Medical Policy Manual Laboratory, Policy No. 60

Multimarker and Proteomics-based Serum Testing Related to Ovarian Cancer

Effective: October 1, 2025

Next Review: December 2025
Last Review: September 2025

IMPORTANT REMINDER

Medical Policies are developed to provide guidance for members and providers regarding coverage in accordance with contract terms. Benefit determinations are based in all cases on the applicable contract language. To the extent there may be any conflict between the Medical Policy and contract language, the contract language takes precedence.

PLEASE NOTE: Contracts exclude from coverage, among other things, services or procedures that are considered investigational or cosmetic. Providers may bill members for services or procedures that are considered investigational or cosmetic. Providers are encouraged to inform members before rendering such services that the members are likely to be financially responsible for the cost of these services.

DESCRIPTION

Multimarker serum tests have been proposed as a method for assessing risk of ovarian malignancy in adnexal masses prior to surgery.

MEDICAL POLICY CRITERIA

The use of multimarker serum tests for assessing risk of ovarian malignancy is considered **investigational** for all indications, including but not limited to:

A. Preoperative evaluation of adnexal masses to triage for malignancy
B. Screening for ovarian cancer
C. Selecting patients for surgery for an adnexal mass
D. Evaluation of patients with clinical or radiologic evidence of malignancy
E. Evaluation of patients with nonspecific signs or symptoms suggesting possible malignancy
F. Postoperative testing and monitoring to assess surgical outcome and/or to detect recurrent malignant disease following treatment.

NOTE: A summary of the supporting rationale for the policy criteria is at the end of the policy.

CROSS REFERENCES

1. Analysis of Proteomic and Metabolomic Patterns for Early Detection or Assessing Risk of Cancer, Laboratory, Policy No. 41
2. Investigational Gene Expression, Biomarker, and Multianalyte Testing, Laboratory, Policy No. 77

BACKGROUND

Adults presenting with an adnexal mass have an estimated 68% likelihood of having a benign lesion.[1] About 6% have borderline tumors, 22% have invasive lesions, and 3% have metastatic disease. The mortality rate, for patients with malignant disease depends on three variables: 1) characteristics of the patient; 2) the biology of the tumor (grade, stage, and type); and 3) the quality of treatment (nature of staging, surgery and chemotherapy used).[2] In particular, comprehensive staging and completeness of tumor resection appear to have a positive impact on patient outcome. Racial, ethnic, and socioeconomic disparities in management and outcomes are prominent in patients with ovarian cancer. Compared to non-Hispanic White and Asian patients, Hispanic and non-Hispanic Black patients are more likely to be diagnosed with advanced disease, and are less likely to undergo optimal primary surgery and adjuvant chemotherapy.[3-5] Patients with ovarian cancer from racial and ethnic minorities are also less likely to be enrolled in clinical trials.[6] These are among the contributing factors to worsened overall survival among these racial and ethnic groups.[4, 7, 8] Patients with impediments to access healthcare (eg, those living in underserved areas, with low household income, and/or who are underinsured or uninsured), which frequently intersect with racial and ethnic determinants, also experience longer time to diagnosis, suboptimal treatment, and worse outcomes.[5, 9-11]

A number of studies have evaluated the role of a variety of practice-related factors that may improve health outcomes in patients with ovarian cancer, including specialty treatment by gynecological oncologists.[12-15] These studies have suggested that this specialty treatment may result in improved outcomes, particularly in patients with advanced stage disease.

Surgery is the only way to diagnose ovarian cancer; this is because biopsy of an ovary with suspected ovarian cancer is usually not performed due to the risk of spreading cancer cells. Serum-based tests have been proposed to triage patients with malignant versus benign adnexal masses. A suggested use of the tests is to identify patients who have a higher likelihood of malignant disease and may benefit from referral to a gynecologic-oncology specialist. These tests are combinations of several separate lab tests known as multi-analyte assays with algorithmic analyses (MAAA) and are performed on a blood sample by a reference laboratory using a proprietary algorithm.

The OvaWatch℠ test uses seven serum biomarkers; CA 125, prealbumin, apolipoprotein A-1, beta 2 microglobulin, transferrin, human epididymis protein 4 (HE4) and follicle stimulating hormone (FSH), along with patient age and menopausal status to predict the risk of malignancy in adnexal masses that are deemed benign or indeterminate on clinical assessment.

The OVA1® test algorithm uses five serum biomarkers, CA 125, prealbumin, apolipoprotein A-1, beta 2 microglobulin, and transferrin. A second-generation test called Overa™ replaces prealbumin and beta 2 microglobulin with HE4 and FSH.

The Ova1Plus® test is a reflex test in which the Overa™ test is automatically performed after Ova1® when the Ova1® result indicates intermediate risk.

The Risk of Ovarian Malignancy Algorithm (ROMA™) test combines two biomarkers, HE4 and CA 125, along with menopausal status.

REGULATORY STATUS

NOTE: On December 10, 2011, the U.S. Food and Drug Administration (FDA) published an amendment to the regulation for classifying ovarian adnexal mass assessment score test systems to restrict these devices so that a prescribed warning statement that addresses off-label risks be highlighted by a black box warning.[16] The warning is intended to mitigate the risk to health associated with off-label use as a screening test, stand-alone diagnostic test, or as a test to determine whether or not to proceed with surgery.

The OvaWatch℠ test

In December 2022 Aspira Women's Health (Austin, TX; formerly Vermilion, Inc.) introduced the OvaWatch℠ test. According to the company website, the test has not been cleared for marketing by the FDA.[17]

The Overa™ test

In March 2016, the Overa™ test (Vermillion, Inc. Fremont, CA) was cleared for marketing by the FDA through the 510(k) process, where it was submitted as the OVA1 Next Generation test. This test was predicated on the OVA1® test, to which the Overa™ test was considered substantially equivalent.

The intended use carried a boxed warning: "PRECAUTION: The OVA1 Next Generation test should not be used without an independent clinical and imaging evaluation and is not intended to be a screening test or to determine whether a patient should proceed to surgery. Incorrect use of the OVA1 Next Generation test carries the risk of unnecessary testing, surgery, and/or delayed diagnosis."

The OVA1® test

On July 16, 2009, the OVA1® test (Vermillion, Inc. Fremont, CA) was cleared for market by the FDA as a 510(k) submission. No predicate was identified, and the review decision was based on the de novo 510(k) review process which allows novel products with moderate or low-risk profiles and without predicates which would ordinarily require premarket approval as a class III device to be down-classified in an expedited manner and brought to market with a special control as a class II device.

The intended use carried a boxed warning: "PRECAUTION: The OVA1® test should not be used without an independent clinical/radiological evaluation and is not intended to be a screening test or to determine whether a patient should proceed to surgery. Incorrect use of the OVA1® test carries the risk of unnecessary testing, surgery, and/or delayed diagnosis."

The Ova1Plus® test

The Ova1Plus® test combines two existing FDA-approved tests, Ova1® and Overa™, into a reflex test.

The ROMA™ test

On September 1, 2011, the Risk of Ovarian Malignancy Algorithm (ROMA™ test, Fujirebio Diagnostics, Inc.) was cleared by the FDA as a 510(k) submission. Because the OVA1 test had been found to be a class II medical device by virtue of the July 2009 clearance, ROMA was found to be substantially equivalent to that predicate device.

The intended use carried a boxed warning: “PRECAUTION: The ROMA (HE4 EIA+ARCHITECT CA 125 II) should not be used without an independent clinical/radiological evaluation and is not intended to be a screening test or to determine whether a patient should proceed to surgery. Incorrect use of the ROMA (HE4 EIA+ARCHITECT CA 125 II) carries the risk of unnecessary testing, surgery, and/or delayed diagnosis.”

EVIDENCE SUMMARY

Assessment of a diagnostic technology typically focuses on three parameters: 1) technical performance; 2) diagnostic performance (sensitivity, specificity, and positive [PPV] and negative predictive value [NPV]) in appropriate populations of patients; and 3) demonstration that the diagnostic information can be used to improve patient outcomes (clinical utility).

Technical performance of a device is typically assessed with two types of studies, those that compare test measurements with a gold standard and those that compare results taken with the same device on different occasions (test-retest). Demonstration of technical performance should include an assessment of its reproducibility and precision.

Diagnostic performance is evaluated by the ability of a test to accurately diagnose a clinical condition in comparison with the gold standard. The sensitivity of a test is the ability to detect a disease when the condition is present (true-positive), while specificity indicates the ability to detect patients who are suspected of disease but who do not have the condition (true-negative). Evaluation of diagnostic performance, therefore, requires independent assessment by the two methods in a population of patients who are suspected of disease but who do not all have the disease.

Evidence related to improvement of clinical outcomes with use of this testing assesses the data linking use of a test to changes in health outcomes (clinical utility). While in some cases, tests can be evaluated adequately using technical and diagnostic performance, when a test identifies a new or different group of patients with a disease; randomized trials are needed to demonstrate impact of the test on the net health outcome.

TECHNICAL PERFORMANCE

Evidence on the technical performance of these tests was evaluated by the U.S. Food and Drug Administration (FDA) and is available through the FDA website. The final analysis indicated acceptable technical performance of Overa™, OVA1®, and ROMA™ for use in clinical care.[18-20]

Analytical performance for the OVA1® test demonstrated good test precision (coefficient of variation (CV) ranging from 1% to 7.4%, depending on the sample levels studied) and good reproducibility (CV from 2.8% to 8.9%). The test appears linear, reagent and samples stable, and there was no observed interference evaluating common endogenous substances (hemoglobin, bilirubin, etc.)

The second-generation Overa™ test, had an improved analytical performance, with an overall precision CV of 1.54% across all days and sample pools, in contrast to the OVA1®'s overall precision CV of 4.09%. The reproducibility of Overa™ was also slightly improved over the OVA1®, and it demonstrated a similar lack of interference from common endogenous substances. Analytical performance for the ROMA also exhibited good precision with a total CV ranging from 0.49% to 7.72%, depending on both sample values and menopausal status. The reproducibility of the test was acceptable, with a CV that ranged from 0.98 to 25.9%, with highest values observed in patients with low scores, as expected. The reagents are variably stable, and users are instructed to follow package inserts for stability on each analyte used. The test was unaffected by interference with hemoglobin, bilirubin, lipids, or human anti-mouse antibodies (HAMA). However, high levels of rheumatoid factor (more than 500 IU/mL) did appear to cause elevations in test values and testing in patients with elevated rheumatoid factor is not recommended.

According to the company website, the OvaWatch^{SM} test was validated using a study of 2000 women with adnexal masses, of whom 4.9% had ovarian cancer. The PPV of OvaWatch^{SM} is estimated to be 22.5% and NPV is 99.4%. It is important to note that the study cited by the website involved patients with a surgical treatment plan for the adnexal mass. The intended use of the OvaWatch^{SM} test is to assess the risk of malignancy in adnexal masses that are to be monitored clinically, and not referred for surgery.[17, 21]

DIAGNOSTIC PERFORMANCE

Systematic Reviews

A BlueCross BlueShield Association Technology Evaluation Center (TEC) Assessment was completed in 2012 on "Multi-analyte testing for the evaluation of adnexal masses."[22] The Assessment included evaluation of both the OVA1® and ROMA tests and their impact on health outcomes. The single existing study assessing OVA1® was selected. Studies were selected showing the diagnostic characteristics of ROMA using prespecified cutoff values that assessed diagnostic performance for all types of malignancy, and that did not include healthy subjects as non-malignant control subjects. The TEC Assessment concluded that OVA1® appears to improve sensitivity for detection of malignancy, however specificity declines so much that most patients test positive, and that ROMA does not appear to improve the sensitivity of testing to a great extent.

Wang (2014) published a meta-analysis of studies evaluating the diagnostic accuracy of the ROMA algorithm and comparing it to the performance of single markers HE4 and CA 125.[23] To be included in the meta-analysis, studies had to investigate both HE4 and CA 125 or calculate ROMA, include women with ovarian cancer and benign gynecologic disease, use pathology diagnosis as the reference standard, and collect blood samples before treatment was initiated. A total of 32 studies met these inclusion criteria; 6 of these were conducted in the United States. Findings of the overall pooled analysis of diagnostic accuracy are presented in Table 1. Findings were similar when diagnostic performance in premenopausal women and postmenopausal women were evaluated separately. ROMA had similar or higher sensitivity than HE4 and CA 125, and HE4 had the highest specificity.

Table 1. Diagnostic Performance of ROMA compared with HE4 and CA 125 from Wang (2014)

	No. Studies	Sensitivity % (95% CI)	Specificity %, 95% CI
ROMA	14	85.3 (81.2-88.6)	82.4 (77.4-86.5)
HE4	28	76.3 (72.0-80.1)	93.6 (90.0-95.9)
CA 125	28	79.2 (74.0-83.6)	82.1 (76.6- 86.5)

Dayyani (2016) conducted a meta-analysis comparing ROMA with HE4 and CA 125 in patients with suspected ovarian cancer.[24] Six studies met the inclusion criteria, four of which were included in the Wang (2014) meta-analysis. Two studies were published in 2014 or later. ROMA had higher area under the curve (AUC) values than either HE4 or CA 125 alone (0.921, 95% confidence interval [CI] 0.855 to 0.960, vs 0.899, 95% CI 0.835 to 0.943, and 0.883, 95% CI 0.771 to 0.950, for HE4 and CA 125, respectively). Findings of the pooled analysis of diagnostic accuracy are shown in Table 2.

A sensitivity analysis conducted by Suri (2021) found ROMA had better diagnostic accuracy in postmenopausal women (sensitivity 88%, specificity 83%) than premenopausal women (sensitivity 80%, specificity 80%), and better discrimination (AUROC 0.94 [SE 0.01]) and 0.88 [SE 0.01], respectively).[25] The review found no evidence of publication bias, nor did it find differential results when analyses were limited to blinded studies.

Table 2. Diagnostic Performance of ROMA compared with HE4 and CA 125 from Dayyani (2016)

	No. Studies	Sensitivity % (95% CI)	Specificity %, 95% CI
ROMA	6	87.3 (75.2-94.0)	85.5 (71.9-93.2)
HE4	6	68.2 (69.3-90.1)	85.1 (71.6-92.8)
CA 125	6	79.6 (66.3-88.5)	82.5 (66.2- 91.9)

Chacon (2019) conducted a meta-analysis comparing ROMA with RMI for detecting ovarian cancer.[26] Among the 2,662 women included in the meta-analysis, 50 percent were premenopausal and 50 percent were postmenopausal. Mean ovarian cancer prevalence was 29% in premenopausal women and 51% in postmenopausal women. The majority of studies were conducted at a single center. Although pooled sensitivities for ROMA (Table 3) were similar to those reported in previous systematic reviews that compared ROMA to HE4 and CA 125, specificities for ROMA were somewhat lower in this meta-analysis compared with the Wang (2014) and Dayyani (2016) analyses. However, findings from this meta-analysis should be interpreted with caution due to important limitations including a high-risk of selection bias in most studies and significant unexplained statistical heterogeneity in the meta-analyses.

Table 3. Diagnostic Performance of ROMA compared with RMI from Chacon (2019)

	No. Studies	Sensitivity % (95% CI)		Specificity %, 95% CI	
		Premenopausal	Postmenopausal	Premenopausal	Postmenopausal
ROMA	8	80 (70-88)	87 (78-93)	78 (69-85)	75 (66- 83)
RMI	8	73 (62-81)	77 (65-86)	89 (83-93)	85 (73-92)

Clinical Studies

OVA1®

Diagnostic performance of the OVA1® test was evaluated in a prospective, double-blind clinical study using 27 demographically mixed subject enrollment sites.[18] Patients underwent a complete clinical evaluation prior to surgical intervention, and only patients with planned surgical intervention were included in the study. The pre-surgical process for identifying patients for surgery and for establishing a preliminary diagnosis as benign or malignant were not specifically described but were noted to be "based on a variety of clinical assessments." The study did require at least one imaging test be performed within 12 weeks of surgery. Presumably, use of this somewhat non-standardized diagnostic methodology provides information on how the test works in conjunction with real-world decision making. The study enrolled a total of 743 patients with 146 subjects used in the training set and 516 in the testing set. Seventy-four patients were excluded because of missing information or samples. All patients had adnexal masses and were scheduled for surgery. The final prevalence of cancer in the population was 27%.

Using pathologic diagnosis as the gold standard, test performance, when combined with presurgical assessment for benign disease, was as follows in the hands of non-gynecological oncologists:

	Clinical assessment alone	Clinical assessment with OVA1
Sensitivity	72%*	92%
Specificity	83%	42%
Positive predictive value	61%	37%
Negative predictive value	89%	93%

* Confidence intervals not provided.

OVA1® appeared to improve sensitivity for detection of malignancy; however, specificity declined so much that most patients tested positive.

Bristow (2013) reported on a prospective non-randomized study of 494 patients evaluated for multivariate index assay (OVA1®), CA 125-II, and clinical impression.[27] Patients were all scheduled to undergo surgery for an adnexal mass and all were recruited from non-gynecological oncology practices. Authors sought to assess the OVA1 test in determining the need for gynecological oncology referral by comparing OVA1® to clinical assessment and CA 125-II in identifying women with ovarian cancer. Sensitivity, specificity, positive predictive value, negative predictive value, and likelihood ratios were calculated to estimate the performance of OVA1®, CA 125, clinical impression, and OVA1 with clinical impression. For ovarian malignancies, authors reported a sensitivity of 95.7% when combined with clinical impression. The negative predictive value was reported at 98.1%. However, both clinical impression and CA 125-II were more accurate in identifying benign disease. As in the previous study, although sensitivity improved with the OVA1® test, specificity declined and was reported as 53.5% with OVA1® alone and 50.7% with OVA1® combined with clinical impression compared to 92.5% with clinical impression and 86.1% and 94.5% for CA 125-II (using two different cut-off values) in predicting disease.

Further analysis of combined data from the Bristow (2013) study[27] and another published OVA1® study, the OVA500 trial[28] were reported in two additional articles; one analyzed the performance of clinical assessment with versus without OVA1® testing[29] and the other "was undertaken to better understand the impact of ovarian imaging on the clinical interpretation of the MIA score."[30] The sensitivity, specificity, PPV, and NPV were reported for a number of comparisons and combinations of OVA1® testing with other clinical/imaging findings. However, the authors noted that both the OVA1® and OVA500 trials were designed to measure

accuracy in prediction of malignancy rather than the test's effect on patient referral for sub-specialty care.

Grenache (2015) evaluated the diagnostic performance of both the OVA1® and ROMA tests in a prospective case series of 146 women with an adnexal mass.[31] Although performance characteristics of both tests were determined and compared, the study did not evaluate diagnostic performance in conjunction with clinical assessment, as the both tests were intended to be used. OVA1® was 97% sensitive and 55% specific, and ROMA was 87% sensitive and 83% specific. This means that with clinical assessment (as intended to be used), the OVA1® test would be no worse than 97% sensitive and no better than 55% specific but cannot be determined from the study. The same conclusions can be drawn regarding the ROMA test.

ROMA

Chacon (2023) published a single-center prospective study of 82 consecutive cases (80 patients) of inconclusive adnexal mass that compared ROMA analysis, ultrasound exam by a gynecologic sonologist, and MRI interpreted by radiologist to determine best next step in evaluation.[32] The ROMA analysis was the least sensitive test (26%), and MRI was least specific (77%). Expert ultrasound was much more sensitive (100%) than ROMA ($p<0.001$) and ultrasound specificity (91%) was similar to ROMA (93%). The authors concluded that the expert ultrasound performed best as a second-step test to evaluate inconclusive adnexal masses.

Davenport (2022) published a Cochrane Review to establish and compare the accuracy of test combinations that use menopausal status, ultrasound, and biomarkers for the diagnosis of ovarian cancer in pre-and post-menopausal women with symptoms suspicious for ovarian cancer.[33] The specific purpose of the tests is to determine whether patients should receive further evaluation and management in a generalist vs. specialized gynecologic oncology setting. The review included 59 studies and compared four tests:

- Risk of Malignancy Index (RMI, ultrasound and CA 125 test)
- ROMA test
- International Ovarian Tumor Analysis (IOTA) Logistic Regression model 2 ultrasound
- Assessment of Different NEoplasias (ADNEX, CA 125 test and ultrasound)

Of the 59 studies, 42 evaluated the ROMA test and included the Wang (2014), Chacon (2019), Grenache (2015), Liest (2019), Nikolova (2017), Al Musalhi (2016), and Terlikowska (2016) studies cited below.[23, 26, 31, 34-37] The review found that ROMA and ADNEX demonstrated higher sensitivity than RMI in both pre- and post-menopausal patients and lower specificity in pre-menopausal patients. ROMA had similar specificity as RMI. Davenport (2022) found a high or unclear risk of bias in most studies and a higher prevalence of ovarian cancer than would be expected in the community-based settings that are the intended clinical environment for the tests. Another potential source of bias is the involvement of borderline ovarian tumors in the included studies. Borderline ovarian tumors, which account for about 15% of ovarian tumors, were either excluded from studies or not clearly categorized, which could have led to overestimations of test sensitivity. The review found that the studies did not provide sufficient information to determine whether the test combinations can be used to inform referral decisions for patients with symptoms suspicious for ovarian cancer.

Multiple studies have described the use of ROMA in different populations of women for whom decisions to pursue surgery had been made,[38-41] including Al Musalhi (2016, n=213 cases),[36] Cho (2015, n=90 cases),[42] Terlikowska (2016, n=224 cases),[37] and Minar (2017, n=267).[43]

The largest is Braicu (2022), a prospective study in which 965 patients had HE4, CA 125, and ROMA to determine if adding the tests to transvaginal ultrasound improved diagnostic accuracy prior to surgery for an adnexal mass.[44] Surgical pathology revealed 161 malignant ovarian tumors, which included 43 borderline tumors, and 804 benign tumors. ROMA and the other biomarkers predicted stage 3-4 ovarian cancer well with all tests achieving an area under the curve (AUC) analyses >0.92. But less accuracy was seen in stage 1-2 ovarian cancer with overall performance of all markers at AUC <0.77. In pre-menopausal patients with early-stage ovarian cancer there was no significant difference between HE4 (AUC 0.74), CA 125 (AUC 0.73) and ROMA (AUC 0.74). In post-menopausal patients, CA 125 (AUC 0.77) and ROMA (AUC 0.74) were more accurate that HE4 (AUC 0.62), but adding the biomarkers to transvaginal ultrasound did not improve detection of early-stage ovarian cancer.

Diagnostic performance of the ROMA test was evaluated for FDA approval in a prospective, blinded clinical trial using 13 demographically mixed subject enrollment sites with company sponsorship.[19] Patients all presented with an adnexal mass and were scheduled to undergo surgery. An Initial Cancer Risk Assessment (ICRA) was performed to determine the detection of benign versus malignant lesions before testing. The prevalence of cancer was 15%.

Using pathologic diagnosis as the gold standard, test performance, when combined with presurgical assessment for benign disease, was as follows in the hands of a mixed population of generalist and specialist physicians:

	ICRA alone	**ICRA with ROMA testing**
Sensitivity (95% CI)	77% (66% to 86%)	91% (81% to 96%)
Specificity (95% CI)	84% (80% to 88%)	67% (61% to 71%)
Positive predictive value (95% CI)	46% (17% to 56%)	33% (26% to 40%)
Negative predictive value (95% CI)	96% (93% to 97%	98% (95% to 99%)

Both tests, when added to pre-testing clinical assessment, produced a fall in the positive predictive value of diagnosis with a small increase in the negative predictive value. The changes observed in the negative predictive value were of uncertain statistical and clinical significance.

It is important to note that all of the above literature assessed ROMA as a stand-alone test and did not evaluate diagnostic performance in conjunction with clinical assessment, as the test was intended to be used. Therefore, the ability to draw conclusions regarding the test's diagnostic performance is limited.

Han (2019) compared ROMA to CA125 and HE4 in 876 women with ovarian cysts, with separate analyses for premenopausal (n=532) and postmenopausal n=344) women.[45] The overall sensitivity and specificity of ROMA in this group was 66.7% and 86.8%, respectively. The diagnostic accuracy of ROMA was lower than that of HE4 in premenopausal patients and lower than CA 125 in postmenopausal patients, and similar results were seen for area under the curve (AUC) analyses.

Leist (2019) compared the performance of ROMA with that of the risk of malignancy index (RMI) in a prospective study of 784 women from nine Swedish hospitals who were scheduled

to have a pelvic mass removed.[34] HE4 and CA125 were also evaluated. There were no significant differences between the RMI and ROMA using a fixed specificity of 75%.

A study by Moore (2014) evaluated ROMA in conjunction with clinical assessment, using either positive clinical assessment or positive ROMA as a positive test (similar to the recommended usage for OVA1®).[46] Using this method of combining tests guarantees a higher sensitivity and lower specificity for the combined test than for either test alone. Used in this way, ROMA would only need to be evaluated in patients with a negative clinical assessment. In this study, 461 women were enrolled, of whom a total of 86 (19%) had a malignancy. Combined assessment improved sensitivity from 77.9% to 89.7%, but worsened specificity from 84.3% to 67.2%.

Nikolova (2017) published a case-control study that compared ROMA, Copenhagen Index (CPH-I), HE4, CA 125, RMI, and Morphology Index (MI) for differentiation between ovarian endometriosis and epithelial ovarian cancer.[35] The study included 164 patients: 37 with ovarian endometriosis, 57 with other benign pelvic masses, 11 with ovarian cancer, and 59 controls. The sensitivity, specificity and accuracy to distinguish endometriosis from ovarian cancer for HE4 were: 81.82%, 100%, and 95.83%; for CA 125: 81.82%, 48.65%, and 56.25%; for ROMA: 90.91%, 83.78%, and 85.42%; for CPH-I: 81.82%, 97.30%, and 93.75%; for RMI: 90.91%, 35.14%, and 47.92%; and for MI: 100%, 75.68%, and 81.25%, respectively. The authors concluded that HE4 and CPH-1 had the best ability to discriminate these two conditions, and that MI had the highest sensitivity for ovarian cancer.

Overa™

In a study submitted to the FDA, the clinical validity of the Overa™ test was evaluated in a nonconcurrent prospective study of 493 preoperatively collected serum specimens from premenopausal and postmenopausal women presenting with an adnexal mass requiring surgical intervention.[20] Overa™ test scores were determined based on the analysis of archived serum specimens from a previous study,[27] and patients were stratified into a low- or high-risk groups for finding malignancy on surgery. The analysis examined whether patient referral to a gynecologic oncologist was supported when dual assessment was determined to be positive (either Overa™ or clinical assessment was positive, or both were positive). A dual assessment was considered negative when both Overa™ and clinical assessment were negative.

Using pathologic diagnosis as the criterion standard, Overa™ test performance, when combined with a clinical assessment by nongynecologic oncologists, was as follows:

	Clinical assessment alone	**Dual assessment with Overa™**
Sensitivity (95% CI)	74% (64% to 82%)	94% (87% to 97%)
Specificity (95% CI)	93% (90% to 95%)	65% (60% to 70%)
Positive predictive value (95% CI)	70% (62% to 77%)	38% (35% to 41%)
Negative predictive value (95% CI)	94% (92% to 96%)	98% (95% to 99%)

The method used for combining clinical assessment and Overa™ test result was to consider the test positive if either clinical assessment or Overa™ test was positive. Thus, in practice, the Overa™ testing would not be necessary if clinical assessment alone indicated cancer. Using Overa™ testing in this manner guarantees that Overa™ testing will be more sensitive and less specific than clinical assessment alone, even if it has no better than chance capability of detecting ovarian cancer. Sensitivity improved from 74% to 94%, and specificity decreased from 93% to 65%.

Ova1Plus®

No clinical studies were identified that specifically evaluated the Ova1Plus® test.

OvaWatch℠

No clinical studies were identified that evaluated the OvaWatch℠ test.

Section Summary

There are a limited number of studies comparing diagnostic accuracy of the ROMA, OVA1®, and Overa™ proteomic tests. Studies have found that ROMA has similar or lower accuracy to other risk prediction measures that use components of the standard workup, such as the RMI and the LR2 measures. Use of these tests in addition to other clinical assessment tools appears to increase sensitivity but decreases specificity. Further prospective studies are needed for both assays to understand their proper role in patient care.

CLINICAL UTILITY

The ideal study design to evaluate clinical utility of proteomics-based testing is a randomized controlled trial comparing patient management decisions (e.g., referral patterns) and/or health outcomes (e.g., mortality) in patients managed with proteomic tests with those managed according to best current clinical practices.

Systematic Review

The 2012 TEC Assessment[22] concluded that studies of OVA1® and ROMA showed improvements in sensitivity and worsening of specificity with the use of the tests in conjunction with clinical assessment., and that there are problems in concluding that this results in improved health outcomes. The clinical assessment performed in the studies was not well characterized. The evidence regarding the effect of OVA1® and ROMA on health outcomes is indirect, and based on studies of diagnostic performance of the tests in patients undergoing surgery for adnexal masses. The authors noted that there were no prospective studies on the use of these tests in patients who presented with an adnexal mass or studies that reported the impact of testing on referral patterns or the impact on health outcomes.

Randomized Controlled Trials

No RCTs on the effects of these tests on patient management decisions and healthcare outcomes compared with current assessment were identified.

Nonrandomized Studies

No prospective outcome studies have been performed using the OVA1®, Overa™, or ROMA test. Kaijser (2014) published a retrospective cohort study included 101 newly diagnosed cases of biopsy-proven invasive ovarian cancer, which provided information relevant to outcomes.[47] Blood samples obtained before treatment were analyzed; HE4 and CA 125 levels were measured and the ROMA algorithm was calculated. Median overall survival in the study cohort was 3.7 years. In a multivariate analysis controlling for confounding variables, neither HE4 levels nor ROMA were independently associated with progression-free survival (PFS) or disease-specific survival (DSS). For example, for ROMA and the outcome of PFS, the adjusted hazard ratio (HR) for each 10% increase in risk was 0.98(95% CI) 0.88 to 1.11). Patients were not prospectively managed according to their HE4 levels or ROMA score and thus the actual

impact of these tests on PFS and DSS cannot be determined from this study. It is not clear what impact either test would have on long-term health outcomes or referral patterns to specialty physicians. The use of proteomic testing to triage patients for malignancy may be only one of many factors in decision making about where treatment should be delivered.

A retrospective study by Dunton (2020) reported on referral patterns for patients who had a low-risk results from the OVA1 test.[48] Of the 146 patients with low-risk OVA1 results, 82 (56%) patients had surgery and 17 of these were referred for specialty care. No cases of invasive malignancy were found.

Section summary

Although current studies show improvements in sensitivity and worsening of specificity with the use of the OVA1®, Overa™, and ROMA tests in conjunction with clinical assessment, there are problems in concluding that this results in improved health outcomes. The clinical assessment performed in the studies is not well characterized. In addition, there is indirect evidence from studies of diagnostic accuracy which suggest that the ROMA test would not improve the accuracy of triage compared to existing measures and is unlikely to improve the accuracy of referral to a specialist and is therefore, unlikely to improve outcomes.

PRACTICE GUIDELINE SUMMARY

THE SOCIETY FOR GYNECOLOGIC ONCOLOGY (SGO)

In May 2013, the Society for Gynecologic Oncology (SGO) issued the following consensus based statement on multiplex serum testing for women with pelvic masses:[49]

> "Blood levels of five proteins in women with a known ovarian mass have been reported to change when ovarian cancer is present. Tests measuring these proteins may be useful in identifying women who should be referred to a gynecologic oncologist. Recent data have suggested that such tests, along with physician clinical assessment, may improve detection rates of malignancies among women with pelvic masses planning surgery. Results from such tests should not be interpreted independently, nor be used in place of a physician's clinical assessment. Physicians are strongly encouraged to reference the American Congress of Obstetricians and Gynecologists' 2011 Committee Opinion "The Role of the Obstetrician-Gynecologist in the Early Detection of Epithelial Ovarian Cancer" to determine an appropriate care plan for their patients."

The SGO Position Statements regarding the use of the OVA1 or ROMA test is not based upon scientific evidence and does not recommend the use of the OVA1 test as part of their referral guidelines for women who present with an adnexal mass.

THE AMERICAN CONGRESS OF OBSTETRICIANS AND GYNECOLOGISTS (ACOG)

In 2017, with reaffirmation in 2019, the American College of Obstetricians and Gynecologists (ACOG) opinion on the role of the obstetrician-gynecologist in the early detection of epithelial ovarian cancer addressed using multimarker serum testing.[50] The opinion states that multimarker panels lack strong evidence for use in asymptomatic women without adnexal masses and do not improve early detection and survival rates in average-risk women. The Society for Gynecologic Oncology endorsed this ACOG opinion.

ACOG additionally published a 2016 practice bulletin on the Evaluation and Management of

Adnexal Masses, which states:[51]

> Serum biomarker panels may be used as an alternative to CA 125 level alone in determining the need for referral to or consultation with a gynecologic oncologist when an adnexal mass requires surgery. These biomarker panels are not recommended for use in the initial evaluation of an adnexal mass, but may be helpful in assessing which women would benefit from referral to a gynecologic oncologist. Trials that have evaluated the predictive value of these panels show potential for improved specificity, especially for evaluation of premenopausal women. However, comparative research has not yet defined the best testing approach.

NATIONAL COMPREHENSIVE CANCER NETWORK (NCCN)

The NCCN ovarian cancer guidelines (v.3.2025)[52] include the following statement:

> "There are a number of prediction algorithms that combine multiple factors, such as symptoms, imaging results, biomarkers, and patient characteristics, to predict the likelihood of malignancy among patients who have an undiagnosed adnexal mass (i.e., a mass detected by clinical exam or imaging that has not yet been resected and definitively diagnosed by pathology). These algorithms were developed with the goal of reducing the number and/or extent of unnecessary surgeries…the NCCN Guidelines do not endorse any of these methods."

SUMMARY

There is not enough research to show that multimarker serum testing to determine the risk of ovarian cancer can improve health outcomes for patients with pelvic masses. Clinical guidelines based on research do not currently recommend the use of these tests for patients with pelvic masses. Therefore, all uses of these tests, including use as a screening tool for ovarian cancer, are considered investigational.

REFERENCES

1. Van Holsbeke C, Van Belle V, Leone FP, 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 : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology.* 2010;36(1):81-7. PMID: 20217895
2. du Bois A, Rochon J, Pfisterer J, et al. Variations in institutional infrastructure, physician specialization and experience, and outcome in ovarian cancer: a systematic review. *Gynecologic oncology.* 2009;112(2):422-36. PMID: 18990435
3. Matthews BJ, Qureshi MM, Fiascone SJ, et al. Racial disparities in non-recommendation of adjuvant chemotherapy in stage II-III ovarian cancer. *Gynecologic oncology.* 2022;164(1):27-33. PMID: 34785030
4. Zhang C, Zhang C, Wang Q, et al. Differences in Stage of Cancer at Diagnosis, Treatment, and Survival by Race and Ethnicity Among Leading Cancer Types. *JAMA Netw Open.* 2020;3(4):e202950. PMID: 32267515

5. Joslin CE, Brewer KC, Davis FG, et al. The effect of neighborhood-level socioeconomic status on racial differences in ovarian cancer treatment in a population-based analysis in Chicago. *Gynecologic oncology.* 2014;135(2):285-91. PMID: 25173584
6. Mattei LH, Robb L, Banning K, et al. Enrollment of Individuals From Racial and Ethnic Minority Groups in Gynecologic Cancer Precision Oncology Trials. *Obstetrics and gynecology.* 2022;140(4):654-61. PMID: 36075065
7. Zhong P, Yang B, Pan F, et al. Temporal trends in Black-White disparities in cancer surgery and cancer-specific survival in the United States between 2007 and 2015. *Cancer Med.* 2023;12(3):3509-19. PMID: 35968573
8. Ellis L, Canchola AJ, Spiegel D, et al. Racial and Ethnic Disparities in Cancer Survival: The Contribution of Tumor, Sociodemographic, Institutional, and Neighborhood Characteristics. *J Clin Oncol.* 2018;36(1):25-33. PMID: 29035642
9. Albright BB, Nasioudis D, Craig S, et al. Impact of Medicaid expansion on women with gynecologic cancer: a difference-in-difference analysis. *American journal of obstetrics and gynecology.* 2021;224(2):195.e1-95.e17. PMID: 32777264
10. Bodurtha Smith AJ, Applebaum J, Fader AN. Association of the Affordable Care Act's Medicaid Expansion With 1-Year Survival Among Patients With Ovarian Cancer. *Obstetrics and gynecology.* 2022;139(6):1123-29. PMID: 35675609
11. Smick AH, Holbert M, Neff R. Association of Physician Densities and Gynecologic Cancer Outcomes in the United States. *Obstetrics and gynecology.* 2022;140(5):751-57. PMID: 36201771
12. Vernooij F, Heintz P, Witteveen E, et al. The outcomes of ovarian cancer treatment are better when provided by gynecologic oncologists and in specialized hospitals: a systematic review. *Gynecologic oncology.* 2007;105(3):801-12. PMID: 17433422
13. Giede KC, Kieser K, Dodge J, et al. Who should operate on patients with ovarian cancer? An evidence-based review. *Gynecologic oncology.* 2005;99(2):447-61. PMID: 16126262
14. Tingulstad S, Skjeldestad FE, Hagen B. The effect of centralization of primary surgery on survival in ovarian cancer patients. *Obstetrics and gynecology.* 2003;102(3):499-505. PMID: 12962932
15. Grabowski JP, Harter P, Buhrmann C, et al. Re-operation outcome in patients referred to a gynecologic oncology center with presumed ovarian cancer FIGO I-IIIA after sub-standard initial surgery. *Surgical oncology.* 2012;21(1):31-5. PMID: 20875732
16. Medical Devices: Ovarian adnexal mass assessment score test system; Labeling; Black box restrictions. 21 CFR Part 866, Federal Register 2011;76(251):82128-82123.
17. OvaWatch. [cited 8/20/2025]. 'Available from:' https://aspirawh.com/ovawatch/.
18. U.S. Food and Drug Administration. 510(k) Substantial Equivalence Determination Decision Summary: OVA1™ Test (K081754). [cited 8/20/2025]. 'Available from:' http://www.accessdata.fda.gov/cdrh_docs/reviews/K081754.pdf.
19. U.S. Food and Drug Administration. 510(k) Substantial Equivalence Determination Decision Summary: ROMA™ Test (K103358). [cited 8/20/2025]. 'Available from:' http://www.accessdata.fda.gov/cdrh_docs/reviews/K103358.pdf.
20. U.S. Food and Drug Administration. 510(k) Substantial Equivalence Determination Decision Summary: Overa™ Test (K150588). [cited 8/20/2025]. 'Available from:' http://www.accessdata.fda.gov/cdrh_docs/pdf15/k150588.pdf.
21. Reilly G, Bullock RG, Greenwood J, et al. Analytical Validation of a Deep Neural Network Algorithm for the Detection of Ovarian Cancer. *JCO Clin Cancer Inform.* 2022;6:e2100192. PMID: 35671415

22. TEC Assessment 2012. "Multi-analyte Testing for the Evaluation of Adnexal Masses." BlueCross BlueShield Association Technology Evaluation Center, Vol. 27, Tab TBA.
23. 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 biology : the journal of the International Society for Oncodevelopmental Biology and Medicine.* 2014;35(6):6127-38. PMID: 24627132
24. Dayyani F, Uhlig S, Colson B, et al. Diagnostic Performance of Risk of Ovarian Malignancy Algorithm Against CA125 and HE4 in Connection With Ovarian Cancer: A Meta-analysis. *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society.* 2016;26(9):1586-93. PMID: 27540691
25. Suri A, Perumal V, Ammalli P, et al. Diagnostic measures comparison for ovarian malignancy risk in Epithelial ovarian cancer patients: a meta-analysis. *Sci Rep.* 2021;11(1):17308. PMID: 34453074
26. Chacon E, Dasi J, Caballero C, et al. Risk of Ovarian Malignancy Algorithm versus Risk Malignancy Index-I for Preoperative Assessment of Adnexal Masses: A Systematic Review and Meta-Analysis. *Gynecologic and obstetric investigation.* 2019;84(6):591-98. PMID: 31311023
27. Bristow RE, Smith A, Zhang Z, et al. Ovarian malignancy risk stratification of the adnexal mass using a multivariate index assay. *Gynecologic oncology.* 2013;128(2):252-9. PMID: 23178277
28. Ueland FR, Desimone CP, Seamon LG, et al. Effectiveness of a multivariate index assay in the preoperative assessment of ovarian tumors. *Obstetrics and gynecology.* 2011;117:1289-97. PMID: 21606739
29. Longoria TC, Ueland FR, Zhang Z, et al. Clinical performance of a multivariate index assay for detecting early-stage ovarian cancer. *American journal of obstetrics and gynecology.* 2014;210(1):78 e1-9. PMID: 24055582
30. Goodrich ST, Bristow RE, Santoso JT, et al. The effect of ovarian imaging on the clinical interpretation of a multivariate index assay. *American journal of obstetrics and gynecology.* 2014;211(1):65 e1-65 e11. PMID: 24530816
31. 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. *Clinica chimica acta; international journal of clinical chemistry.* 2015;438:358-63. PMID: 25283731
32. Chacon E, Arraiza M, Manzour N, et al. Ultrasound examination, MRI, or ROMA for discriminating between inconclusive adnexal masses as determined by IOTA Simple Rules: a prospective study. *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society.* 2023;33(6):951-56. PMID: 37055169
33. Davenport C, Rai N, Sharma P, et al. Menopausal status, ultrasound and biomarker tests in combination for the diagnosis of ovarian cancer in symptomatic women. *Cochrane Database Syst Rev.* 2022;7(7):Cd011964. PMID: 35879201
34. Liest AL, Omran AS, Mikiver R, et al. RMI and ROMA are equally effective in discriminating between benign and malignant gynecological tumors: A prospective population-based study. *Acta obstetricia et gynecologica Scandinavica.* 2019;98(1):24-33. PMID: 30216407
35. Nikolova T, Zivadinovic R, Evtimovska N, et al. Diagnostic performance of human epididymis protein 4 compared to a combination of biophysical and biochemical markers to differentiate ovarian endometriosis from epithelial ovarian cancer in premenopausal women. *The journal of obstetrics and gynaecology research.* 2017;43(12):1870-79. PMID: 29027715

36. Al Musalhi K, Al Kindi M, Al Aisary F, et al. Evaluation of HE4, CA-125, Risk of Ovarian Malignancy Algorithm (ROMA) and Risk of Malignancy Index (RMI) in the Preoperative Assessment of Patients with Adnexal Mass. *Oman medical journal.* 2016;31(5):336-44. PMID: 27602187
37. Terlikowska KM, Dobrzycka B, Witkowska AM, et al. Preoperative HE4, CA125 and ROMA in the differential diagnosis of benign and malignant adnexal masses. *Journal of ovarian research.* 2016;9(1):43. PMID: 27436085
38. Choi HJ, Lee YY, Sohn I, et al. Comparison of CA 125 alone and risk of ovarian malignancy algorithm (ROMA) in patients with adnexal mass: A multicenter study. *Current problems in cancer.* 2019:100508. PMID: 31708114
39. Shen Y, Zhao L, Lu S. Diagnostic performance of HE4 and ROMA among Chinese women. *Clinica chimica acta; international journal of clinical chemistry.* 2020;500:42-46. PMID: 31626761
40. Dochez V, Randet M, Renaudeau C, et al. Efficacy of HE4, CA125, Risk of Malignancy Index and Risk of Ovarian Malignancy Index to Detect Ovarian Cancer in Women with Presumed Benign Ovarian Tumours: A Prospective, Multicentre Trial. *Journal of clinical medicine.* 2019;8(11). PMID: 31699959
41. Carreras-Dieguez N, Glickman A, Munmany M, et al. Comparison of HE4, CA125, ROMA and CPH-I for Preoperative Assessment of Adnexal Tumors. *Diagnostics (Basel).* 2022;12(1). PMID: 35054393
42. Cho HY, Park SH, Park YH, et al. Comparison of HE4, CA125, and Risk of Ovarian Malignancy Algorithm in the Prediction of Ovarian Cancer in Korean Women. *Journal of Korean medical science.* 2015;30(12):1777-83. PMID: 26713052
43. Minar L, Felsinger M, Cermakova Z, et al. Comparison of the Copenhagen Index versus ROMA for the preoperative assessment of women with ovarian tumors. *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics.* 2017. PMID: 29086914
44. Braicu EI, Krause CL, Torsten U, et al. HE4 as a serum biomarker for the diagnosis of pelvic masses: a prospective, multicenter study in 965 patients. *BMC Cancer.* 2022;22(1):831. PMID: 35907794
45. Han KH, Park NH, Kim JJ, et al. The power of the Risk of Ovarian Malignancy Algorithm considering menopausal status: a comparison with CA 125 and HE4. *Journal of gynecologic oncology.* 2019;30(6):e83. PMID: 31576682
46. Moore RG, Hawkins DM, Miller MC, 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. PMID: 25449569
47. Kaijser J, Van Belle V, Van Gorp T, et al. Prognostic value of serum HE4 levels and risk of ovarian malignancy algorithm scores at the time of ovarian cancer diagnosis. *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society.* 2014;24(7):1173-80. PMID: 24987915
48. Dunton CJ, Eskander RN, Bullock RG, et al. Low-risk multivariate index assay scores, physician referral and surgical choices in women with adnexal masses. *Current medical research and opinion.* 2020;36(12):2079-83. PMID: 33107342
49. Society of Gynecologic Oncologists. Multiplex Serum Testing for Women with Pelvic Mass (May 2013). [cited 8/20/2025]. 'Available from:' https://www.sgo.org/newsroom/position-statements-2/multiplex-serum-testing-for-women-with-pelvic-mass/.

50. Committee Opinion No. 716: The Role of the Obstetrician-Gynecologist in the Early Detection of Epithelial Ovarian Cancer in Women at Average Risk. *Obstetrics and gynecology.* 2017;130(3):e146-e49. PMID: 28832487
51. Practice Bulletin No. 174: Evaluation and Management of Adnexal Masses. *Obstetrics and gynecology.* 2016;128(5):e210-e26. PMID: 27776072
52. National Comprehensive Cancer Network (NCCN). Clinical Practice Guidelines in Oncology™. Ovarian Cancer Including Fallopian Tube Cancer and Primary Peritoneal Cancer v.3.2025. [cited 8/20/2025]. 'Available from:' https://www.nccn.org/professionals/physician_gls/pdf/ovarian.pdf.

CODES

Codes	Number	Description
CPT	0003U	Oncology (ovarian) biochemical assays of five proteins (apolipoprotein A-1, CA 125 II, follicle stimulating hormone, human epididymis protein 4, transferrin), utilizing serum, algorithm reported as a likelihood score
	0375U	Oncology (ovarian), biochemical assays of 7 proteins (follicle stimulating hormone, human epididymis protein 4, apolipoprotein A-1, transferrin, beta-2 macroglobulin, prealbumin [ie, transthyretin], and cancer antigen 125), algorithm reported as ovarian cancer risk score
	0577U	Oncology (ovarian), serum, analysis of 39 glycoproteins by liquid chromatography with tandem mass spectrometry (LC-MS/MS) in multiple reaction monitoring mode, reported as likelihood of malignancy
	81500	Oncology (ovarian), biochemical assays of two proteins (CA-125 and HE4), utilizing serum, with menopausal status, algorithm reported as a risk score (This code is for reporting the ROMA™ test)
	81503	Oncology (ovarian), biochemical assays of five proteins (CA-125, apoliproprotein A1, beta-2 microglobulin, transferrin and pre-albumin), utilizing serum, algorithm reported as a risk score (This code is for reporting the OVA1™ test)
HCPCS	None	

Date of Origin: *November 2012*