future perspectives

Future research should prioritize the discovery of biomarkers that can detect ovarian cancer at an early stage, as well as combinations of multiple biomarkers to increase sensitivity for early-stage disease while maintaining specificity. Numerous novel candidates (from protein panels to genetic and cell-free DNA markers) are being investigated, but so far none has surpassed the performance of CA125 and HE4(9).

5. References

1. Torre LA, Trabert B, DeSantis CE, Miller KD, Samimi G, Runowicz CD, et al. Ovarian cancer statistics, 2018. CA Cancer J Clin. 2018 Jul;68(4):284–96.

2. Hong MK, Ding DC. Early Diagnosis of Ovarian Cancer: A Comprehensive Review of the Advances, Challenges, and Future Directions. Vol. 15, Diagnostics. Multidisciplinary Digital Publishing Institute (MDPI); 2025.

3. Lyratzopoulos G, Neal RD, Barbiere JM, Rubin GP, Abel GA. Variation in number of general practitioner consultations before hospital referral for cancer: fi ndings from the 2010 National Cancer Patient Experience Survey in England. 2012; Available from: www.thelancet.com/oncology

4. Ovarian cancer: recognition and initial management Clinical guideline [Internet]. 2011. Available from: www.nice.org.uk/guidance/cg122

5. Prat J. Ovarian carcinomas: Five distinct diseases with different origins, genetic alterations, and clinicopathological features. Vol. 460, Virchows Archiv. 2012. p. 237–49.

6. 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. Vol. 198, American Journal of Obstetrics and Gynecology. 2008. p. 351–6.

7. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018 Nov;68(6):394–424.

8. Jacobs IJ, Menon U, Ryan A, Gentry-Maharaj A, Burnell M, Kalsi JK, et al. Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): A randomised controlled trial. The Lancet. 2016 Mar 5;387(10022):945–56.

9. Mukama T, Fortner RT, Katzke V, Hynes LC, Petrera A, Hauck SM, et al. Prospective evaluation of 92 serum protein biomarkers for early detection of ovarian cancer. Br J Cancer. 2022 May 18;126(9):1301–9.

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

11. 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. BJOG [Internet]. 1990 Oct 1 [cited 2025 Apr 27];97(10):922–9. Available from: /doi/pdf/10.1111/j.1471-0528.1990.tb02448.x

12. Skates SJ. Ovarian cancer screening: Development of the risk of ovarian cancer algorithm (ROCA) and ROCA screening trials. In: International Journal of Gynecological Cancer. 2012.

13. Fekete JT, Győrffy B. MetaAnalysisOnline.com: Web-Based Tool for the Rapid Meta-Analysis of Clinical and Epidemiological Studies. J Med Internet Res [Internet]. 2025 Mar 6;27:e64016. Available from: http://www.ncbi.nlm.nih.gov/pubmed/39928123

14. Menon U, Gentry-Maharaj A, Burnell M, Singh N, Ryan A, Karpinskyj C, et al. Ovarian cancer population screening and mortality after long-term follow-up in the UK Collaborative Trial of

Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. The Lancet. 2021 Jun 5;397(10290):2182–93.

15. Buys SS, Partridge E, Greene MH, Prorok PC, Reding D, Riley TL, et al. Ovarian cancer screening in the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial: Findings from the initial screen of a randomized trial. Am J Obstet Gynecol. 2005 Nov;193(5):1630–9.

16. 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. 2008 May;18(3):414–20.

17. Gentry-Maharaj A, Burnell M, Dilley J, Ryan A, Karpinskyj C, Gunu R, et al. Serum HE4 and diagnosis of ovarian cancer in postmenopausal women with adnexal masses. In: American Journal of Obstetrics and Gynecology. Mosby Inc.; 2020. p. 56.e1-56.e17.

18. Sagi-Dain L, Lavie O, Auslander R, Sagi S. CA 19-9 in evaluation of adnexal mass: Retrospective cohort analysis and review of the literature. International Journal of Biological Markers. 2015 Jul 1;30(3):e333–40.

19. Kim B, Park Y, Kim B, Ahn HJ, Lee KA, Chung JE, et al. Diagnostic performance of CA 125, HE4, and risk of Ovarian Malignancy Algorithm for ovarian cancer. J Clin Lab Anal. 2019 Jan 1;33(1).

20. Funston G, Hamilton W, Abel G, Crosbie EJ, Rous B, Walter FM. The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: A population-based cohort study. PLoS Med. 2020 Oct 28;17(10).

21. Chacón E, Dasí J, Caballero C, Alcázar JL. Risk of Ovarian Malignancy Algorithm versus Risk Malignancy Index-I for Preoperative Assessment of Adnexal Masses: A Systematic Review and Meta-Analysis. Gynecol Obstet Invest. 2019 Nov 1;84(6):591–8.

22. Suri A, Perumal V, Ammalli P, Suryan V, Bansal SK. Diagnostic measures comparison for ovarian malignancy risk in Epithelial ovarian cancer patients: a meta-analysis. Sci Rep. 2021 Dec 1;11(1).

Figures

PRISMA Flow Diagram

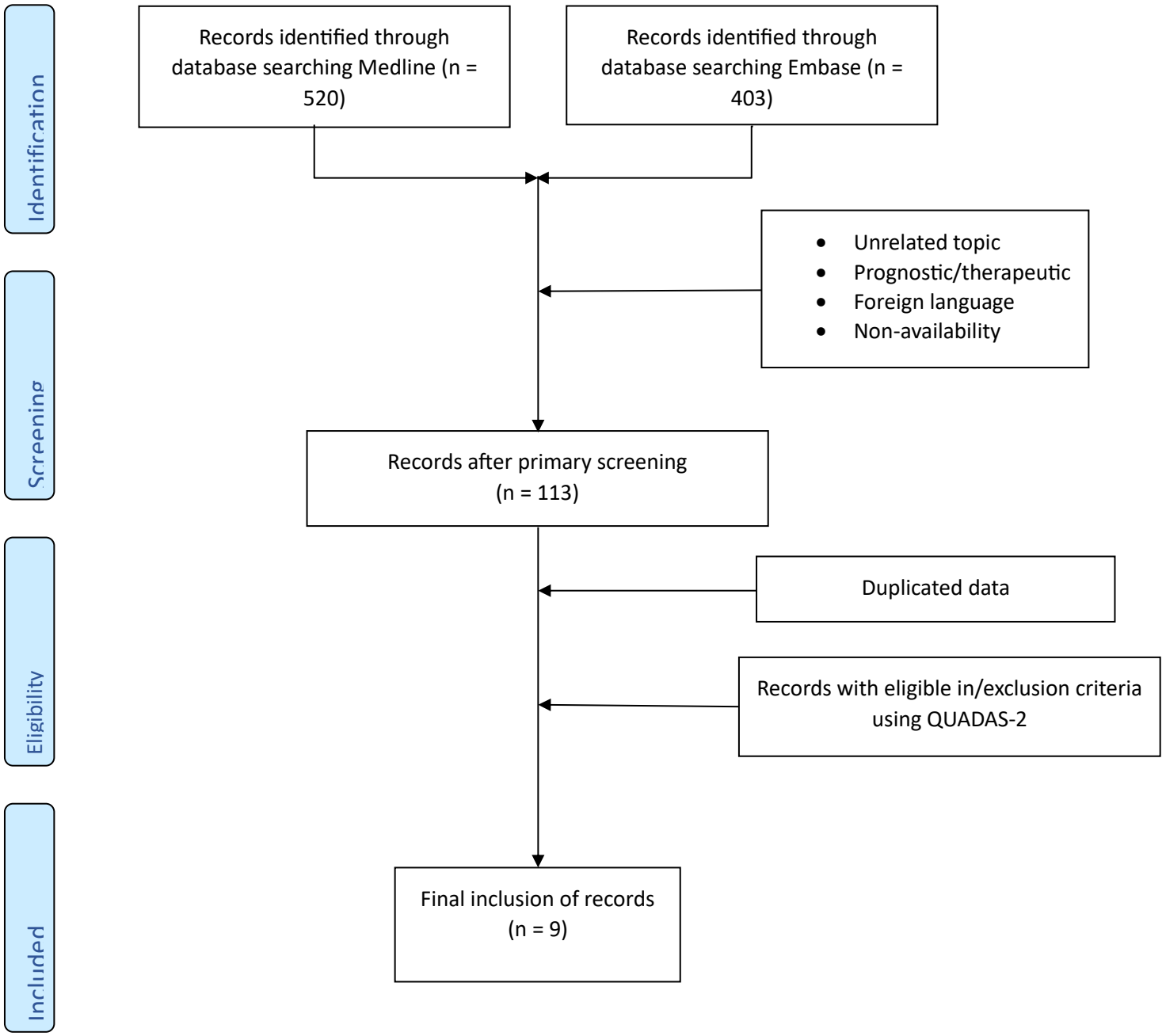


Figure 1. Prisma Flow Diagram

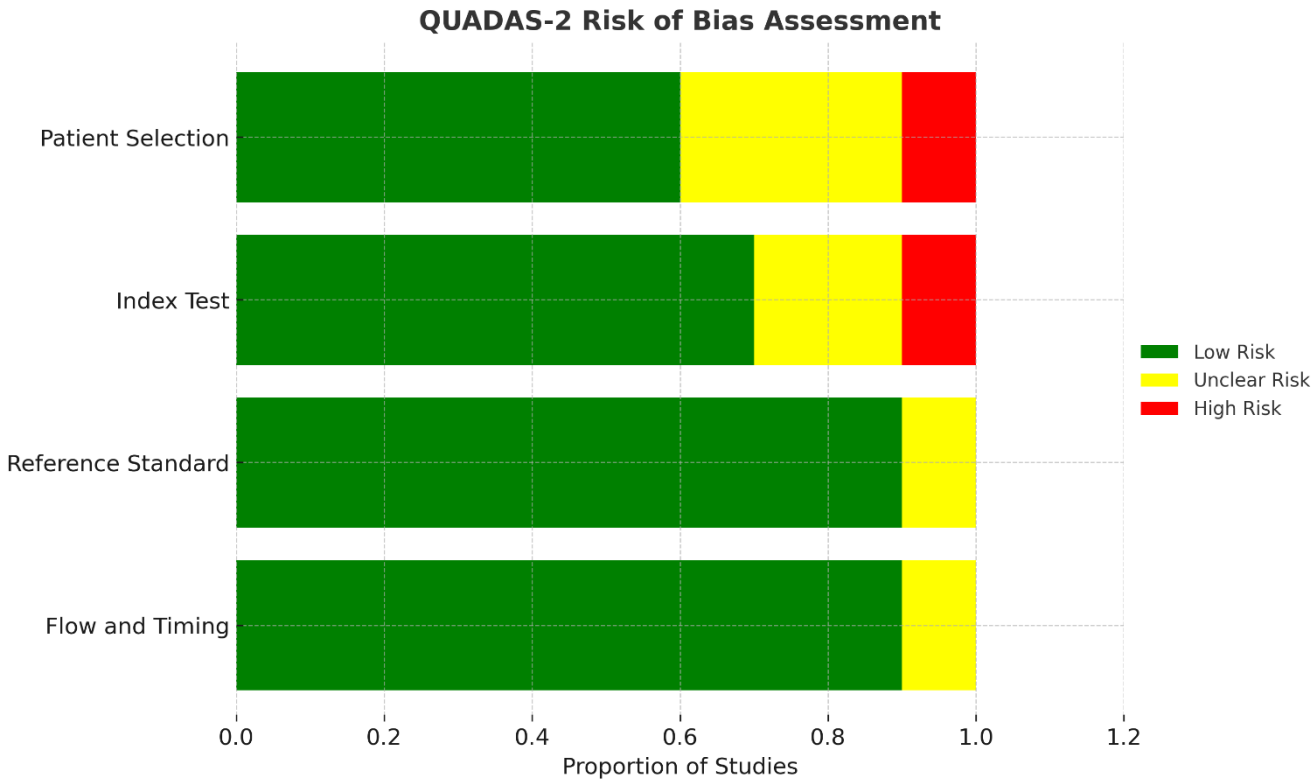


Figure 2. QUADAS-2 analysis of the included studies

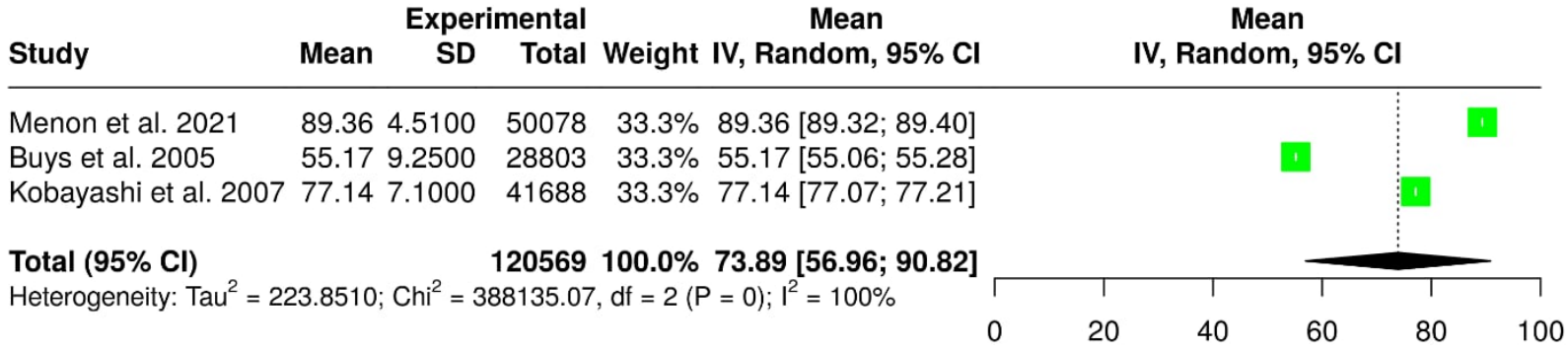


Figure 3. Separate and pooled results for CA125 sensitivity

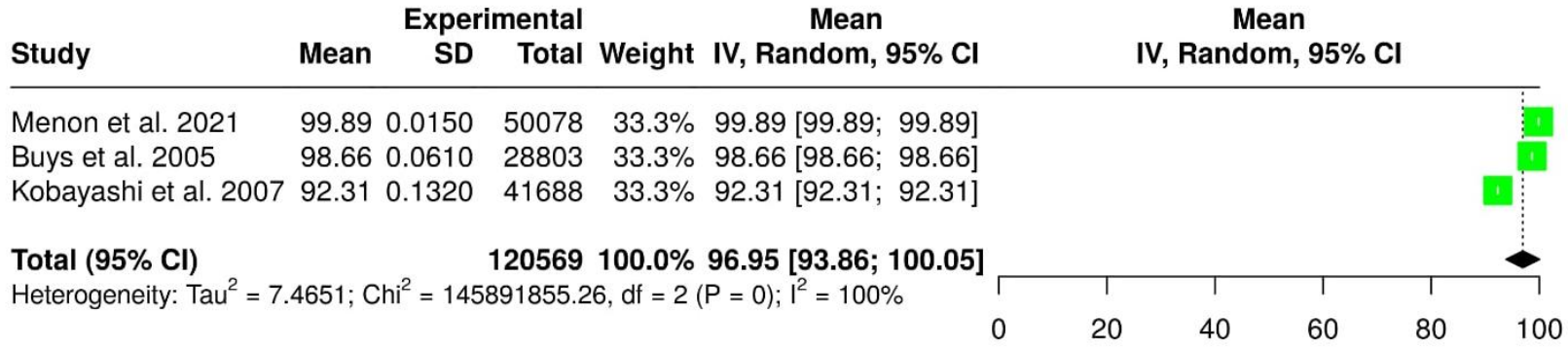


Figure 4. Separate and pooled results for CA125 specificity

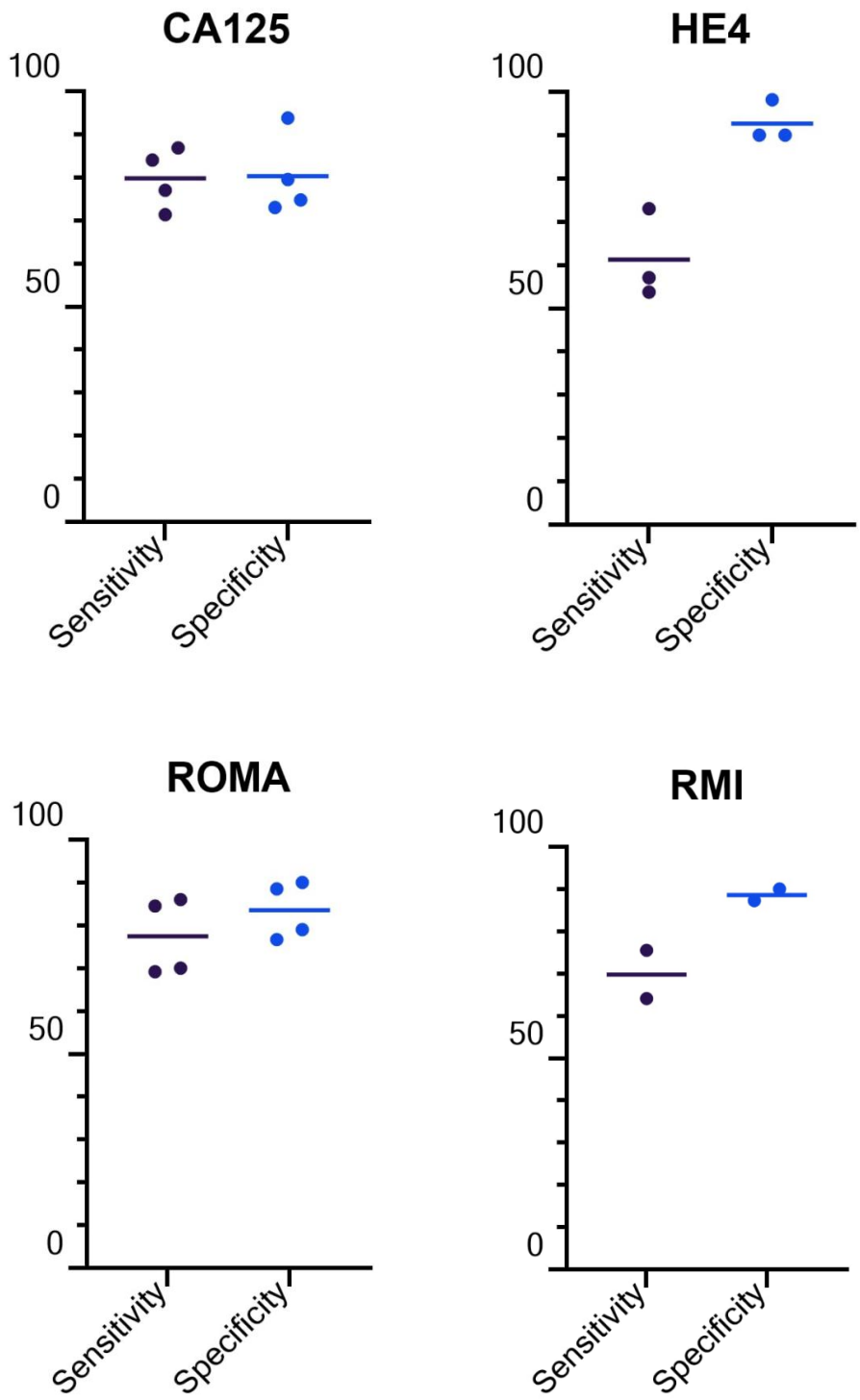


Figure 5. Sensitivity and specificity of CA125, HE4, ROMA, and RMI for pre-operative differentiation of ovarian masses across included studies

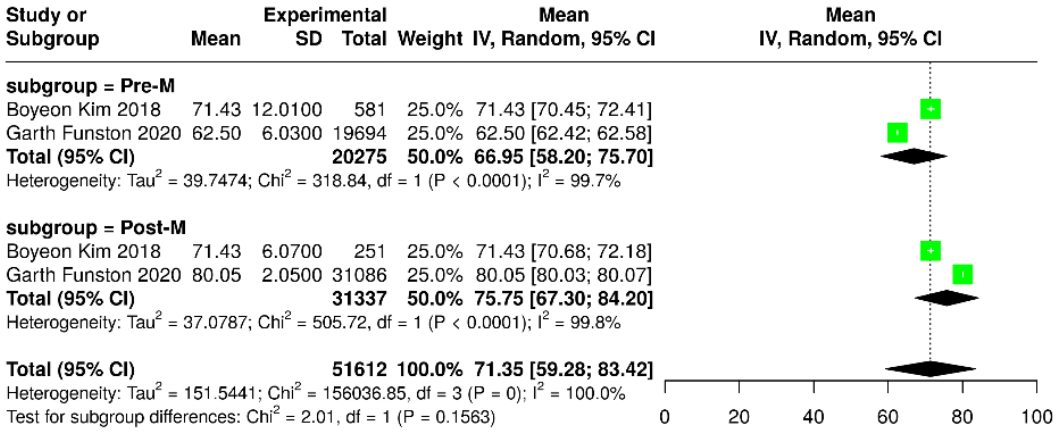


Figure 6A: forest plot with subgroup analysis for sensitivity of CA125 as a screening method for malignancy of adnexal masses in premenopausal (pre-M) and postmenopausal (post-M) women

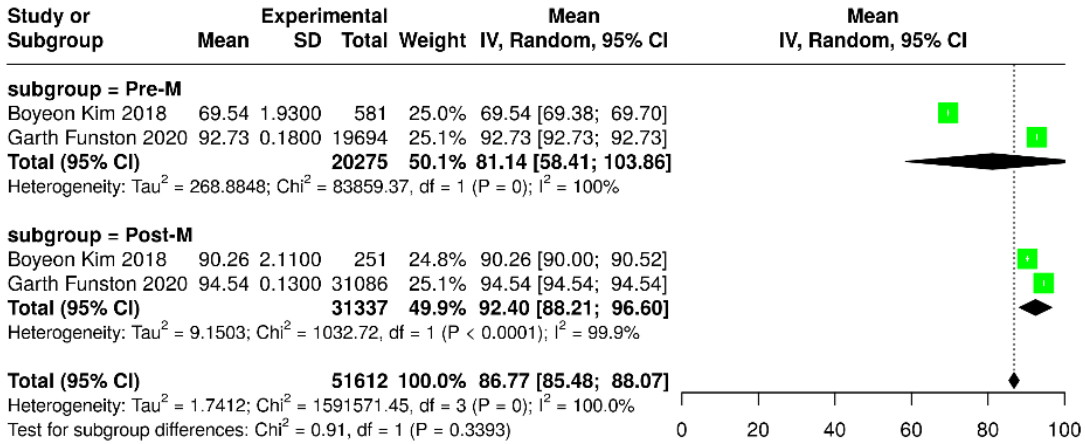


Figure 6B: forest plot with subgroup analysis for specificity of CA125 as a screening method for malignancy of adnexal masses in premenopausal (pre-M) and postmenopausal (post-M) women

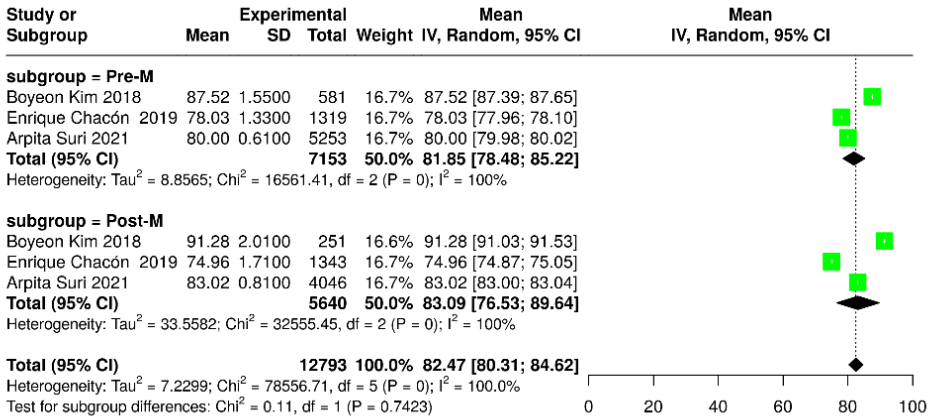


Figure 6C: forest plot with subgroup analysis for sensitivity of ROMA as a screening method for malignancy of adnexal masses in premenopausal (pre-M) and postmenopausal (post-M) women

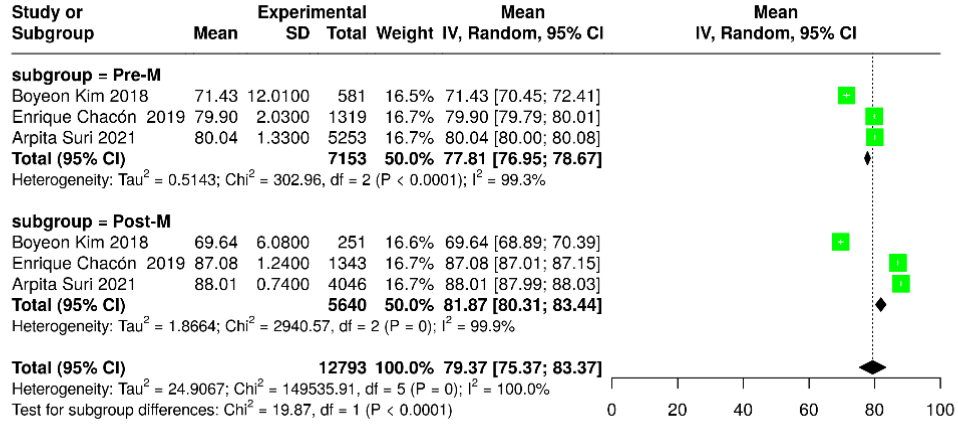


Figure 6D: forest plot with subgroup analysis for specificity of ROMA as a screening method for malignancy of adnexal masses in premenopausal (pre-M) and postmenopausal (post-M) women

Author	Year	N of patients	Study design	Multicentre study	Biomarker	Control group	Follow-up	Menstrual status	Care setting
U. Menon et al.(14)	2021	202.562	RCT	Yes	CA125	Healthy	7 years	Postmenopausal	Tertiary
S. S. Buys et al.(15)	2005	28.803	RCT	Yes	CA125	Healthy	13 years	Postmenopausal	Tertiary
H. Kobayashi et al.(16)	2007	82.827	RCT	Yes	CA125	Healthy	9.2 years	Postmenopausal	Tertiary
A. Gentry-Maharaj et al.(17)	2020	202.562	Retrospective Cohort within RCT	Yes	CA125	Healthy	10.9 years (median)	Postmenopausal	Tertiary
L. Sagi-Dain et al.(18)	2015	503	Retrospective cohort	No	CA125	Benign tumour	/	Both	Secondary
B. Kim et al.(19)	2018	832	Retrospective cohort	No	CA125, HE4, ROMA	Benign tumour	/	Both	Tertiary
G. Funston et al.(20)	2020	50.780	Retrospective cohort	Yes	CA125	Benign tumour	1 year	Both	Primary
E. Chacón et al.(21)	2019	8 studies	Systematic review	No	ROMA, RMI	Benign tumour	/	both	Mostly tertiary
A. Suri et al.(22)	2021	32 studies	Meta-analysis	No	CA125, HE4, ROMA	Benign tumour	/	Both	Mostly tertiary

Table 1. Characteristics of included studies

Authors
Jannick Achtergaele and Jan Verbakel

Conflict of Interest
The authors declare no conflict of interest.

Financial disclosure
Financial benefits for this study were not received.

Funding
Funding for this study was not received.