## Policy Statement

All uses of the OVA1, Overa, and ROMA tests are considered **investigational**, including but not limited to:

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

## Policy Guidelines

OVA1, Overa, and ROMA tests are combinations of several separate lab tests and involve proprietary algorithms for determining risk (i.e., what American Medical Association’s CPT calls multianalyte assays with algorithmic analyses [MAAAs]).

### Coding

The following CPT category I MAAA code is specific for ROMA test:

- **81500**: Oncology (ovarian), biochemical assays of two proteins (CA-125 and HE4), utilizing serum, with menopausal status, algorithm reported as a risk score

The following CPT category I MAAA code is specific for OVA1:

- **81503**: Oncology (ovarian), biochemical assays of five proteins (CA-125, apolipoprotein A1, beta-2 microglobulin, transferrin, and pre-albumin), utilizing serum, algorithm reported as a risk score

CPT instructs that these codes cannot be reported with the component tests (i.e., codes 86304 and 86305 cannot be reported with 81500, and codes 82172, 82232, 83695, 83700, 84134, 84466, and 86304 cannot be reported with 81503).

The following code is specific to Overa:

- **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

## Description

A variety of serum biomarkers have been studied for their association with ovarian cancer. Of particular interest have been tests that integrate results from multiple analytes into a risk score to predict the presence of disease. Three tests based on this principle, OVA1, Overa (the second-generation OVA1 test), and ROMA have been cleared by the U.S. Food and Drug Administration. The intended use of OVA1 and Overa is to use them as an aid to further assess whether malignancy is present even when the physician’s independent clinical and radiologic evaluation does not indicate malignancy. The intended use of ROMA is to use it as an aid, in conjunction with clinical assessment, to assess whether a premenopausal or a postmenopausal woman who presents with an ovarian adnexal mass is at high or low likelihood of finding malignancy on surgery.
Related Policies

- Serum Biomarker Human Epididymis Protein 4

Benefit Application

Benefit determinations should be based in all cases on the applicable contract language. To the extent there are any conflicts between these guidelines and the contract language, the contract language will control. Please refer to the member's contract benefits in effect at the time of service to determine coverage or non-coverage of these services as it applies to an individual member.

Some state or federal mandates (e.g., Federal Employee Program [FEP]) prohibits plans from denying Food and Drug Administration (FDA)-approved technologies as investigational. In these instances, plans may have to consider the coverage eligibility of FDA-approved technologies on the basis of medical necessity alone.

Regulatory Status

In July 2009, the OVA1® test (Aspira Labs [Austin, TX]) was cleared for marketing by the FDA through the 510(k) process. OVA1® was designed as a tool to further assess the likelihood that malignancy is present when the physician’s independent clinical and radiologic evaluation does not indicate malignancy.

In September 2011, the Risk of Ovarian Malignancy Algorithm (ROMA™ test; Fujirebio Diagnostics [Sequin, TX]) was cleared for marketing by the FDA through the 510(k) process. The intended use of ROMA™ is as an aid, in conjunction with clinical assessment, in assessing whether a premenopausal or postmenopausal woman who presents with an ovarian adnexal mass is at high or low likelihood of finding malignancy on surgery.

In March 2016, a second-generation test called Overa™ (also referred as next-generation OVA1®), in which 2 of the 5 biomarkers in OVA1® are replaced with human epididymis secretory protein 4 and follicle stimulating hormone, was cleared for marketing by the FDA through the 510(k) process. Similar to OVA1®, Overa™ generates a low or high risk of malignancy on a scale from 0 to 10.

Black Box Warning

In December 2011, the FDA amended its regulation for classifying ovarian adnexal mass assessment score test systems. The change required that off-label risks be highlighted using a black box warning. 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 to proceed with surgery. Considering the history and currently unmet medical needs for ovarian cancer testing, the FDA concluded that there is a risk of off-label use of this device. To address this risk, the FDA requires that manufacturers provide notice concerning the risks of off-label uses in the labeling, advertising, and promotional material of ovarian adnexal mass assessment score test systems. Manufacturers must address the following risks:

- Women without adnexal pelvic masses (i.e., for cancer “screening”) are not part of the intended use population for the ovarian adnexal mass assessment score test systems. Public health risks associated with false-positive results for ovarian cancer screening tests are well described in the medical literature and include morbidity or mortality associated with unneeded testing and surgery. The risk from false-negative screening results also includes morbidity and mortality due to failure to detect and treat ovarian malignancy.
- Analogous risks, adjusted for prevalence and types of disease, arise if test results are used to determine the need for surgery in patients who are known to have ovarian adnexal masses.
If used outside the “OR” rule that is described in this special control guidance, results from ovarian adnexal mass assessment score test systems pose a risk for morbidity and mortality due to nonreferral for oncologic evaluation and treatment.

**Rationale**

**Background**

**Epithelial Ovarian Cancer**

The term epithelial ovarian cancer collectively includes high-grade serous epithelial ovarian, fallopian tubal, and peritoneal carcinomas due to their shared pathogenesis, clinical presentation, and treatment. We use epithelial ovarian cancer to refer to this group of malignancies in the discussion that follows. There is currently no serum biomarker that can distinguish between these types of carcinoma. An estimated 22,440 women in the United States are expected to be diagnosed in 2017 with ovarian cancer, and approximately 14,080 will die of the disease. The mortality rate depends on 3 variables: (1) patient characteristics; (2) tumor biology (grade, stage, type); and (3) treatment quality (nature of staging, surgery, and chemotherapy used). In particular, comprehensive staging and completeness of tumor resection appear to have a positive impact on patient outcome.

In 1997, the Society of Surgical Oncology recommended ovarian cancer surgery and follow-up treatment be performed by physicians with ovarian cancer disease expertise. Numerous articles have been published on the application of this recommendation examining long- and short-term outcomes as well as process measures (e.g., types of treatment such as complete staging or tumor debulking). At least 2 meta-analyses have concluded that outcomes are improved when patients with ovarian cancer are treated by gynecologic oncologists. The available data are most convincing for patients with advanced-stage disease.

Adult women presenting with an adnexal mass have an estimated 68% likelihood of having a benign lesion. About 6% of women with masses have borderline tumors; 22% possess invasive malignant lesions, and 3% have metastatic 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. Most clinicians agree that women with masses that have a high likelihood of malignancy should undergo surgical staging by a gynecologic oncologist. However, women with clearly benign masses do not require a referral to see a specialist. Therefore, criteria and tests that help differentiate benign from malignant pelvic masses are desirable.

In 2005, the American College of Obstetricians and Gynecologists and the Society of Gynecologic Oncologists jointly released referral guidelines that addressed criteria for referring women with pelvic masses suspicious for ovarian cancer to gynecologic oncologists. Separate criteria were developed for premenopausal and postmenopausal women. In premenopausal women, referral criteria included at least one of the following: elevated cancer antigen 125 (CA 125; >200 U/mL), ascites, evidence of abdominal or distant metastasis, or a positive family history. The referral criteria for postmenopausal women were similar, except that a lower threshold for an elevated CA 125 test was used (35 U/mL); moreover, a nodular or fixed pelvic mass was an added criterion.

Three multimarker serum-based tests specific to ovarian cancer have been cleared by the Food and Drug Administration (FDA) with the intended use of triaging patients with adnexal masses (see Regulatory Status section). They are summarized in Table 1. The proposed use of the tests is to identify women with a substantial likelihood of malignant disease who may benefit from referral to a gynecologic oncology specialist. Patients with positive results may be considered candidates for referral to a gynecologic oncologist for treatment. The tests have been developed and evaluated only in patients with adnexal masses and planned surgeries. Other potential uses, such as selecting patients to have surgery, screening asymptomatic patients, and...
monitoring treatment, have not been investigated. Furthermore, the tests are not intended to be used as stand-alone tests, but in conjunction with clinical assessment.

Other multimarker panels and longitudinal screening algorithms are under development; however, these are not yet commercially available.8,9.

Table 1. Summary of FDA-Approved Multimarker Serum-Based Tests Specific to Ovarian Cancer

<table>
<thead>
<tr>
<th>Variables</th>
<th>OVA1</th>
<th>Overa</th>
<th>ROMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approved</td>
<td>2009</td>
<td>2016</td>
<td>2011</td>
</tr>
<tr>
<td>Manufacturer</td>
<td>Quest Diagnostics</td>
<td>Vermillion</td>
<td>Roche Diagnostics</td>
</tr>
<tr>
<td>Biomarkers used</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA 125 II</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>b2-microglobulin</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transferrin</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Transthryretin</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apolipoprotein Al</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>HE4</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>FSH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score range</td>
<td>0-10</td>
<td>0-10</td>
<td>0-10</td>
</tr>
<tr>
<td>Risk categorization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Premenopausal</td>
<td>&lt;5.0: low</td>
<td>≥5.0: high</td>
<td>≥1.3: high</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postmenopausal</td>
<td>&lt;4.4: low</td>
<td>≥4.4: high</td>
<td>≥2.77: high</td>
</tr>
</tbody>
</table>


**Literature Review**

Evidence reviews assess whether a medical test is clinically useful. A useful test provides information to make a clinical management decision that improves the net health outcome. That is, the balance of benefits and harms is better when the test is used to manage the condition than when another test or no test is used to manage the condition.

The first step in assessing a medical test is to formulate the clinical context and purpose of the test. The test must be technically reliable, clinically valid, and clinically useful for that purpose. The following is a summary of the key findings to date.

**Multimarker Serum Testing Related to Ovarian Cancer**

**Clinical Context and Test Purpose**

The purpose of multimarker serum testing of individuals over age 18 with an ovarian adnexal mass for which surgery is planned and not yet referred to an oncologist is to use the test as an aid to further assess the probability that malignancy is present, even when the physician’s independent clinical and radiologic evaluation does not indicate malignancy.

The questions addressed in this evidence review are: (1) Is there evidence that multimarker serum testing of individuals described above has clinical validity? and (2) Does multimarker serum testing of such individuals change patient management in a way that improves outcomes as a result of testing?

The following PICOTS were used to select literature to inform this review.

**Patients**

The relevant population of interest is individuals who:
Multimarker Serum Testing Related to Ovarian Cancer

- Are over age 18
- Have ovarian adnexal mass for which surgery is planned
- Have not yet been referred to an oncologist
- A physician’s independent clinical and radiologic evaluation does not indicate malignancy.

**Interventions**

The relevant interventions are 3 commercially multimarker serum genetic tests (e.g., OVA1, Overa, ROMA).

**Comparators**

The comparator of interest is standard clinical assessment.

**Outcomes**

The potential beneficial outcomes of primary interest in the case of a true negative would be the avoidance of unnecessary surgery and its associated consequences (e.g., morbidity, mortality, resource utilization, patient anxiety). The potential harms from a false-positive could be inappropriate assessment and improper management of patients with ovarian malignancies, which could result in the following: inappropriate surgical decisions, high frequency of unnecessary further testing, and unnecessary patient anxiety. The potential harms from a false-negative could be a determination that the patient does not have ovarian malignancy, which would lead to a delay in surgery and tumor diagnosis.

Off-label use of the test (e.g., in patients who have not already been identified as needing surgery for pelvic mass, or patients without reference to an independent clinical and radiologic evaluation), might lead to a high frequency of unnecessary testing and surgery due to false-positive results, or to a delay in tumor diagnosis due to false-negative results.

**Timing**

Multimarker serum testing for related to ovarian cancer may be performed at any point when an individual presents with an ovarian adnexal mass for which surgery is planned, but physician’s independent clinical and radiologic evaluation does not indicate malignancy and referral to an oncologist is being considered.

**Setting**

Most patients are likely to be tested in an outpatient setting.

**Study Selection Criteria**

Below are selection criteria for studies to assess whether a test is clinically valid:

1. The study population represents the population of interest. Eligibility and selection are described.
2. The test is compared with a credible reference standard.
3. If the test is intended to replace or be an adjunct to an existing test; it should also be compared with that test.
4. Studies should report sensitivity, specificity, and predictive values. Studies that completely report true- and false-positive results are ideal. Studies reporting other measures (e.g., ROC, AUROC, c-statistic, likelihood ratios) may be included but are less informative.
5. Studies should also report reclassification of diagnostic or risk category.

Technically Reliable Assessment of technical reliability focuses on specific tests and operators and requires review of unpublished and often proprietary information. Review of specific tests, operators, and unpublished data are outside the scope of this evidence review and alternative sources exist. This evidence review focuses on the clinical validity and clinical utility.
Clinically Valid
A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

OVA1 Test
Descriptions of the developmental process for the OVA1 test have been published in the FDA documents and in a perspective by Fung (2010). Candidate biomarkers were selected based on initial studies using mass spectroscopy but were converted to standard immunoassays to improve analytic performance. Seven final markers were evaluated, none of which individually appeared to be highly specific for malignant ovarian disease. However, the choice of five of these (CA 125, prealbumin, apo AI, β2-microglobulin, transferrin) produced a composite profile that did appear to have discriminatory ability. The test, as cleared by the FDA, is performed on a blood sample, which is sent to a reference laboratory for testing using the 5 immunoassays previously described. Results of the 5 determinations are entered manually into an Excel spreadsheet used by the OvaCalc software. This software contains an algorithm that combines the 5 discrete values into a single unitless numeric score from 0.0 to 10.0.

Details of the algorithm appear proprietary, but the development is described as an empirical process; it is a process based on several different factors: the use of banked samples from academic partners; a small prospective study of samples from Europe; and a designated subset of samples from the clinical study used to support the submission to the FDA. It appears that, at an undisclosed point in the developmental process as a result of interaction with the FDA, separate cutpoints were developed for premenopausal and postmenopausal women.

The clinical validity was evaluated in a prospective, double-blind, clinical study using 27 enrollment sites. The study was supported by the commercial sponsor of the test. Patients underwent a complete clinical evaluation before surgical intervention, and only patients with adnexal masses who had a planned surgical intervention were included. The study enrolled 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. The final prevalence of cancer in the population was 27%.

Using pathologic diagnosis as the criterion standard, OVA1 test performance, when combined with a clinical assessment by non-gynecologic oncologists, was as follows (see Table 2). The method used for combining clinical assessment and OVA1 result was to consider the test positive if either clinical assessment or OVA1 test was positive. Thus, in practice, OVA1 testing would not be necessary if clinical assessment alone indicated cancer. Using OVA1 testing in this manner guarantees that OVA1 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 72% to 92%, and specificity decreased from 83% to 42%.

Table 2. Clinical Validity of the OVA1 Test Among 269 Patients Evaluated by Nongynecologic Oncologists

<table>
<thead>
<tr>
<th>Diagnostic Characteristics</th>
<th>Clinical Assessment Alone, %</th>
<th>Clinical Assessment with OVA1 Test, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>72</td>
<td>92</td>
</tr>
<tr>
<td>Specificity</td>
<td>83</td>
<td>42</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>61</td>
<td>37</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>89</td>
<td>93</td>
</tr>
</tbody>
</table>

One additional 2015 study (by Grenache et al) was identified; it evaluated the diagnostic performance of the OVA1 test. However, it did not evaluate diagnostic performance in conjunction with clinical assessment, as the test was intended to be used. By itself, OVA1 was 97% sensitive and 55% specific. This means that with clinical assessment (as intended to be used), the test would be no worse than 97% sensitive and no better than 55% specific, but these characteristics cannot be determined from the study.

### Table 3. Summary of Key Study Characteristics

<table>
<thead>
<tr>
<th>Study; Trial</th>
<th>Countries</th>
<th>Sites</th>
<th>Dates</th>
<th>Participants</th>
<th>Interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grenache (2015)</td>
<td>US</td>
<td>2009-2011</td>
<td>Women with an adnexal mass (n=146)</td>
<td>OVA1</td>
<td>ROMA</td>
</tr>
</tbody>
</table>

### Table 4. Summary of Key Study Results

<table>
<thead>
<tr>
<th>Study</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>PPV (95% CI)</th>
<th>NPV (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grenache (2015)</td>
<td>96.8% (83.3-99.9)</td>
<td>54.8% (45.2-64.1)</td>
<td>36.6% (26.2-48.0)</td>
<td>98.4% (91.6-99.9)</td>
</tr>
<tr>
<td>OVA1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROMA</td>
<td>83.9% (66.3-94.6)</td>
<td>83.5% (75.4-89.8)</td>
<td>57.8% (42.2-72.3)</td>
<td>95.1% (88.8-98.4)</td>
</tr>
</tbody>
</table>

CI: confidence interval; PPV: positive predictive value; NPV: negative predictive value.

The purpose of the gaps tables (see Tables X and X) is to display notable gaps identified in each study. This information is synthesized as a summary of the body of evidence following each table and provides the conclusions on the sufficiency of evidence supporting the position statement.

### Table 5. Relevance Gaps

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Intervention</th>
<th>Comparator</th>
<th>Outcomes</th>
<th>Follow-Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDA(k) OVA1 Test K081754</td>
<td>1. Some patients were not evaluated by a gynecologic oncologist; 2. Unclear how patients were recruited; 3. Enrollment was limited to patients with planned surgical intervention</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grenache et al (2015)</td>
<td>1. Patients were not evaluated by a gynecologic oncologist; 4. Enrollment included only patients with planned surgical intervention, due to the small number of women with malignant adnexal masses, the strength of</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The evidence gaps stated in this table are those notable in the current review; this is not a comprehensive gaps assessment.

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Intervention</th>
<th>Comparator</th>
<th>Outcomes</th>
<th>Follow-Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDA(k) OVA1 Test K081754</td>
<td>1. Subjects were not allocated randomly</td>
<td>1.10% of subjects were eliminated due to missing information or lack of sample</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grenache et al (2015)</td>
<td>1. Subjects were not allocated randomly</td>
<td>1.2. Treatment assignment and outcome assessment were not blinded</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The evidence gaps stated in this table are those notable in the current review; this is not a comprehensive gaps assessment.

d Follow-Up key: 1. High loss to follow-up or missing data; 2. Inadequate handling of missing data; 3. High number of crossovers; 4. Inadequate handling of crossovers; 5. Inappropriate exclusions; 6. Not intent to treat analysis (per protocol for noninferiority trials).
e Power key: 1. Power calculations not reported; 2. Power not calculated for primary outcome; 3. Power not based on clinically important difference.
f Statistical key: 1. Intervention is not appropriate for outcome type: (a) continuous; (b) binary; (c) time to event; 2. Intervention is not appropriate for multiple observations per patient; 3. Confidence intervals and/or p values not reported; 4. Comparative treatment effects not calculated.

Overa Test
Descriptions of the developmental process for the Overa test have been published in the FDA documents. The FDA documents do not provide details on how biomarkers were selected. The test, as cleared by the FDA, is performed on a blood sample, which is to be sent to a reference laboratory for testing using the 5 immunoassays previously described. Results of the 5 determinations are entered into a proprietary algorithm, called OvaCalc software (v4.0.0), which combines the 5 discrete values into a single unitless numeric score from 0.0 to 10.0.

Clinical validity 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. Overa test scores were determined based on the analysis of archived serum specimens from a previous study, and the patient was stratified...
into a low- or high-risk group 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 clinical assessment by nongynecologic oncologists, was as follows (see Table 3). 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, 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%.

**Table 7. Clinical Validity of the Overa Test Among 493 Patients Evaluated by Nongynecologic Oncologists**

<table>
<thead>
<tr>
<th>Diagnostic Characteristics</th>
<th>Clinical Assessment Alone, %</th>
<th>Dual Assessment with Overa Test, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity (95% CI)</td>
<td>74 (64 to 82)</td>
<td>94 (87 to 97)</td>
</tr>
<tr>
<td>Specificity (95% CI)</td>
<td>93 (90 to 95)</td>
<td>65 (60 to 70)</td>
</tr>
<tr>
<td>Positive predictive value (95% CI)</td>
<td>70 (62 to 77)</td>
<td>38 (35 to 41)</td>
</tr>
<tr>
<td>Negative predictive value (95% CI)</td>
<td>94 (92 to 96)</td>
<td>98 (95 to 99)</td>
</tr>
<tr>
<td>Prevalence</td>
<td></td>
<td>19 (92/493)</td>
</tr>
</tbody>
</table>

Adapted from U.S. Food and Drug Administration (FDA). 510(k) Substantial Equivalence Determination Decision Summary: OVA1™ Next Generation Test (K150588). n.d.; CI: confidence interval.

The purpose of the gap tables (see Tables X and X) is to display notable gaps identified in each study. This information is synthesized as a summary of the body of evidence following each table and provides the conclusions on the sufficiency of evidence supporting the position statement.

**Table 8. Relevance Gaps**

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Intervention</th>
<th>Comparator</th>
<th>Outcomes</th>
<th>Follow-Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDA 510(k) OVA1 K1505881</td>
<td>4. 70.3% of subjects were white</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The evidence gaps stated in this table are those notable in the current review; this is not a comprehensive gaps assessment.

- Population key: 1. Intended use population unclear; 2. Clinical context is unclear; 3. Study population is unclear; 4. Study population not representative of intended use.
- Intervention key: 1. Not clearly defined; 2. Version used unclear; 3. Delivery not similar intensity as comparator; 4. Not the intervention of interest.
- Comparator key: 1. Not clearly defined; 2. Not standard or optimal; 3. Delivery not similar intensity as intervention; 4. Not delivered effectively.

**Table 9. Study Design and Conduct Gaps**

<table>
<thead>
<tr>
<th>Study</th>
<th>Allocation</th>
<th>Blinding</th>
<th>Selective Reporting</th>
<th>Follow-Up</th>
<th>Power</th>
<th>Statistical</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDA 510(k) OVA1 K1505881</td>
<td>1. Subjects were not</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The evidence gaps stated in this table are those notable in the current review; this is not a comprehensive gaps assessment.


\( b \) Blinding key: 1. Not blinded to treatment assignment; 2. Not blinded outcome assessment; 3. Outcome assessed by treating physician.

\( c \) Selective Reporting key: 1. Not registered; 2. Evidence of selective reporting; 3. Evidence of selective publication.

\( d \) Follow-Up key: 1. High loss to follow-up or missing data; 2. Inadequate handling of missing data; 3. High number of crossovers; 4. Inadequate handling of crossovers; 5. Inappropriate exclusions; 6. Not intent to treat analysis (per protocol for noninferiority trials).

\( e \) Power key: 1. Power calculations not reported; 2. Power not calculated for primary outcome; 3. Power not based on clinically important difference.

\( f \) Statistical key: 1. Intervention is not appropriate for outcome type: (a) continuous; (b) binary; (c) time to event; 2. Intervention is not appropriate for multiple observations per patient; 3. Confidence intervals and/or p values not reported; 4. Comparative treatment effects not calculated.

**ROMA Test**

Moore et al (2008) described the development of the ROMA test.17 The authors studied 9 biomarkers and chose HE4 and CA 125 because these markers in tandem produced the best performance. The algorithm developed was subsequently modified to include menopausal status and was independently validated.18 Again, separate cutoffs were used for premenopausal and postmenopausal women.

In 2014, Wang et al published a meta-analysis of studies evaluating the clinical validity of the ROMA test algorithm and comparing it with the performance of single biomarkers HE4 and CA 125.19 To be included in the meta-analysis, studies had to investigate both HE4 and CA 125 or calculate ROMA, enroll women with ovarian cancer and benign gynecologic disease, use pathology diagnosis as the reference standard, and collect blood samples before treatment was initiated. Thirty-two studies met these inclusion criteria; six were conducted in the United States. Findings of the overall pooled analysis of diagnostic accuracy are presented in Table 4.

**Table X. Characteristics of Systematic Reviews**

<table>
<thead>
<tr>
<th>Study</th>
<th>Studies Included</th>
<th>Study Populations Included</th>
<th>Study Designs Included</th>
<th>Study Reference Standards Included</th>
<th>No Domains</th>
<th>1-2 Domains</th>
<th>&gt;2 Domains</th>
<th>Domains With &gt;33% Studies at High Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang (2014)</td>
<td>32</td>
<td>Women with ovarian cancer and benign gynecologic disease</td>
<td>Blinded and unblinded</td>
<td>CA125, HE4, and ROMA</td>
<td></td>
<td></td>
<td></td>
<td>•</td>
</tr>
<tr>
<td>Dayyani (2016)</td>
<td>6</td>
<td>Women with ovarian cancer</td>
<td>All</td>
<td>CA125, HE4, and ROMA</td>
<td></td>
<td></td>
<td></td>
<td>•</td>
</tr>
</tbody>
</table>
Table 10. Meta-Analytic Findings for Diagnostic Performance of the ROMA Test vs HE4 and CA 125

<table>
<thead>
<tr>
<th>Test</th>
<th>No. Studies</th>
<th>Sensitivity (95% CI), %</th>
<th>Specificity (95% CI), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROMA test</td>
<td>14</td>
<td>85.3 (81.2 to 88.6)</td>
<td>82.4 (77.4 to 86.5)</td>
</tr>
<tr>
<td>Human epididymis secretory protein 4</td>
<td>28</td>
<td>76.3 (72.0 to 80.1)</td>
<td>93.6 (90.0 to 95.9)</td>
</tr>
<tr>
<td>Cancer antigen 125</td>
<td>28</td>
<td>79.2 (74.0 to 83.6)</td>
<td>82.1 (76.6 to 86.5)</td>
</tr>
</tbody>
</table>

Adapted from Wang et al (2014).19
CI: confidence interval.

Findings were similar when diagnostic performance in premenopausal women and postmenopausal women were evaluated separately. ROMA had similar or higher sensitivity than HE4 and CA125, and HE4 had the highest specificity.

In 2016, Dayyani et al conducted a meta-analysis comparing ROMA with HE4 and CA 125 in patients with suspected ovarian cancer.20 Six studies met the inclusion criteria, four of which were included in the 2014 Wang meta-analysis. Two studies were published in 2014 or later. Based on area under the curve analysis, ROMA had higher values than either HE4 (0.921; 95% CI, 0.855 to 0.960) or CA 125 alone (0.899; 95% CI, 0.835 to 0.943) and HE4 plus CA 125 (0.883; 95% CI, 0.771 to 0.950). Findings of the pooled analysis of diagnostic accuracy are shown in Table 5.

Table 11. Meta-Analytic Findings for Diagnostic Performance of the ROMA Test vs HE4 and CA 125

<table>
<thead>
<tr>
<th>Test</th>
<th>No. Studies</th>
<th>Sensitivity (95% CI), %</th>
<th>Specificity (95% CI), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROMA test</td>
<td>6</td>
<td>87.3 (75.2 to 94.0)</td>
<td>85.5 (71.9 to 93.2)</td>
</tr>
<tr>
<td>Human epididymis secretory protein 4</td>
<td>6</td>
<td>68.2 (69.3 to 90.1)</td>
<td>85.1 (71.6 to 92.8)</td>
</tr>
<tr>
<td>Cancer antigen 125</td>
<td>6</td>
<td>79.6 (66.3 to 88.5)</td>
<td>82.5 (82.5 to 91.9)</td>
</tr>
</tbody>
</table>

Adapted from Dayyani et al (2016).20
CI: confidence interval.

The point estimates for sensitivity and specificity were lower in pre- and postmenopausal women, with wider confidence intervals.

Since the Wang and Dayyani meta-analyses, multiple studies have described the use of the ROMA test in populations of women in whom decisions to pursue surgery had been made, including Al Musalhi et al (2016; n=213 cases),21 Cho et al (2015; n=90 cases),22 and Terlikowska et al (2016; n=224 cases).23

The FDA labelling for ROMA, unlike that for OVA1, does not indicate how ROMA is to be used in conjunction with clinical assessment. All previously cited literature assessed ROMA as a stand-alone test for ovarian cancer and did not provide a comparison with clinical assessment alone. The study by Moore et al (2014) evaluated ROMA in conjunction with clinical assessment, using either a positive clinical assessment or a positive ROMA as a positive test (similar to the recommended usage for OVA1).24 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 given to patients with a negative clinical assessment. In this study, 461 women were enrolled, of whom 86 (19%) had a malignancy. Combined assessment improved sensitivity from 77.9% to 89.7%, but specificity worsened from 84.3% to 67.2% (see Table 6).
Table X. Summary of Key Study Characteristics

<table>
<thead>
<tr>
<th>Study; Trial</th>
<th>Countries</th>
<th>Sites</th>
<th>Dates</th>
<th>Participants</th>
<th>Interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moore (201)</td>
<td>US</td>
<td>13</td>
<td>2009-2010</td>
<td>Women with an ovarian cyst or pelvic mass (n=461)</td>
<td>ICRA+ROMA</td>
</tr>
</tbody>
</table>

Adapted from Moore et al (2014).24

The purpose of the gap tables (see Tables X and X) is to display notable gaps identified in each study. This information is synthesized as a summary of the body of evidence following each table and provides the conclusions on the sufficiency of evidence supporting the position statement.

Table 12. Diagnostic Performance of the ROMA Test for All Malignancy

<table>
<thead>
<tr>
<th>Diagnostic Characteristics</th>
<th>Clinical Assessment Alone (95% CI), %</th>
<th>Clinical Assessment with ROMA (95% CI), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>77.9 (66.2 to 87.1)</td>
<td>89.7 (79.9 to 95.8)</td>
</tr>
<tr>
<td>Specificity</td>
<td>84.3 (80.2 to 87.8)</td>
<td>67.2 (62.2 to 71.9)</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>47.3 (37.8 to 57.0)</td>
<td>33.2 (26.4 to 40.5)</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>95.5 (92.6 to 97.4)</td>
<td>97.3 (94.5 to 98.9)</td>
</tr>
</tbody>
</table>

Adapted from Moore et al (2014).24

CI: confidence interval.

Table 13. Relevance Gaps

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Intervention</th>
<th>Comparator</th>
<th>Outcomes</th>
<th>Follow-Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moore et al (2014)</td>
<td>4. 84.8% of subjects were white; 60.4% of subjects were EOC grade 3; 66.7% had stage III epithelial ovarian cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The evidence gaps stated in this table are those notable in the current review; this is not a comprehensive gaps assessment.

a Population key: 1. Intended use population unclear; 2. Clinical context is unclear; 3. Study population is unclear; 4. Study population not representative of intended use.

b Intervention key: 1. Not clearly defined; 2. Version used unclear; 3. Delivery not similar intensity as comparator; 4. Not the intervention of interest.

c Comparator key: 1. Not clearly defined; 2. Not standard or optimal; 3. Delivery not similar intensity as intervention; 4. Not delivered effectively.

d Outcomes key: 1. Key health outcomes not addressed; 2. Physiologic measures, not validated surrogates; 3. No CONSORT reporting of harms; 4. Not establish and validated measurements; 5. Clinical significant difference not prespecified; 6. Clinical significant difference not supported.

e Follow-Up key: 1. Not sufficient duration for benefit; 2. Not sufficient duration for harms.

Table 14. Study Design and Conduct Gaps

<table>
<thead>
<tr>
<th>Study</th>
<th>Allocation</th>
<th>Blinding</th>
<th>Selective Reporting</th>
<th>Follow-Up</th>
<th>Power</th>
<th>Statistical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moore et al (2014)</td>
<td>1. Subjects were not allocated randomly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The evidence gaps stated in this table are those notable in the current review; this is not a comprehensive gaps assessment.


Section Summary: Clinically Valid
Evidence for the clinical utility for the OVA1 and Overa tests include prospective, double-blind studies that have evaluated the clinical validity of these tests in predicting the likelihood of malignancy in women who are planning to have surgery for an adnexal mass. They have not been studied for ovarian cancer screening. The prospective studies showed that, in patients with adnexal mass who had a planned surgical intervention, use of OVA1 and Overa in conjunction with a clinical assessment by nongynecologic oncologists increased the sensitivity but decreased the specificity compared with clinical assessment alone. When used with clinical assessment in this manner, the sensitivity to ovarian malignancy was 92% and the specificity was 42%. ROMA is intended for use in conjunction with clinical assessment, but no specific method has been defined. One study, which used clinical assessment and ROMA results, showed a sensitivity of 90% and a specificity of 67%. Two meta-analysis reported less than 90% sensitivity and specificity with ROMA testing.

Clinically Useful
A test is clinically useful if the use of the results informs management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.

The ideal study design to evaluate the clinical utility of multimarker serum-based test would be a randomized controlled trial comparing health outcomes (e.g., mortality) in patients managed using the tests with those managed according to best current clinical practices. According to the chain of logic, greater numbers of persons referred for initial surgical treatment with ovarian cancer should result in improved overall health outcomes. No randomized or nonrandomized studies with these comparisons were identified.

Although OVA1, Overa, and ROMA, when used in conjunction with clinical assessment, improve the sensitivity for detection of malignancy, the specificity declines. In studies using either positive ROMA or clinical assessment as a positive test, sensitivity improved—but it was still less than 90%. It is uncertain whether there is meaningful clinical benefit from using a test that avoids a high number of referrals and does not contain sensitive data (even though incrementally better). Because there is no established or recommended method for using ROMA in conjunction with clinical assessment, diagnostic performance characteristics are uncertain because it would vary depending on how it is used.

It is also uncertain whether the incremental yield of malignancy resulting from the use of the tests would result in improved patient outcomes. Although prior studies revealed an improvement of outcomes when women with ovarian cancer were initially managed by gynecologic oncologists, it is uncertain whether improved outcomes would occur in the additional cases detected by use of these tests. These additional cancer cases may differ from other cases detected by clinical assessment alone. If they tend to be earlier stage cancers or biologically less aggressive cancers, initial treatment by a gynecologic oncologist may not provide incremental benefit.

Section Summary: Clinically Useful
There is no direct evidence in terms of assessing patient outcomes based on the use of such testing prior to undergoing surgery. It is uncertain whether discrimination is sufficient to alter...
decision-making based on clinical assessment alone, thus offering a meaningful benefit to patients. Therefore, the chain of evidence supporting improved outcomes is incomplete.

Summary of Evidence
For individuals who have adnexal mass(es) undergoing surgery for possible ovarian cancer who receive multimarker serum testing with clinical assessment preoperatively to assess ovarian cancer risk, the evidence includes studies assessing the technical performance and diagnostic accuracy. Relevant outcomes are overall survival and test accuracy. OVA1 and Overa are intended for use in patients for whom clinical assessment does not indicate cancer. When used in this manner, sensitivity for ovarian malignancy was 92% and specificity was 42% with OVA1; with Overa, sensitivity was 94% and specificity was 65%. ROMA is intended for use with clinical assessment, but no specific method has been defined. One study, which used clinical assessment and ROMA results, showed a sensitivity of 90% and specificity of 67%. However, there is no direct evidence in terms of assessing patient outcomes based on the use of such testing prior to undergoing surgery. Moreover, it is uncertain whether discrimination is sufficient to alter decision making based on clinical assessment alone and, therefore, it is uncertain whether patients will find the testing to be of meaningful benefit. Thus, the chain of evidence supporting improved outcomes is incomplete. The evidence is insufficient to determine the effects of the technology on health outcomes.

Supplemental Information
Clinical Input from Physician Specialty Societies and Academic Medical Centers
While the various physician specialty societies and academic medical centers may collaborate with and make recommendations during this process, through the provision of appropriate reviewers, input received does not represent an endorsement or position statement by the physician specialty societies or academic medical centers, unless otherwise noted.

In response to requests from Blue Cross Blue Shield Association, input was received in 2012. Input was mixed in support of these tests as a tool for triaging patients with an adnexal mass. Reviewers agreed that the evidence was insufficient to determine the impact of these tests on referral patterns. For indications other than triaging patients with an adnexal mass, there was a lack of support for the use of these tests.

Practice Guidelines and Position Statements
American Congress of Obstetricians and Gynecologists
The American Congress of Obstetricians and Gynecologists (ACOG) addressed the use of the OVA1 test in its 2011 guidelines on the role of the obstetrician-gynecologist in the early detection of epithelial ovarian cancer, and these guidelines were confirmed in 2017. In 2013, the Society for Gynecologic Oncology endorsed these ACOG guidelines. This ACOG document included the following comments, which were not specific guidelines about the use of the test:

- The OVA1 test “appears to improve the predictability of ovarian cancer in women with pelvic masses.”
- “This is not a screening test, but it may be useful for evaluating women with a pelvic mass.”
- “Clinical utility is not yet established.”

Further, in 2016, an ACOG Practice Bulletin addressed the evaluation and management of adnexal masses makes a level B recommendation (based on limited or inconsistent scientific evidence) that consultation with or referral to a gynecologic oncologist is recommended for premenopausal or postmenopausal with an elevated score on a formal risk assessment test such as the multivariate index assay, risk of malignancy index, or the Risk of Ovarian Malignancy Algorithm, or one of the ultrasound-based scoring systems from the International Ovarian Tumor Analysis group.
National Institute for Health and Care Excellence

The National Institute for Health and Care Excellence issued guidance in 2011 on the identification and management of ovarian cancer. This guidance is currently being updated and is under review.

National Comprehensive Cancer Network

National Comprehensive Cancer Network (NCCN) guidelines on ovarian cancer (v.4.2017) include the following statement:

“It has been suggested that specific biomarkers (serum HE4 [human epididymis secretory protein 4] and CA-125 [cancer antigen 125]) along with an algorithm (Risk of Ovarian Malignancy Algorithm [ROMA]) may be useful for determining whether a pelvic mass is malignant or benign. The FDA [Food and Drug Administration] has approved the use of HE4 and CA-125 for estimating the risk of ovarian cancer in women with a pelvic mass. Currently, the NCCN Panel does not recommend the use of these biomarkers for determining the status of an undiagnosed pelvic mass.”

Regarding the OVA1 test, NCCN guidelines state:

“The OVA1 test uses 5 markers (including transthyretin, apolipoprotein A1, transferrin, beta-2 microglobulin, and CA-125) to assess who should undergo surgery by an experienced gynecologic oncologist and who can have surgery in the community.... Based on data documenting an increased survival, NCCN Guidelines Panel Members recommend that all patients should undergo surgery by an experienced gynecologic oncologist (category 1).”

U.S. Preventive Services Task Force Recommendations

In 2012, the U.S. Preventive Services Task Force recommended against screening women for ovarian cancer (D recommendation). The Task Force has not addressed multimarker serum testing related to ovarian cancer. The 2012 statement is currently in update.

Medicare National Coverage

There is no national coverage determination. In the absence of a national coverage determination, coverage decisions are left to the discretion of local Medicare carriers.

Ongoing and Unpublished Clinical Trials

A search of ClinicalTrials.gov in November 2018 did not identify any ongoing or unpublished trials that would likely influence this review.

References


**Documentation for Clinical Review**

- No records required

**Coding**

This Policy relates only to the services or supplies described herein. Benefits may vary according to product design; therefore, contract language should be reviewed before applying the terms of the Policy. Inclusion or exclusion of codes does not constitute or imply member coverage or provider reimbursement.

**IE The following services may be considered investigational.**

<table>
<thead>
<tr>
<th>Type</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPT®</td>
<td>81500</td>
<td>Oncology (ovarian), biochemical assays of two proteins (CA-125 and HE4), utilizing serum, with menopausal status, algorithm reported as a risk score</td>
</tr>
<tr>
<td>CPT®</td>
<td>81503</td>
<td>Oncology (ovarian), biochemical assays of five proteins (CA-125, apolipoprotein A1, beta-2 microglobulin, transferrin, and pre-albumin), utilizing serum, algorithm reported as a risk score</td>
</tr>
<tr>
<td>CPT®</td>
<td>0003U</td>
<td>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</td>
</tr>
</tbody>
</table>
### Definitions of Decision Determinations

**Medically Necessary:** A treatment, procedure, or drug is medically necessary only when it has been established as safe and effective for the particular symptoms or diagnosis, is not investigational or experimental, is not being provided primarily for the convenience of the patient or the provider, and is provided at the most appropriate level to treat the condition.

**Investigational/Experimental:** A treatment, procedure, or drug is investigational when it has not been recognized as safe and effective for use in treating the particular condition in accordance with generally accepted professional medical standards. This includes services where approval by the federal or state governmental is required prior to use, but has not yet been granted.

**Split Evaluation:** Blue Shield of California/Blue Shield of California Life & Health Insurance Company (Blue Shield) policy review can result in a split evaluation, where a treatment, procedure, or drug will be considered to be investigational for certain indications or conditions, but will be deemed safe and effective for other indications or conditions, and therefore potentially medically necessary in those instances.

### Prior Authorization Requirements (as applicable to your plan)

Within five days before the actual date of service, the provider must confirm with Blue Shield that the member's health plan coverage is still in effect. Blue Shield reserves the right to revoke an
authorization prior to services being rendered based on cancellation of the member's eligibility. Final determination of benefits will be made after review of the claim for limitations or exclusions.

Questions regarding the applicability of this policy should be directed to the Prior Authorization Department. Please call (800) 541-6652 or visit the provider portal at www.blueshieldca.com/provider.

Disclaimer: This medical policy is a guide in evaluating the medical necessity of a particular service or treatment. Blue Shield of California may consider published peer-reviewed scientific literature, national guidelines, and local standards of practice in developing its medical policy. Federal and state law, as well as contract language, including definitions and specific contract provisions/exclusions, take precedence over medical policy and must be considered first in determining covered services. Member contracts may differ in their benefits. Blue Shield reserves the right to review and update policies as appropriate.