Proteomics-based Testing Related to Ovarian Cancer

Description

There are a variety of gene-based biomarkers that have been studied in 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. Two tests based on this principle have now been cleared by FDA for use in women with adnexal masses (Ova1™ test and ROMA™ test) as an aid to further assess the likelihood that malignancy is present.

Background

In 2009, it was estimated that more than 21,000 women in the U.S. were diagnosed with ovarian cancer and more than 14,000 died of this disease. (1) The mortality rate 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.

In 1997, the Society of Surgical Oncology recommended ovarian cancer surgery and follow-up treatment be performed by physicians with ovarian cancer disease expertise. (3) To date, dozens of articles have been published on the application of this recommendation looking at long-term outcomes, short-term outcomes, and process measures (e.g., types of treatment such as complete staging or tumor debulking). At least 2 meta-analyses have concluded that outcomes are better in patients with ovarian cancer when they are treated by gynecologic oncologists. (4,5) Data have been 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.(6) About 6% have borderline tumors, 22%, invasive malignant lesions, and 3%, metastatic disease. Clinicians generally agree that women with masses that have a high likelihood of malignancy should undergo surgical staging by gynecologic oncologists. However, women with clearly benign masses do not require referral to a specialist. Criteria and tests that help differentiate benign from malignant pelvic masses are thus desirable.

In 2005, the American College of Obstetricians and Gynecologists and the Society of Gynecologic Oncologists jointly released referral guidelines that address criteria for referring women with pelvic masses that are suspicious for ovarian cancer to gynecologic oncologists. (7) Separate criteria were
developed for premenopausal and postmenopausal women. In premenopausal women, referral criteria included at least one of the following: elevated CA125 (greater than 200 U/mL), ascites, evidence of abdominal or distant metastasis, or a positive family history. The referral criteria in postmenopausal women were similar, except that a lower threshold for an elevated CA125 test was used (35 U/mL) and nodular or fixed pelvic mass was an additional criterion.

Two proteomic tests have now been cleared by FDA with the intended use to triage patients with adnexal masses. A suggested use of the test is to identify women with a positive test who have a higher likelihood of malignant disease and 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.

Regulatory Status

On July 16, 2009, the OVA1TM test (Vermillion Inc., Fremont, CA) was cleared for market by FDA as a 510(k) submission. On September 1, 2011, the Risk of Ovarian Malignancy Algorithm (ROMA test; Fujirebio Diagnostics Inc., Malvern, PA) was cleared by 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. Intended use of OVA1 is as an aid to further assess the likelihood that malignancy is present when the physician’s independent clinical and radiological evaluation does not indicate malignancy. Intended use of ROMA is as an aid 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. Neither test is FDA-cleared as a screening or stand-alone diagnostic assay.

FDA product code: ONX

Black Box Warning: On December 10, 2011, the 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. 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.

Related Policies

2.04.66 Serum Biomarker Human Epidiymis Protein 4 (HE4)

Policy

*This policy statement applies to clinical review performed for pre-service (Prior Approval, Precertification, Advanced Benefit Determination, etc.) and/or post-service claims.

Use of the OVA1 and ROMA tests are considered not medically necessary for preoperative evaluation of adnexal masses to triage for malignancy.

All other uses of the OVA1 and ROMA tests are investigational including but not limited to:
Policy Guidelines

OVA1 and ROMA tests are combinations of several separate lab tests and involve a proprietary algorithm for determining risk (i.e., they are what the American Medical Association’s CPT calls “Multianalyte Assays with Algorithmic Analyses” [MAAAs]).

Rationale

Assessment of a diagnostic technology typically focuses on 3 parameters: 1) technical performance; 2) diagnostic performance (sensitivity, specificity, and positive [PPV] and negative predictive value [NPP]) 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 2 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).

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 2 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.

Literature Review

Technical Performance

Descriptions of the developmental process for the OVA1 test have been published in FDA documents and in a perspective by Fung in 2010. (9-11) Candidate biomarkers were selected based on initial studies using mass spectroscopy but were converted to standard immunoassays to improve analytical
performance. Seven final markers were evaluated, none of which individually appeared to be highly specific for malignant ovarian disease. However, the choice of 5 of these (CA125, prealbumin, apolipoprotein A-I, 2-microglobulin, transferrin) produced a composite profile that did appear to have discriminatory ability. The test, as cleared by 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 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 numerical score from 0.0 to 10.0.

Details of the algorithm appear proprietary but development is described as an empiric process based on use of banked samples from academic partners, on a small prospective study of samples from Europe and using a designated subset of samples from the clinical study used to support submission to the FDA. It appears at an undisclosed point in the developmental process as a result of interaction with FDA, separate cut-points were developed for premenopausal and postmenopausal women.

A similar developmental process was described for ROMA by Moore et al. (12) They studied 9 biomarkers and chose human epididymis secretory protein 4 (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. (13) Separate cut-offs were again used for premenopausal and postmenopausal women.

The OVA1 is a qualitative serum test that combines immunoassay results for the 5 analytes described above (CA 125, prealbumin, apolipoprotein A-1, beta 2 microglobulin, and transferrin) into a single numerical score. Analytical performance for the 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 ROMA test is also a qualitative serum test that combines 2 analytes HE4 EIA and the ARCHITECT CA 125, along with menopausal status into a numerical score. 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.

**Diagnostic Performance**

Diagnostic performance of the OVA1 test was evaluated in a prospective, double-blind clinical study using 27 enrollment sites. (11) 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 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. The final prevalence of cancer in the population was 27%.

Using pathological 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. See Table 1.

**Table 1. Diagnostic Performance of OVA1**

<table>
<thead>
<tr>
<th></th>
<th>Clinical assessment alone</th>
<th>Clinical assessment with OVA1</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>

* Confidence intervals not provided.

No additional studies evaluating the diagnostic performance of the OVA1 test were identified in literature updates.

In 2014, Wang et al 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 CA125. (14) To be included in the meta-analysis, studies had to investigate both HE4 and CA125 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 2.

**Table 2. Diagnostic Performance of ROMA compared with HE4 and CA125: Meta-analysis findings**

<table>
<thead>
<tr>
<th></th>
<th>No. Studies</th>
<th>Sensitivity % (95% CI)</th>
<th>Specificity %, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROMA</td>
<td>14</td>
<td>85.3 (81.2-88.6)</td>
<td>82.4 (77.4-86.5)</td>
</tr>
<tr>
<td>HE4</td>
<td>28</td>
<td>76.3 (72.0-80.1)</td>
<td>93.6 (90.0-95.9)</td>
</tr>
<tr>
<td>CA125</td>
<td>28</td>
<td>79.2 (74.0-83.6)</td>
<td>82.1 (76.6-86.5)</td>
</tr>
</tbody>
</table>

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.
The study with the largest sample size was published by Karlsen and colleagues in 2012. (15) The study included 1218 women presenting with pelvic masses. Prior to diagnosis, HE4 and CA125 levels were obtained, and ROMA and the Risk of Malignancy Index (an index consisting of ultrasound findings, menopausal status and CA125 levels) were calculated. At a fixed sensitivity of 94.4%, the specificity of ROMA was 76.5% and the specificity of RMI was 81.5%. At a fixed specificity of 75%, the sensitivity of ROMA was 94.8% and the sensitivity of RMI was 96.0%. Accuracy of ROMA and RMI were not compared statistically, but appeared to be similar. In another study, Kaijser and colleagues evaluated 360 women with pelvic masses who were scheduled for surgery. (16) The study compared the diagnostic accuracy of ROMA and an ultrasound-based prediction model (LR2) developed by the International Ovarian Tumor Analysis Study (IOTA). Histology was used as the reference standard. The overall performance of LR2 (94% sensitivity and 82% specificity) was significantly better than ROMA (84% sensitivity and 80% specificity).

Clinical Utility

The ideal study design to evaluate clinical utility of proteomics-based testing would be a randomized controlled trial comparing patient management decisions (e.g., referral patterns) and/or health outcomes (e.g., mortality) in patients managed with the tests with those managed according to best current clinical practices. No randomized or nonrandomized studies with these comparisons were identified.

A TEC Assessment was completed in 2012 on “Multi-analyte testing for the evaluation of adnexal masses.” (17) The Assessment included evaluation of both the OVA1 and ROMA tests in regards to their impact on health outcomes. The following conclusions were made:

“The evidence regarding the effect of OVA1 and ROMA and effects on health outcomes is indirect and based on studies of diagnostic performance of the tests in patients undergoing surgery for adnexal masses. Although the studies show improvements in sensitivity and worsening of specificity with the use of the 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. Although OVA1 improves sensitivity, specificity declines so much that most patients test positive. ROMA does not improve the sensitivity of testing to a great extent. Underlying these issues is uncertainty regarding whether there would be actual health benefits based on altering patient referral based on these tests...Whether use of OVA1 or ROMA improves the net health outcome or is as beneficial as other diagnostic strategies has not been demonstrated. Referring all patients to a physician with expertise in staging and debulking ovarian cancer is a reasonable clinical alternative with no harm.”

In 2014, Kaijser et al published a study that did not directly evaluate clinical utility but that provides relevant information. (18) It was a retrospective cohort study that included 101 newly diagnosed cases of biopsy-proven invasive ovarian cancer. Blood samples obtained before treatment were analyzed; HE4 and CA125 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% confidence interval (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.

### Clinical Practice Guidelines and Position Statements

In May 2013, the Society for Gynecologic Oncology (SGO) issued the following statement on multiplex serum testing for women with pelvic masses: (19)

“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. It is important to note that no such test has been evaluated for use as, nor cleared by, the FDA as a screening tool for ovarian cancer. SGO does not formally endorse or promote any specific products or brands.”

The American Congress of Obstetricians and Gynecologists addressed the use of the OVA1 test in their guidelines on the role of the obstetrician-gynecologist in the early detection of epithelial ovarian cancer. (20) This document made the following statements:

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

The National Cancer Institute included a discussion of proteomics in their publication on the Genetics of Breast and Ovarian Cancer. (21) The following statement was included in their discussion on proteomics:

- These (proteomic) studies have generally been small case-control studies that are limited by sample size and the number of early-stage cancer cases included. Further evaluation is needed to determine whether any additional markers identified in this fashion have clinical utility for the early detection of ovarian cancer in the unselected clinical population of interest. Level of Evidence: 5

The National Institute for Health and Clinical Excellence (NICE) issued a publication in 2011 on the recognition and management of ovarian cancer. (22) These guidelines made the following recommendations:
The evidence suggests that the combination of HE4 and serum CA125 is more specific, but less sensitive than either marker in isolation.

There was no evidence to suggest that multiple tumour markers were much better than the two marker combination of serum CA125 and HE4.

The routine use of CA 125 is recommended; the data on other serum markers is not substantial enough to recommend their use.

The National Comprehensive Cancer Network (NCCN) guideline on ovarian cancer (V.3.2014) includes the following statement: (23)

“It has been suggested that specific biomarkers (serum HE4 and CA-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 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.”

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). (24) The task force has not addressed proteomics-based testing related to ovarian cancer.

Summary
The OVA1 and ROMA tests have both been analytically validated and clinical performance has been reported in prospective multicenter clinical studies. Changes in the observed sensitivity and negative predictive value of testing compared with clinical assessment has been small and of uncertain diagnostic value. Studies on the diagnostic accuracy of these tests compared with other diagnostic tools have had mixed findings, but do not report that ROMA is superior to other risk prediction tools that use standard clinical information or single markers. No studies have been performed that directly evaluated the impact on patient management e.g., referral patterns, and no studies have evaluated the impact on health outcomes. Clinical input from academic medical centers and specialty societies did not show consensus that this test improved outcomes when used as a tool to triage patients with adnexal masses. As a result of the evidence and clinical input, these tests are considered not medically necessary pending more information about their performance and impact on outcomes. All other use of these tests are considered investigational.

Medicare National Coverage
No national coverage determination.

References


8. Medical Devices: Ovarian adnexal mass assessment score test system; Labeling; Black box restrictions. 21 CFR Part 866, Federal Register 2011;76(251):82128-82123. PMID 9


<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Section:</strong></td>
<td>Medicine</td>
<td><strong>Effective Date:</strong></td>
</tr>
<tr>
<td><strong>Subsection:</strong></td>
<td>Pathology/Laboratory</td>
<td><strong>Original Policy Date:</strong></td>
</tr>
<tr>
<td><strong>Subject:</strong></td>
<td>Proteomics-based Testing Related to Ovarian Cancer</td>
<td><strong>Page:</strong></td>
</tr>
</tbody>
</table>


17. Blue Cross and Blue Shield Association Technology Evaluation Center (TEC). Multi-analyte testing for the evaluation of adnexal masses. *TEC Assessment Program*. 2012; Volume 27, Tab 8. PMID


Section: Medicine  Effective Date: April 15, 2015
Subsection: Pathology/Laboratory  Original Policy Date: December 7, 2011
Subject: Proteomics-based Testing Related to Ovarian Cancer  Page: 11 of 11

Policy History

<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>December 2011</td>
<td>New Policy</td>
<td></td>
</tr>
<tr>
<td>March 2013</td>
<td>Update Policy</td>
<td>Policy update with literature search, results of TEC assessment. References 7, 13, and 17-28 added, Policy statement changed to not medically necessary for pre-operative evaluation.</td>
</tr>
<tr>
<td>March 2014</td>
<td>Update Policy</td>
<td>Policy updated with literature search. References 14, 15 and 20 added; other references renumbered or removed. No change to policy statement. Title changed to Proteomic-based Testing Related to Ovarian Cancer.</td>
</tr>
</tbody>
</table>

Keywords

Proteomic-based testing, ovarian
ROMA, Risk of Ovarian malignancy
OVA1

This policy was approved by the FEP® Pharmacy and Medical Policy Committee on March 20, 2015 and is effective April 15, 2015.

Signature on file

Deborah M. Smith, MD, MPH