# Serum Biomarker Human Epididymis Protein 4

**Policy Number:** 2.04.66  
**Origination:** 5/2015  
**Last Review:** 5/2019  
**Next Review:** 5/2020

## Policy

Blue Cross and Blue Shield of Kansas City (Blue KC) will not provide coverage for Serum Biomarker Human Epididymis Protein 4. This is considered investigational.

### When Policy Topic is covered

Not Applicable

### When Policy Topic is not covered

Measurement of HE4 is **investigational** for all indications.

## Description of Procedure or Service

<table>
<thead>
<tr>
<th>Populations</th>
<th>Interventions</th>
<th>Comparators</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individuals:</strong></td>
<td></td>
<td><strong>Comparators of interest are:</strong></td>
<td><strong>Relevant outcomes include:</strong></td>
</tr>
<tr>
<td>- With ovarian cancer</td>
<td>- Serum biomarker human epididymis protein 4 level testing</td>
<td>- Measurement of cancer antigen 125 test</td>
<td>- Overall survival</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Measurement of combination cancer antigen 125 plus human epididymis protein 4</td>
<td>- Disease-specific survival</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Test validity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Other test performance measures</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Change in disease status</td>
</tr>
<tr>
<td><strong>Individuals:</strong></td>
<td></td>
<td><strong>Comparators of interest are:</strong></td>
<td><strong>Relevant outcomes include:</strong></td>
</tr>
<tr>
<td>- With adnexal masses</td>
<td>- Serum biomarker human epididymis protein 4 levels testing</td>
<td>- Measurement of cancer antigen 125 test</td>
<td>- Overall survival</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Measurement of combination cancer antigen 125 plus human epididymis protein 4</td>
<td>- Disease-specific survival</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Test validity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Other test performance measures</td>
</tr>
</tbody>
</table>
Human epididymis protein 4 (HE4) is a novel biomarker that has been cleared by the U.S. Food and Drug Administration (FDA) for monitoring patients with epithelial ovarian cancer. HE4 is proposed as a replacement for or a complement to carbohydrate antigen 125 (CA-125) for monitoring disease progression and recurrence. HE4 has also been proposed as a test to evaluate women with ovarian masses and to screen for ovarian cancer in asymptomatic women.

For individuals who have ovarian cancer who receive measurement of serum biomarker HE4, the evidence includes several retrospective studies comparing the diagnostic accuracy of HE4 and CA 125 for predicting disease progression and/or recurrence. Relevant outcomes are overall survival, disease-specific survival, test accuracy and validity, other test performance measures, and change in disease status. Data submitted to FDA for approval of commercial HE4 tests found that HE4 was not inferior to CA 125 for detecting ovarian cancer recurrence. However, the superiority of HE4 to CA 125 (alone or in combination), the key question in the evidence review, was not demonstrated in the available literature. In addition, there is no established cutoff in HE4 levels for monitoring disease progression, and cutoffs in studies varied. There is no direct evidence from prospective controlled studies on the impact of HE4 testing on health outcomes and no clear chain of evidence that changes in management based on HE4 would lead to improved health outcome. The evidence is insufficient to determine the effects of the technology on health outcomes.

For individuals who have adnexal masses who receive measurement of serum biomarker HE4, the evidence includes diagnostic accuracy studies and meta-analyses. Relevant outcomes are overall survival, disease-specific survival, test accuracy and validity, and other test performance measures. Meta-analyses have generally found that HE4 and CA 125 have similar overall diagnostic accuracy (ie, sensitivity, specificity) and several found that HE4 has significantly higher specificity than CA 125 but not sensitivity. Two meta-analyses had mixed findings on whether the combination of HE4 and CA 125 is superior to CA 125 alone for the initial diagnosis of ovarian cancer. The number of studies evaluating the combined test is relatively low and publication bias in studies of HE4 has been identified. In addition, studies have not found that HE4 improves diagnostic accuracy beyond that of subjective assessment of transvaginal ultrasound. There is no direct evidence from prospective controlled studies on the impact of HE4 testing on health outcomes and no clear chain of evidence that changes in management based on HE4 would lead to improved health outcome. The evidence is insufficient to determine the effects of the technology on health outcomes.
For individuals who are asymptomatic and not at high risk of ovarian cancer who receive screening with serum biomarker HE4, the evidence includes several retrospective comparative studies and no prospective studies comparing health outcomes in asymptomatic women managed with and without HE4 screening. Relevant outcomes are overall survival, disease-specific survival, test accuracy and validity, and other test performance measures. The retrospective studies found that levels of HE4 increased over time in women ultimately diagnosed with ovarian cancer. Prospective comparative studies are needed to definitively determine whether HE4 is a useful screening tool. The evidence is insufficient to determine the effects of the technology on health outcomes.

**Background**

**Ovarian Cancer**

Ovarian cancer is the fifth most common cause of cancer mortality among U.S. women. According to Surveillance Epidemiology and End Results (SEER) data, in 2013, an estimated 22,440 women would be diagnosed with ovarian cancer and 14,080 women would die of the disease. Stage at diagnosis is an important predictor of survival; however, most women are not diagnosed until the disease has spread. For the period 1999 to 2006, 62% of women with ovarian cancer were diagnosed when the disease had distant metastases (stage IV), and this was associated with a 5-year survival rate of 28.9%. In contrast, the 14.8% of women diagnosed with localized cancer (stage I) had a 5-year survival rate of 92.5%. Epithelial ovarian tumors account for 85% to 90% of ovarian cancers.

**Treatment**

The standard treatment for epithelial ovarian cancer is surgical staging and primary cytoreductive surgery followed by chemotherapy in most cases. There is a lack of consensus about an optimal approach to follow-up of patients with ovarian cancer after or during primary treatment. Patients undergo regular physical examinations and may have imaging studies. In addition, managing patients with serial measurement of the biomarker cancer antigen 125 (CA 125) to detect early recurrence of disease is common. A rising CA 125 level has been found to correlate with disease recurrence and has been found to detect recurrent ovarian cancer earlier than clinical detection. However, a survival advantage of initiating treatment based on early detection with CA 125 has not been demonstrated to date. For example, a 2010 randomized controlled trial (RCT) with women having ovarian cancer that was in complete remission did not find a significant difference in overall survival when treatment for remission was initiated after CA 125 concentration exceeded twice the limit of normal compared with delaying treatment initiation until symptom onset.

Human epididymis protein 4 (HE4) is a protein that circulates in the serum and has been found to be overexpressed in epithelial ovarian cancer (EOC), lung adenocarcinoma, breast cancer, pancreatic cancer, endometrial cancer, and bladder cancer. HE4 is made up of 2 whey acidic proteins with a 4 disulfide core domain and has been proposed as a biomarker for monitoring patients with epithelial ovarian cancer.
Evaluation of Adnexal Masses
This evidence review also addresses the use of the HE4 as a stand-alone test for evaluating women with ovarian masses who have not been diagnosed with ovarian cancer. Such patients undergo a diagnostic workup to determine whether the risk of malignancy is sufficiently high to warrant surgical removal. In patients for whom surgery is indicated, further evaluation may be warranted to determine if a surgical referral to a specialist with expertise in ovarian cancer is warranted. The Risk of Ovarian Malignancy Algorithm (ROMA) combines HE4, CA 125, and menopausal status into a numeric score. ROMA has been cleared by FDA for predicting the risk that an adnexal mass is malignant; this test is considered separately in a separate policy (multimarker serum testing related to ovarian cancer).

Regulatory Status
Multiple HE4 test kits have been cleared by the Food and Drug Administration through the 510(k) process and summarized in Table 1. The FDA determined that this device was substantially equivalent to a CA 125 assay kit for use as an aid in monitoring disease progression or recurrence in patients with epithelial ovarian cancer. The FDA-approved indication states that serial testing for HE4 should be done in conjunction with other clinical methods used for monitoring ovarian cancer and that the HE4 test is not intended to assess the risk of disease outcomes.

Table 1. Serum HE4 Tests Cleared by FDA

<table>
<thead>
<tr>
<th>Test</th>
<th>Manufacturer</th>
<th>Location</th>
<th>Date Cleared</th>
<th>510(k) No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE4 EIA Kit</td>
<td>Fujirebio Diagnostics</td>
<td>Malvern, PA</td>
<td>06/09/2008</td>
<td>K072939</td>
</tr>
<tr>
<td>ARCHITECT HE4 assay (CMIA)</td>
<td>Fujirebio Diagnostics</td>
<td>Malvern, PA</td>
<td>03/18/2010</td>
<td>K093957</td>
</tr>
<tr>
<td>ELECSYS HE4 (CMIA)</td>
<td>Roche Diagnostics</td>
<td>Indianapolis, IN</td>
<td>09/10/2012</td>
<td>K112624</td>
</tr>
<tr>
<td>Lumipulse G HE4 Immunoreaction Cartridges</td>
<td>Fujirebio Diagnostics</td>
<td>Malvern, PA</td>
<td>11/24/2015</td>
<td>K151378</td>
</tr>
</tbody>
</table>

CMIA: chemoluminescent microparticle immunoassay; HE4: human epididymis protein 4; EIA: enzymatic immunoassay; FDA: Food and Drug Administration.

Rationale
The evidence review was created in August 2010 and has been updated regularly with searches of the MEDLINE database. The most recent literature update was performed through October 30, 2018

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 literature.

**Biomarker Human Epididymis Protein 4 TESTING for Ovarian Cancer**

**Clinical Context and Test Purpose**
The purpose of testing serum biomarker human epididymis protein 4 levels is to provide an alternative to or an improvement on existing testing in patients with ovarian cancer.

The question addressed in this evidence review is: Does the use of testing with HE4 to monitor patient with ovarian cancer result in an improvement in net health outcomes.

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

**Patients**
The relevant population of interest are individuals with epithelial ovarian cancer and who have had primary treatment.

**Interventions**
The test being considered is testing serum biomarker human epididymis protein 4 levels for surveillance of progression (response to primary treatment) or recurrence.

**Comparators**
Comparators of interest include measurement of cancer antigen 125 test and measurement of combination cancer antigen 125 plus human epididymis protein 4 for surveillance of progression or recurrence. Typically, patients undergoing primary chemotherapy after cytoreductive surgery will also have monitoring for response with CT scan. After the completion of primary treatment patients may have other monitoring imaging studies such as positron emission tomography (PET).

**Outcomes**
The general outcomes of interest are overall survival, disease-specific survival, test accuracy, test validity, and change in disease status.

**Timing**
The timing of follow-up after testing HE4 serum levels in patient with ovarian cancer is based on stage of disease, type of prior therapy and guideline recommendations.

**Setting**
Patients with ovarian cancer are managed by oncologists in an outpatient clinical setting.
**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).

**Study Selection Criteria**
For the evaluation of clinical validity of HE4 testing, methodologically credible studies were selected using the following principles:

For the evaluation of clinical validity of the tests, studies that meet the following eligibility criteria were considered:

- Reported on the accuracy of the marketed version of the technology (including any algorithms used to calculate scores)
- Included a suitable reference standard
- Patient/sample clinical characteristics were described
- Patient/sample selection criteria were described
- Included a validation cohort separate from development cohort.

FDA documents included information on the diagnostic performance of HE4 for monitoring the progression and recurrence of ovarian cancer. FDA materials addressed the noninferiority rather than the superiority of HE4 tests to CA 125. A study reported in the 510(k) substantial equivalence determination decision summary for the HE4 EIA assay evaluated whether this test is noninferior to the CA 125 test. The study included samples from 80 women with epithelial ovarian cancer who were undergoing serial surveillance of cancer progression. Blood samples were obtained from a large cancer center in the United States; they were not drawn specifically for this study. A total of 354 samples were obtained for the 80 women (women had multiple visits over time). Receiver operating characteristic (ROC) curve analysis was used to compare the 2 assays, and clinical evidence of progression was used as the reference standard. When a positive change in HE4 level (ie, to indicate disease progression) was defined as a value at least 25% higher than the previous value of the test, the sensitivity of the test was 76 (60.3%) of 126, and the specificity was 171 (75%) of 228. (Note that the unit of analysis was the number of samples rather than the number of women.) The area under the receiver operating characteristic (AUROCC) curves was found to be similar (HE4=0.725 vs CA 125=0.709), with overlap in the confidence intervals.
(CI); according to the authors, this indicated that the HE4 assay was not inferior to the CA 125 assay for detecting cancer progression.

Another analysis estimated the cutoff values and specificities for the HE4 and CA 125 assays across a range of fixed sensitivities, where the sensitivities of the HE4 and CA 125 assays were set at the same values. The specificity values for CA 125 and HE4 did not differ statistically at the respective cutoffs and sensitivities. These data were also said to confirm that the HE4 EIA test was not inferior to the CA 125 test for detecting ovarian cancer progression.

The 510(k) substantial equivalence determination decision summary for the ARCHITECT HE4 assay reported data from a retrospective study using remnant serial samples from 76 women diagnosed with epithelial ovarian cancer being monitored after completion of chemotherapy. The eligibility criteria included availability of at least 3 serial specimens; samples could have been drawn during and/or after treatment. Clinical determination of disease progression was used as the reference standard. A positive test was defined as an HE4 level that was 14% higher than the previous reading. Using this cutoff, the sensitivity of the assay for detecting progressive disease was 53 (53.5%) of 99 events. The specificity of the assay was 260 (78.5%) of 331. Of note, the sensitivity is lower than that previously reported for the HE4 EIA test at a similar specificity, when a cutoff of a 25% increase was used (sensitivity, 60.3%; specificity, 75%).

The FDA documents noted that there is no clinically accepted cutoff for monitoring cancer progression in epithelial ovarian cancer patients using the HE4 assays. As mentioned, a study included in the HE4 EIA assay materials defined a positive test as a level 25% higher than a previous measurement, and a study on the ARCHITECT HE4 test defined a positive test as an increase of at least 14% in the level of HE4. FDA documents further stated that clinicians may decide whether to use the cutoffs in the studies or another cutoff that reflects personal preferences in the tradeoff between sensitivity and specificity.

Nonrandomized Observational Studies
Published observational studies on the diagnostic performance of HE4 for monitoring progression and/or recurrence of epithelial cancer are described next.

A 2013 study by Braicu et al evaluated 275 patients with advanced primary ovarian cancer who underwent cytoreductive surgery and adjuvant platinum-based chemotherapy at a specialized clinic (OVCAD study). In 221 (80.4%) of 275 patients, preoperative HE4 and CA 125 levels as well as data on residual tumor mass after debulking were available. For HE4 levels, the area under the curve (AUC) for residual tumor mass was 0.634. At an HE4 cutoff of 235 pM, the sensitivity was 76.6%, and the specificity was 47.4%. When the cutoff for HE4 was 500 pM, the sensitivity was 51.9%, and the specificity was 70.4%. For CA 125, the AUC for residual tumor mass was 0.64, nearly the same as that for HE4. At a cutoff of 500 IU/mL, the sensitivity of CA 125 for predicting complete tumor resection was 69.4%, and the specificity was 52.3%. Using the most accurate cutoffs for HE4 (235 pM) and CA 125 (500 IU/mL), the combination of the 2
markers had a sensitivity of 64.8% and a specificity of 73.5%. Additional analysis of these data was published by Nassir et al in 2016. Ninety-two (33%) of 275 patients, who had preoperative and follow-up plasma samples for analyzing HE4 and CA 125, were included in the analysis. (However, 13 preoperative HE4 samples and 10 postoperative CA 125 samples were missing.) Both preoperative HE4 and CA 125 levels significantly predicted 12-month recurrence or death. Among responders, median overall survival was worse among patients for whom both biomarkers were elevated (hazard ratio, 17.96; 95% CI, 4.00 to 80.85; p<0.001) compared with patients for whom no biomarker was elevated. The confidence interval for the overall survival analysis was wide, indicating an imprecise estimate. There was no significant association with median overall survival when only 1 biomarker was elevated; the sample size may have been inadequate for this analysis.

In 2016, Steffensen et al evaluated the ability of HE4 and CA 125, individually and together, to predict ovarian cancer recurrence after first-line chemotherapy. The study included 88 patients with serum samples drawn at the end of chemotherapy and at least twice during the follow-up period. The median length of follow-up for patients still living was 47 months. During the study, 55 (62.5%) of 88 patients had recurrences, and 38 (43%) died. HE4 levels at the end of chemotherapy classified 70 (84.3%) patients as at high risk of relapse and 13 (15.7%) as at low risk. The sensitivity of HE4 was 90.0% (95% CI, 79.0% to 96.8%) and the specificity was 25.8% (95% CI, 11.9% to 44.6%). The combination of HE4 and CA 125 levels classified 69 (83%) patients as high risk of relapse and 14 (16.9%) as low risk, with a sensitivity of 90.0% (95% CI, 79.0% to 96.8%) and a specificity of 29% (95% CI, 14.2% to 48.0%). Based on analysis of HE4 and CA 125 levels from samples taken 3 months after chemotherapy, an increase of at least 50% relative to baseline was considered the cutoff for predicting a significant worsening of progression-free survival (PFS). Both HE4 and CA 125 as individual markers at 3 and 6 months were significantly associated with poorer PFS. However, on multivariate analysis, HE4 was a significant predictor of PFS at 6 months but not at 3 months, and CA 125 was not significant at 3 or 6 months. For the combination of CA 125 and HE4, there were too few patients positive on both markers at 3 months (n=7) to analyze the combination’s association with PFS.

Vallius et al (2017) reported a study that was designed to assess fluorodeoxyglucose-positron emission tomography/computed tomography (18F-FDG-PET/CT) imaging and serum tumor markers in epithelial ovarian cancer staging and chemotherapy response. A substudy analysis evaluated the use of HE4 profiles to predict treatment outcomes during the first line of chemotherapy after primary cytoreductive surgery. HE4 and CA125 was measured in patients with FIGO III/IV EOC who received primary debulking surgery followed by platinum based chemotherapy or neoadjuvant chemotherapy followed by interval debulking surgery. HE4 at the time of diagnosis was not associated with PFS (p = 0.24), whereas lower CA125 at the time of diagnosis predicted longer PFS (p = 0.01, hazard ratio (HR) = 1.45, 95% confidence interval (CI) = 1.09-1.94). When patients who underwent either surgical approach were combined (N=40), those with no macroscopic residual disease after cytoreductive surgery were more likely
to have lower postoperative HE4 values. Both HE4 and CA 125 nadir values were associated with a greater complete response to chemotherapy. Tables 2-5 summarize findings for this study.

### Table 2. Summary of Observational Study Characteristics

<table>
<thead>
<tr>
<th>Study</th>
<th>Study Type</th>
<th>Country</th>
<th>Dates</th>
<th>Participants</th>
<th>Treatment1</th>
<th>Treatment2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vallius (2017)</td>
<td>Observational cohort</td>
<td>Finland</td>
<td>2009-2014</td>
<td>FIGO Stage III-IV EOC 49</td>
<td>PDS + platinum based chemotherapy 22</td>
<td>NACT + IDS 27</td>
</tr>
</tbody>
</table>

EOC: epithelial ovarian cancer; FIGO: Federation of International Gynecology and Obstetrics; IDS: interval debulking surgery; NACT: neoadjuvant chemotherapy; NCT: National Clinical Trial; PDS: primary debulking surgery;

### Table 3. Summary of Observational Study Results

<table>
<thead>
<tr>
<th>Study</th>
<th>Treatment1</th>
<th>Treatment2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median (range)</td>
<td>Median (range)</td>
</tr>
<tr>
<td>Vallius (2017)</td>
<td>HE4 (pmol/L)</td>
<td>CA125 (U/mL)</td>
</tr>
<tr>
<td>NCT01276574</td>
<td>At diagnosis: 573 (59-1391)</td>
<td>1094 (17-17,992)</td>
</tr>
<tr>
<td></td>
<td>Pre-IDS: 96 (34-856)</td>
<td>181 (32-2023)</td>
</tr>
<tr>
<td></td>
<td>Postoperative Nadir: 48 (25-204)</td>
<td>12 (4-162)</td>
</tr>
<tr>
<td></td>
<td>Post-primarytherapy: 96 (34-856)</td>
<td>12 (4-127)</td>
</tr>
<tr>
<td></td>
<td>CA125 (U/mL)</td>
<td>At diagnosis: 48 (25-431)</td>
</tr>
<tr>
<td></td>
<td>Pre-IDS: 61 (31-175)</td>
<td>15 (4-447)</td>
</tr>
<tr>
<td></td>
<td>Postoperative Nadir: 104 (35-477)</td>
<td>15 (4-37)</td>
</tr>
<tr>
<td></td>
<td>Post-primary therapy: 99 (39-384)</td>
<td>42 (6-589)</td>
</tr>
</tbody>
</table>

CA 125: cancer antigen 125; HE4: human epididymis protein 4; IDS: interval debulking surgery; NACT: neoadjuvant chemotherapy; PDS: primary debulking surgery; (pmol/L): picomole per liter; Treatment 1: PDS + platinum based chemotherapy; Treatment2: NACT + IDS; (U/mL): units per milliliter

Relevance and relevance design and conduct gaps are reported in Tables 4 and 5.
Table 4: Relevance Gaps

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Intervention</th>
<th>Comparator</th>
<th>Outcomes</th>
<th>Duration of Follow-Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vallius et al (2017)</td>
<td>3. Study population is mixed regarding risk factors</td>
<td>2. clinical context for primary debulking surgery + chemotherapy differs from neoadjuvant chemotherapy + interval debulking surgery</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 5: Study Design and Conduct Gaps

<table>
<thead>
<tr>
<th>Study</th>
<th>Selectiona</th>
<th>Blindingb</th>
<th>Delivery of Testc</th>
<th>Selective Reportingd</th>
<th>Completeness of Follow-Upd</th>
<th>Statisticalf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vallius et al (2017)</td>
<td>2. Broad date range for obtaining samples</td>
<td>2. Assessment of residual disease solely based on surgeon evaluation</td>
<td>1. Not uniformly reported</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.


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).

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

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.

Clinically Useful
The available observational studies have used HE4 alone or in combination with CA 125 to predict residual tumor mass and association with recurrence after primary chemotherapy. In addition, HE4 alone or in combination with CA 125 has been
assessed for its association with residual disease and tumor progression during the course of primary chemotherapy after tumor debulking as well during neoadjuvant chemotherapy followed by interval debulking surgery. Improvement in health outcomes would depend on demonstrating that further assessment and management decisions or patients with ovarian cancer were initiated that would improve health outcomes. There is no clear chain of evidence demonstrating that incremental changes in ovarian cancer recurrence detection would lead to improved health outcomes. No prospective studies were identified that compared health outcomes in patients who had ovarian cancer managed with and without HE4 testing, alone or in combination with CA 125 or other disease markers.

**Section Summary: Individuals with Ovarian Cancer**
Several studies, including those submitted to the FDA, have addressed HE4 for monitoring ovarian cancer progression and recurrence. There is insufficient evidence that the diagnostic accuracy of HE4, alone or combined with CA 125, is superior to CA 125 alone. Moreover, there is a lack of clarity about which HE4 cutoff to use to predict disease progression or recurrence. No direct evidence from prospective studies was identified that compared health outcomes in patients who had ovarian cancer managed with and without HE4 testing, alone or in combination with CA 125 or other disease markers. In addition, there is no clear chain of evidence that shows changes in management based on HE4 testing would improve health outcomes.

**Biomarker Human Epididymis Protein 4 TESTING for Adnexal masses**

**Clinical Context and Test Purpose**
The purpose of testing serum biomarker human epididymis protein 4 levels is to provide a diagnostic option that is an alternative to or an improvement on existing testing in patients with adnexal masses.

The question addressed in this evidence review is: does testing with serum HE4 levels as adjunctive testing for diagnosing patients with adnexal masses improve net health outcomes?

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

**Patients**
The relevant population of interest are individuals with adnexal masses.

**Interventions**
The test being considered is testing serum biomarker human epididymis protein 4 levels.

**Comparators**
Comparators of interest include measurement of cancer antigen 125 test and measurement of combination cancer antigen 125 plus human epididymis protein 4.
Outcomes
The general outcomes of interest are overall survival, disease-specific survival, test validity, and other test performance measures.

Timing
The timing of follow-up after testing for serum HE4 for the evaluation of an adnexal mass would be determined by whether or not the patient has surgical management. Typical clinical follow-up in the absence of a pathological diagnosis would be every 6 months.

Setting
Patients with adnexal masses are managed by oncologists in an outpatient clinical setting.

Study Selection Criteria
For the evaluation of clinical validity of HE4 testing, methodologically credible studies were selected using the following principles:

For the evaluation of clinical validity of the tests, studies that meet the following eligibility criteria were considered:

- Reported on the accuracy of the marketed version of the technology (including any algorithms used to calculate scores)
- Included a suitable reference standard
- Patient/sample clinical characteristics were described
- Patient/sample selection criteria were described
- Included a validation cohort separate from development cohort.

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
Because CA 125 is the marker most often recommended for evaluation of adnexal masses, this evidence review addresses whether the diagnostic performance of HE4 is superior to CA 125 and whether combined HE4 and CA 125 is superior to CA 125 alone.
**Table 5. Meta-Analyses of HE4 for Diagnosing Ovarian Cancer**

<table>
<thead>
<tr>
<th>Meta-Analyses (Year)</th>
<th>No. of Studies</th>
<th>Pooled Sensitivity (95% CI), %</th>
<th>Pooled Specificity (95% CI), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dayyani et al (2016)</td>
<td>5</td>
<td>82 (68 to 90)</td>
<td>85 (72 to 93)</td>
</tr>
<tr>
<td>Maceda et al (2014)</td>
<td>45</td>
<td>78 (77 to 79)</td>
<td>86 (85 to 87)</td>
</tr>
<tr>
<td>Wang et al (2014)</td>
<td>28</td>
<td>76 (72 to 80)</td>
<td>93 (90 to 96)</td>
</tr>
<tr>
<td>Zhen et al (2014)</td>
<td>25</td>
<td>74 (72 to 76)</td>
<td>90 (89 to 91)</td>
</tr>
<tr>
<td>Yang et al (2013)</td>
<td>31</td>
<td>73 (71 to 75)</td>
<td>89 (88 to 90)</td>
</tr>
<tr>
<td>Ferraro et al (2013)</td>
<td>14</td>
<td>79 (76 to 81)</td>
<td>93 (92 to 94)</td>
</tr>
<tr>
<td>Yu et al (2012)</td>
<td>12</td>
<td>80 (77 to 83)</td>
<td>92 (90 to 93)</td>
</tr>
</tbody>
</table>

CI, confidence interval; HE4: human epididymis protein 4.

Meta-analyses differed somewhat in their study inclusion criteria, search dates, and other factors, but, as shown in Table 1, had similar results in terms of the diagnostic value of HE4; pooled sensitivities ranged from 73% to 82%, and pooled specificities ranged from 85% to 93%.

Several of the previous meta-analyses also pooled data from studies on the diagnostic accuracy of CA 125, alone and/or in combination with HE4 and findings are shown in Table 2.

**Table 6. Meta-Analyses of CA 125 Alone or Combined With HE4 for Diagnosing Ovarian Cancer**

<table>
<thead>
<tr>
<th>Meta-Analyses (Year)</th>
<th>No. of Studies</th>
<th>Pooled Sensitivity (95% CI), %</th>
<th>Pooled Specificity (95% CI), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA 125 alone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dayyani et al (2016)</td>
<td>5</td>
<td>80 (66 to 89)</td>
<td>83 (66 to 92)</td>
</tr>
<tr>
<td>Wang et al (2014)</td>
<td>28</td>
<td>79 (74 to 84)</td>
<td>82 (77 to 87)</td>
</tr>
<tr>
<td>Zhen et al (2014)</td>
<td>25</td>
<td>74 (72 to 76)</td>
<td>83 (81 to 84)</td>
</tr>
<tr>
<td>Ferraro et al (2013)</td>
<td>13</td>
<td>79 (77 to 82)</td>
<td>78 (76 to 80)</td>
</tr>
<tr>
<td>Yu et al (2012)</td>
<td>10</td>
<td>66 (62 to 70)</td>
<td>87 (85 to 89)</td>
</tr>
<tr>
<td>HE4 and CA 125</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhen et al (2014)</td>
<td>9</td>
<td>90 (87 to 92)</td>
<td>85 (82 to 87)</td>
</tr>
<tr>
<td>Ferraro et al (2013)</td>
<td>4</td>
<td>82 (78 to 86)</td>
<td>76 (72 to 80)</td>
</tr>
</tbody>
</table>


All meta-analyses included in Table 2, except Dayyani et al (2016), reported statistical comparisons between the diagnostic performance of HE4 and CA 125. None found that the performance (combination of sensitivity and specificity) of HE4 and CA 125 differed significantly. However, both Wang et al (2014) and...
Zhen et al (2014) found that the specificity (but not sensitivity) of HE4 was significantly higher than CA 125.

Findings differed in the 2 meta-analyses comparing the diagnostic performance of HE4 and CA 125 with CA 125 alone. Ferraro et al did not find that the sensitivity and specificity of HE4 in combination with CA 125 differed significantly from that of CA 125 alone. Zhen et al found that both the sensitivity and specificity of HE4 combined with CA 125 were significantly better than CA 125 alone. In the subgroup of 9 studies that made direct comparisons in the Zhen meta-analysis, the sensitivity of HE4 plus CA 125 was 90% (95% CI, 87% to 92%) and 74% (95% CI, 69% to 78%) for CA 125 alone, and the specificity of HE4 plus CA 125 was 85% (95% CI, 82% to 87%) and 73% (95% CI, 69% to 76%) for CA 125 alone. In addition, in the Zhen meta-analysis, the overall diagnostic accuracy (measured by the diagnostic odds ratio) was significantly higher for the combination of HE4 and CA 125 than for HE4 alone. Pooled diagnostic odds ratio were 10.31 (95% CI, 6.18 to 17.21) for CA 125 and 53.92 (95% CI, 26.07 to 111.54) for HE4 plus CA 125. Zhen et al noted several limitations to their meta-analysis, including substantial publication bias for HE4, heterogeneity among studies, and a lack of consideration given to clinical factors such as menopausal status.

Nonrandomized Observational Studies

Several studies have evaluated the diagnostic performance of HE4 as a second-line test after subjective assessment of transvaginal ultrasound. The final histologic diagnosis was used as the reference standard. Kaijser et al (2014) enrolled 389 patients with a suspicious pelvic mass who were scheduled for surgery. Data on 360 (93%) patients were available for analysis. Experienced ultrasonographers categorized each mass as benign, borderline, or invasive malignant. Serum samples were obtained before surgery, and HE4 levels were measured, using a cutoff of at least 70 pmol/L to indicate malignancy. Overall, subjective ultrasound evaluation by an experienced examiner had higher sensitivity and specificity than serum HE4. Sensitivity was 97% with subjective assessment ultrasound and 74% with HE4, and specificity was 90% and 85%, respectively. The additional consideration of HE4 levels after sonographers categorized a mass as benign resulted in a slight increase in sensitivity and a large increase in the number of false positives. Moreover, sequential use of serum HE4 after sonographers categorized a mass as malignant resulted in lower sensitivity and an increase in specificity.

Moszynski et al (2013) retrospectively reviewed records on 253 women with adnexal masses. Women were examined with transvaginal ultrasound by an experienced examiner before surgery. The sonographer categorized masses as certainly benign, probably benign, uncertain, probably malignant, and certainly malignant. Tumors in the certainly benign and certainly malignant categories were excluded from further analysis, and the remainder (n=145) were considered suspicious tumors. HE4 and CA 125 levels were measured in serum, and a cutoff of 65 pmol/L was used for HE4. The sensitivity and specificity of ultrasound
evaluation for diagnosing the suspicious tumors were 93.3% and 90.6%, respectively. Neither HE4 nor CA 125 improved the diagnostic accuracy for suspicious tumors. Sensitivity and specificity of HE4 were 80.0% and 91.7%, respectively, and the sensitivity and specificity of CA 125 were 85.8% and 74.7%, respectively. A logistic regression analysis confirmed that neither HE4 nor CA 125 improved the diagnostic accuracy beyond that of subjective assessment of ultrasonography.

Nikolova et al (2017) conducted a study to measure the effectiveness of HE4 compared with CA125 for differentiating ovarian endometriosis from EOC in premenopausal women. In the observational study, 164 patients were divided into three study groups-ovarian endometriosis (n=37), other benign pelvic masses (n=57), and EOCs (n=11)-and a control group (n=59). Analysis of biomarkers in blood samples from all four groups determined that HE4 performed the best at differentiating endometriosis from EOC (specificity=100%, accuracy=95.83%), while CPH-I also performed well (specificity=97.30%, accuracy=93.75%). CA125 was found to have significantly lower specificity and accuracy. Limitations of the study include the relatively small cohort.

### Table 7. Characteristics of Diagnostic Studies Evaluating HE4 as a Test for Ovarian Cancer

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Participants</th>
<th>Evaluated Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaijser et al (2014)</td>
<td>EU</td>
<td>Women with adnexal masses scheduled for surgery (n=289)</td>
<td>HE4</td>
</tr>
<tr>
<td>Moszynski et al (2013)</td>
<td>EU</td>
<td>Women with adnexal masses (n=253)</td>
<td>HE4, CA 125, individually</td>
</tr>
<tr>
<td>Nikolova et al (2017)</td>
<td>Macedonia, Serbia</td>
<td>Women with ovarian endometriosis, benign pelvic masses, EOCs (n=164)</td>
<td>HE4, CA 125, individually</td>
</tr>
</tbody>
</table>

CA 125: cancer antigen 125; EOC: epithelial ovarian cancer; HE4: human epididymis protein 4; ROMA: Risk of Ovarian Malignancy Algorithm

### Table X. Clinical Validity of HE4 as a Test for Ovarian Cancer

<table>
<thead>
<tr>
<th>Study</th>
<th>Initial N</th>
<th>Final N</th>
<th>Excluded Samples</th>
<th>Prevalence of Condition</th>
<th>Clinical Validity (95% Confidence Interval)</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaijser (2014)</td>
<td>389</td>
<td>360</td>
<td>29</td>
<td>40%</td>
<td>HE4</td>
<td>74%</td>
<td>85%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SA</td>
<td>97%</td>
<td>90%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moszynski (2013)</td>
<td>253</td>
<td></td>
<td></td>
<td>41.4%</td>
<td>HE4</td>
<td>80.0%</td>
<td>91.7%</td>
<td>87.3%</td>
<td>86.7%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SA</td>
<td>93.3%</td>
<td>90.6%</td>
<td>87.5%</td>
<td>95.1%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ROMA</td>
<td>88.9% (p=1.0)</td>
<td>89.2% (p&lt;0.001)</td>
<td>68.6%</td>
<td>96.8%</td>
</tr>
<tr>
<td>Nikolova (2017)</td>
<td>164</td>
<td></td>
<td></td>
<td></td>
<td>CA125</td>
<td>81.8% (48.2-97.7)</td>
<td>48.7% (31.9-65.6)</td>
<td>32.1% (15.9-52.4)</td>
<td>90.0% (68.3-98.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HE4</td>
<td>81.8% (48.2-97.7)</td>
<td>100% (90.5-100)</td>
<td>100% (66.4-100)</td>
<td>94.87% (82.7-99.4)</td>
</tr>
</tbody>
</table>
NPV: negative predictive value; PPV: positive predictive value; SA: subjective assessment.

Relevance and relevance design and conduct gaps are reported in Tables 8 and 9.

Table 8: Relevance Gaps

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Intervention</th>
<th>Comparator</th>
<th>Outcomes</th>
<th>Duration of Follow-Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaijser et al (2014)</td>
<td>3. 7% of data unavailable from population</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moszynski et al (2013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nikolova et al (2017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Key
1. Intended use population unclear
2. Clinical context for test is unclear
3. Study population unclear
4. Study population not representative of intended clinical use
5. Study population is subpopulation of intended use

1. Classification thresholds not defined
2. Version used unclear
3. Not version currently in clinical use

1. Classification thresholds not defined
2. Not compared to credible reference standard
3. Not compared to other tests in use for same purpose

1. Study does not directly assess a key health outcome
2. Evidence chain or decision model not explicated
3. Key clinical validity outcomes not reported (sensitivity, specificity, predictive values)
4. Reclassification of diagnostic or risk categories not reported
5. Adverse events of the test not described (excluding minor discomforts and inconvenience of venipuncture or noninvasive tests)

1. Follow-up duration not sufficient with respect to natural history of disease (TP, TN, FP, FN cannot be determined)

Table 9: Relevance Design and Conduct Gaps

<table>
<thead>
<tr>
<th>Study</th>
<th>Selectiona</th>
<th>Blindingb</th>
<th>Delivery of Testc</th>
<th>Selective Reportingd</th>
<th>Completeness of Follow-Up</th>
<th>Statisticalf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaijser et al (2014)</td>
<td>2. Selection retrospective and not randomized</td>
<td>1. Results were not blinded</td>
<td></td>
<td></td>
<td></td>
<td>1. P-values/ CI not reported</td>
</tr>
<tr>
<td>Moszynski et al (2013)</td>
<td>2. Selection retrospective and not randomized</td>
<td>1. Results were not blinded</td>
<td></td>
<td></td>
<td></td>
<td>1. P-values/ CI not reported</td>
</tr>
<tr>
<td>The et al (2018)</td>
<td>2. Selection not randomized</td>
<td>1. Results were not blinded</td>
<td></td>
<td></td>
<td></td>
<td>1. P-values reported</td>
</tr>
<tr>
<td>Nikolova et al (2017)</td>
<td>2. Selection not randomized;</td>
<td>1. Results were not blinded</td>
<td></td>
<td></td>
<td></td>
<td>1. P-values reported</td>
</tr>
</tbody>
</table>
Table 9: Study Design and Conduct Gaps

<table>
<thead>
<tr>
<th>Study</th>
<th>Selectiona</th>
<th>Blindingb</th>
<th>Delivery of Testc</th>
<th>Selective Reportingd</th>
<th>Completeness of Follow-Up*</th>
<th>Statisticalf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaijser et al (2014)</td>
<td>2. Selection retrospective and not randomized</td>
<td>1. Results were not blinded</td>
<td></td>
<td></td>
<td>1. P-values/CI not reported</td>
<td></td>
</tr>
<tr>
<td>Moszynski et al (2013)</td>
<td>2. Selection retrospective and not randomized</td>
<td>1. Results were not blinded</td>
<td></td>
<td></td>
<td>1. P-values/CI not reported</td>
<td></td>
</tr>
<tr>
<td>The et al (2018)</td>
<td>2. Selection not randomized</td>
<td>1. Results were not blinded</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nikolova et al (2017)</td>
<td>2. Selection retrospective and not randomized; small cohort</td>
<td>1. Results were not blinded</td>
<td></td>
<td></td>
<td>1. P-values/CI not reported</td>
<td></td>
</tr>
</tbody>
</table>

Key
1. Selection not described
2. Selection not random nor consecutive (ie, convenience)
1. Not blinded to results of reference or other comparator tests
1. Timing of delivery of index or reference test not described
1. Not registered
2. Evidence of selective reporting
3. Evidence of selective publication
1. Inadequate description of indeterminate and missing samples
2. High number of samples excluded
3. High loss to follow-up or missing data
4. Expertise of evaluators not described
1. Timing of index and comparator tests not same
3. Procedure for interpreting tests not described
4. Procedure for interpreting tests not described
5. Procedure for interpreting tests not described
1. Confidence intervals and/or p values not reported
2. No statistical test reported to compare to alternatives
Clinically Useful
Although HE4 levels are associated with the presence of ovarian cancer, the test does not have high sensitivity or specificity. Thus it cannot be used to rule in or rule out ovarian cancer before surgery. No prospective studies were identified that compared health outcomes in patients with adnexal masses managed with and without HE4 testing, alone or in combination with CA 125 or other disease markers. There is no strong chain of evidence demonstrating that clinical decisions based on HE4 testing would improve patient outcomes.

Section Summary: Individuals with Adnexal Masses
Multiple studies on the diagnostic accuracy of HE4 for evaluating adnexal masses have been published, and there are multiple meta-analyses of these studies. Five meta-analyses have compared the diagnostic accuracy of HE4 and CA 125. Meta-analyses found no significant difference in overall diagnostic accuracy, but 2 meta-analyses found that HE4 had higher specificity than CA 125. Findings differed in the 2 meta-analyses that compared the diagnostic accuracy of HE4 in combination with CA 125 to CA 125 alone. One of the two found that the combined test had significantly higher sensitivity and specificity than CA 125 alone. In addition, studies have not found that HE4 improves diagnostic accuracy beyond that of subjective assessment of transvaginal ultrasound. Moreover, no direct evidence from prospective studies was identified that compared health outcomes in patients with adnexal masses managed with and without HE4 testing, alone or in combination with CA 125 or other disease markers. The chain of evidence supporting the use of HE4 testing is weak. There is no strong chain of evidence demonstrating that clinical decisions based on HE4 testing would improve patient outcomes.

Biomarker Human Epididymis Protein 4 Screening for Asymptomatic individuals not at high risk of ovarian cancer

Clinical Context and Test Purpose
The purpose of screening with a serum biomarker human epididymis protein 4 levels is to provide a diagnostic option that is an alternative to or an improvement on existing testing in patients who are asymptomatic and not at high risk of ovarian cancer.

The question addressed in this evidence review is: does screening with serum HE4 levels in asymptomatic individuals not a high risk of ovarian cancer improve net health outcomes?

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

Patients
The relevant population of interest are asymptomatic individuals not at high risk of ovarian cancer.
**Interventions**
The test being considered is screening with a serum biomarker human epididymis protein 4 levels.

**Comparators**
Comparators of interest include no ovarian cancer screening.

**Outcomes**
The general outcomes of interest are overall survival, disease-specific survival, test accuracy, test validity, and other test performance measures.

**Timing**
Though not completely standardized, follow-up for patients who are asymptomatic and not at high risk of ovarian cancer would typically occur in the years before diagnosis.

**Setting**
Patients who are asymptomatic and not at high risk of ovarian cancer are managed by primary care providers in an outpatient clinical setting.

**Study Selection Criteria**
For the evaluation of clinical validity of the tests, studies that meet the following eligibility criteria were considered:

- Reported on the accuracy of the marketed version of the technology (including any algorithms used to calculate scores)
- Included a suitable reference standard
- Patient/sample clinical characteristics were described
- Patient/sample selection criteria were described
- Included a validation cohort separate from development cohort.

**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**
No published studies identified compared the diagnostic performance of the HE4 biomarker for screening asymptomatic women for ovarian cancer with a reference standard. In addition, no studies prospectively compared health outcomes in asymptomatic women managed with and without HE4 screening.
Several retrospective studies aimed at determining the potential value of using HE4 and other biomarkers in early identification of ovarian cancer in asymptomatic women. In 2010 Anderson et al published data on 34 women with ovarian cancer and 70 matched controls, all of whom were participating in an unrelated randomized controlled trial on smokers at increased risk of lung cancer. Blood samples were available for the women between 0 years and 18 years before ovarian cancer diagnosis. In descriptive analyses, individual serum markers, including HE4, CA 125, and mesothelin, showed increasing accuracy over time approaching the diagnosis of ovarian cancer. Mean concentrations of these markers, which were measured by visually read immunoassays, began to increase approximately 3 years before diagnosis but attained detectable levels only within the final year before diagnosis. The study had a small sample size, limiting the ability to conduct quantitative analysis, and included only heavy smokers and therefore may not be representative of the population of women at risk of ovarian cancer.

In 2011, Urban et al retrospectively reviewed preclinical serum samples to evaluate the potential utility of HE4 and other markers as a secondary screening test in women found to have epithelial ovarian cancer. There were samples from 112 ovarian cancer patients and 706 matched controls. Individuals participated in the Prostate, Lung, Colorectal, and Ovarian (PLCO) trial and had been screened annually for 6 years with CA 125. Serum samples to evaluate potential markers were taken from the year proximate to that in which women were diagnosed with ovarian cancer. (Serum samples were not available for the fourth screen, so they were taken from the third year for the women diagnosed with ovarian cancer between the third and fourth screens.) Investigators evaluated the associations between CA 125, HE4, and levels of 5 other markers with malignancy, accounting for increasing CA 125 levels and adjusting for demographic characteristics. Increase in CA 125 levels was associated with statistically significant increases in all of the markers. Levels of HE4 were most elevated, compared with controls, ie, the highest average HE4 level was 4.26 standard deviations above the mean HE4 level in control samples.

In 2016, Terry et al retrospectively analyzed prospectively collected data from the European Prospective Investigation into Cancer and Nutrition (EPIC) study, a multicenter cohort study investigating the relationship between diet and cancer. The analysis used a nested case-control design. A total of 197 women who developed invasive ovarian cancer were matched with 725 randomly selected ovarian cancer-free controls. Baseline and follow-up blood samples were analyzed for levels of several biomarkers (ie, CA 125, HE4, cancer antigen 15.3, cancer antigen 72.4) and the sensitivity, specificity, and AUROC curve were calculated. CA 125 was best able to discriminate between cases and controls within 6 months of ovarian cancer diagnosis (C statistic, 0.92), followed by HE4 (C=0.84). The ability of the markers to discriminate between cases and controls decreased with longer intervals between blood draws and cancer diagnosis. For example, with a 1- to 2-year time lag, C statistic values were 0.72 for CA 125 and 0.65 for HE4; for a 3- to 6-year time lag, the C statistic was 0.55 for CA 125. (Data on HE4 were not available for the 3- to 6-year timelag analysis.)
Clinically Useful
No RCTs or nonrandomized comparative studies evaluating the clinical utility of screening asymptomatic women with HE4 were identified. The studies have not estimated the sensitivity and specificity of HE4 in the screening setting, and thus the chain of evidence supporting screening is incomplete.

Section Summary: Screening Asymptomatic Individuals Not at High Risk of Ovarian Cancer
There is insufficient evidence from prospective or controlled studies that HE4 is an effective screening tool for identifying ovarian cancer in asymptomatic women. The utility of HE4 as a biomarker to screen for ovarian cancer along with CA 125 needs to be further evaluated in prospective studies and confirmed in RCTs that evaluate the impact of screening on health outcomes.

Summary of Evidence
For individuals who have ovarian cancer who receive a measurement of serum biomarker HE4, the evidence includes a prospective study and several retrospective studies comparing the diagnostic accuracy of HE4 with CA 125 for predicting disease progression and/or recurrence. Relevant outcomes are overall survival, disease-specific survival, test validity, other test performance measures, and change in disease status. Data submitted to the U.S. Food and Drug Administration for approval of commercial HE4 tests found that HE4 was not inferior to CA 125 for detecting ovarian cancer recurrence. However, the superiority of HE4 to CA 125 (alone or in combination), the key question in the evidence review, was not demonstrated in the available literature. In addition, there is no established cutoff in HE4 levels for monitoring disease progression, and cutoffs in studies varied. There is no direct evidence from prospective controlled studies on the impact of HE4 testing on health outcomes, and no clear chain of evidence that changes in management based on HE4 would lead to an improved health outcome. The evidence is insufficient to determine the effects of the technology on health outcomes.

For individuals who have adnexal masses who receive a measurement of serum biomarker HE4, the evidence includes diagnostic accuracy studies and meta-analyses. Relevant outcomes are overall survival, disease-specific survival, test validity, and other test performance measures. Meta-analyses have generally found that HE4 and CA 125 have a similar overall diagnostic accuracy (ie, sensitivity, specificity) and several found that HE4 has significantly higher specificity than CA 125 but not sensitivity. Two meta-analyses had mixed findings on whether the combination of HE4 and CA 125 is superior to CA 125 alone for the initial diagnosis of ovarian cancer. The number of studies evaluating the combined test is relatively low, and publication bias in studies of HE4 has been identified. In addition, studies have not found that HE4 improves diagnostic accuracy beyond that of subjective assessment of transvaginal ultrasound. There is no direct evidence from prospective controlled studies on the impact of HE4 testing on health outcomes, and no clear chain of evidence that changes in management
based on HE4 would lead to an improved health outcome. The evidence is insufficient to determine the effects of the technology on health outcomes.

For individuals who are asymptomatic and not at high risk of ovarian cancer who receive screening with a serum biomarker HE4 test, the evidence includes several retrospective comparative studies and no prospective studies comparing health outcomes in asymptomatic women managed with and without HE4 screening. Relevant outcomes are overall survival, disease-specific survival, test validity, and other test performance measures. The retrospective studies found that HE4 levels increased over time in women ultimately diagnosed with ovarian cancer. Prospective comparative studies are needed to determine definitively whether HE4 testing is a useful screening tool. The evidence is insufficient to determine the effects of the technology on health outcomes.

SUPPLEMENTAL INFORMATION

Practice Guidelines and Position Statements

National Comprehensive Cancer Network
National Comprehensive Cancer Network (NCCN) ovarian cancer guidelines (v.2.2018) state that, for monitoring and follow-up of patients with stage I to IV ovarian cancer with a complete response to initial treatment, “CA-125 [cancer antigen 125] or other tumor marker” should be used at “every visit if initially elevated”. The guidelines do not specify any marker other than CA 125 for monitoring patients after treatment.

NCCN guidelines state the following on evaluating undiagnosed pelvic masses: “The FDA has approved the use of HE4 [human epididymis protein 4] and CA-125 for estimating the risk for 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.”

NCCN guidelines state the following on screening for ovarian cancer:

The literature does not support routine screening for ovarian cancer in the general population, and routine screening is not currently recommended by any professional society. Some physicians follow women with high-risk factors (eg, those with BRCA mutations, those with a family history) using cancer antigen 125 (CA-125) monitoring and endovaginal ultrasound; however, prospective validation of these tests remains elusive.

National Institute for Health and Care Excellence
The National Institute for Health and Care Excellence (NICE) issued guidance in 2011 on the detection and initial management of ovarian cancer. The guidance included the following recommendations:

- “Measure serum CA125 in primary care in women with symptoms that suggest ovarian cancer.
- If serum CA125 is 35 IU/ml or greater, arrange an ultrasound scan of the abdomen and pelvis.
- If the ultrasound suggests ovarian cancer, refer the woman urgently for further investigation.
- For any woman who has normal serum CA125 (less than 35 IU/ml), or CA125 of 35 IU/ml or greater but a normal ultrasound:
  - assess her carefully for other clinical causes of her symptoms and investigate if appropriate
  - if no other clinical cause is apparent, advise to return to her GP if her symptoms become more frequent and/or persistent.

Malignancy indices

- Calculate a risk of malignancy index I (RMI I) score (after performing an ultrasound...). [The RMI I combines CA 125, menopausal status and the ultrasound score].”

The guidance did not mention HE4.

**U.S. Preventive Services Task Force Recommendations**

The U.S. Preventive Services Task Force updated its recommendations for screening for ovarian cancer in February 2018.26. The Task Force recommended against screening for ovarian cancer in asymptomatic women (D recommendation). HE4 was not specifically discussed.

**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**

Some currently unpublished trials that might influence this review are listed in Table 3.

**Table 10. Summary of Key Trials**

<table>
<thead>
<tr>
<th>NCT No.</th>
<th>Trial Name</th>
<th>Planned Enrollment</th>
<th>Completion Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ongoing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCT01121640</td>
<td>A Randomized Controlled Trial Using Novel Markers to Predict Malignancy in Elevated-Risk Women</td>
<td>1208</td>
<td>Jun 2018</td>
</tr>
<tr>
<td>NCT02687321</td>
<td>The Role of HE4 in the Follow-up of Advanced Ovarian, Fallopian Tube and Primary Peritoneal Cancer</td>
<td>150</td>
<td>Jan 2020</td>
</tr>
</tbody>
</table>

NCT: national clinical trial.

**REFERENCES**


Billing Coding/Physician Documentation Information

86305 Human epididymis protein 4 (HE4)

If HE4 is performed as part of the Risk of Ovarian Malignancy Algorithm (ROMA™) test, the following code would be reported:

81500 Oncology (ovarian), biochemical assays of 2 proteins (CA-125 and HE4), utilizing serum, with menopausal status, algorithm reported as a risk score

Additional Policy Key Words
N/A

Policy Implementation/Update Information

5/1/15 New Policy, considered investigational.
5/1/16 No policy statement changes.
5/1/17 No policy statement changes.
5/1/18 No policy statement changes.
5/1/19 No policy statement changes.

State and Federal mandates and health plan contract language, including specific provisions/exclusions, take precedence over Medical Policy and must be considered first in determining eligibility for coverage. The medical policies contained herein are for informational purposes. The medical policies do not constitute medical advice or medical care. Treating health care providers are independent contractors and are neither employees nor agents Blue KC and are solely responsible for diagnosis, treatment and medical advice. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, photocopying, or otherwise, without permission from Blue KC.