EDUCATIONAL COMMENTARY - HE4 AND CA125—BIOMARKERS FOR OVARIAN CANCER

Educational commentary is provided through our affiliation with the American Society for Clinical Pathology (ASCP). To obtain FREE CME/CMLE credits click on Earn CE Credits under Continuing Education on the left side of the screen.

Learning Objectives

On completion of this exercise, the participant should be able to

- characterize the utility of diagnostic sensitivity/specificity, positive/negative predictive values, and receiver operating characteristic (ROC) curves on the choice of one biomarker over another;
- compare and contrast HE4, CA125, and the RMI/ROCA/ROMA/ICRA algorithms as tumor biomarkers for ovarian cancer; and
- discuss the reasons to use individual tumor markers or multimarker algorithms for monitoring treatment success, rather than for diagnosis.

The Challenge of Using Tumor Markers: Statistical Tools

The National Cancer Institute defines a tumor marker as “a substance found in tissue, blood, or other body fluids that may be a sign of cancer or certain benign (noncancerous) conditions.” This statement provides both the major advantage and the major disadvantage of such substances: they are made by both normal and cancer cells, and can be found in elevated amounts in both benign and malignant conditions. Currently, no tumor marker is accepted as being so specific that it points to only one type of cancer. This reality can be demonstrated with an initial exploration of the most critical terms associated with such medical decisions: diagnostic sensitivity, diagnostic specificity, positive predictive value, and negative predictive value.

When used in a selected population, the vast majority of clinical laboratory tests will demonstrate a bimodal pattern, with one curve showing the test results for “healthy” individuals and the other curve showing those with the disease in question. It is common for those two curves to intersect (Figure 1A). The crucial aspect here is the placement of the cutoff value, which will distinguish the true-positives (TP, persons with the disease) from the true-negatives (without the disease, TN). Inevitably, however, the placement of the cutoff value will result in some healthy individuals being called “positive” (false-positive, FP), and some with disease being called “negative” (false-negative, FN). Moving that cutoff point to the left (Figure 1B), to maximize the number of true-positive results identified by the test, automatically increases the number of false-positives; accordingly, moving the cutoff to the right (Figure 1C) will increase the number of false-negatives.

Ideally, the closer that a laboratory test can get to correctly identifying all those with disease as positive and all those without disease as negative, the more diagnostically useful that test will be. This can be determined mathematically by setting a test cutoff at a certain level, and then counting the number of the
Tested population who fit into TP, FP, TN, and FN. The diagnostic sensitivity of a test is the percentage of those who have the disease (TP+FN) who test positive (TP). The diagnostic specificity of a test is the percentage of those who do not have the disease (TN+FP) who test negative (TN). Viewed another way, the positive predictive value (PPV) of a test is the percentage of all those with a positive test result (TP+FP) who actually have the disease (TP), and the negative predictive value (NPV) of the test is the percentage of those with a negative result (TN+FN) who actually do not have the disease (TN).

Diagnostic sensitivity and specificity are the criteria used for choosing whether to offer a screening test (“How much am I willing to risk missing someone who actually has the disease by using this test?”), while PPV/NPV become the diagnostic tools when obtaining the test result before knowing the final diagnosis and then asking “How reliably will this result predict that this patient truly has or truly does not have the condition?” Another critical consideration in interpreting these calculations is prevalence, the number of cases of a disease that are present in a particular population at a given time. Prevalence is sometimes confused with incidence, which “refers to the number of new cases that develop in a given period of time.” Assuming all other factors remain constant, the PPV increases with increasing prevalence, and NPV decreases with increase in prevalence.

Positive predictive value is an important measure of the diagnostic utility of a laboratory test, particularly for screening tests for cancer. Getting a cancer diagnosis, whether true (TP) or not true (FP), subjects patients to either end of the same emotional, financial, and follow-up spectrum. In addition, it is well known that the earlier a diagnosis is made, the greater the possibility for successful treatment. The quest for a screening laboratory test that will accurately predict, in a general female population (with perhaps a small prevalence of cancer), and as early as possible, the individuals with the greatest probability of ovarian cancer, is still the impossible dream as of this writing. Using that test in a population with the greatest historical risk for ovarian cancer, and therefore an increased prevalence, is the focus of many of the research efforts for a new tumor biomarker for this cancer.
ROC Plots

The statistics above can be visualized in a tool that was originally used in World War II to establish the bias of radar receiver operators in distinguishing between enemy aircraft and noise. These receiver operating characteristic (ROC) plots were introduced into medical decision making as early as the 1950s, and soon were found useful in many areas, including psychology and medical imaging. Lasko et al. summarized the benefits and limitations of using ROC plots in biomedical informatics, including the importance of setting up the initial comparative experiment correctly, as an "inappropriate approach can easily lead to incorrect conclusions" in interpreting ROC curves. These curves display the data obtained when multiple sequential cutoff test values, as illustrated in Figure 2, determine the proportion of the tested population who will be TP and FP. In one example using two biomarkers for ovarian cancer (HE4 and CA125), the diagonal line represents a test that would provide no diagnostic information, whereas the closer the test lines are to the upper left corner (high TP, low FP), the more "perfect" the test is. Both tests, according to these data, are diagnostically useful, with HE4 showing greater differentiation of TP from FP. A calculation associated with ROC curves is the area under the curve (AUC). The closer this value is to 1.00, the better the test. In this example, HE4 was the better diagnostic test, with an AUC of 0.96 compared with CA125’s AUC of 0.82. Such statistical tools have become common and expected practice when comparing a new test, like a biomarker, with a “gold standard.”

<table>
<thead>
<tr>
<th>Marker</th>
<th>AUC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE4</td>
<td>0.96 (0.90-1)</td>
</tr>
<tr>
<td>CA125</td>
<td>0.82 (0.70-0.94)</td>
</tr>
<tr>
<td>p</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Figure 2. ROC plot and AUC for CA125 and HE4 for patients with epithelial ovarian cancer and benign diseases. Abbreviations: CI, Confidence Interval; p, statistical probability. Reprinted with permission from BioMed Central Ltd.
EDUCATIONAL COMMENTARY - HE4 AND CA125—BIOMARKERS FOR OVARIAN CANCER (cont.)

Ovarian Cancer Statistics

The American Cancer Society estimates that in 2015, approximately 21,290 women will have been newly diagnosed with ovarian cancer (OC), and 14,180 will have died of the condition. Ovarian cancer ranks fifth in cancer deaths in females; the risk that a woman will have OC in her lifetime, or will die of it, are 1 in 75 and 1 in 100, respectively. Risk factors for developing OC include age (63+ years old), obesity, never pregnant or first pregnancy after 35, taking the fertility drug clomiphene citrate for longer than 1 year, and a family or personal history of OC, breast cancer, or colorectal cancer.

There are multiple OC subtypes: serous, mixed epithelial, clear cell, stromal, endometrioid, mucinous, transitional cell, germ cell, adenocarcinoma, and undifferentiated. The relative five-year survival rate (compared to people without OC) for all types of OC is 45%, with younger women faring better than those older than 65 years. For those whose cancer is localized to the ovary (stages IA and IB), 92% will survive 5 years, but 85% of all those with OC will already have experienced metastasis before diagnosis is made.

The Search for the Best Tumor Marker for Ovarian Cancer: CA125

In 2012, Ozer et al. used immunohistochemistry to discover markers on 68 ovarian tumors of various subtypes. They found that antibodies to p53, p21, bax, c-kit, and metallothionein could be used for typing ovarian tumors, but bcl-2 and telomerase staining could not. Histologic staining, however, requires surgical removal of an established tumor and therefore is a poor candidate for an early tumor screening test.

The first recognized OC biomarker in the blood was CA125 (cancer antigen, carbohydrate antigen). First described by Bast et al. in 1981, CA125 is a glycoprotein that is normally expressed by the ovary, cervix, endometrium, fallopian tube, pleura, pericardium, and peritoneum, as well as epithelial tissues of the colon, pancreas, lung, kidney, prostate, breast, stomach, and gallbladder. This widespread presence leads to the major limitation of CA125 as an OC tumor marker. It can be increased in other tumors (pancreas, breast, colon, lung) as well as in non-gynecologic disease (inflammatory peritoneum/pleura/pericardium or pancreatitis, hepatitis, cirrhosis, and tuberculosis). Studies by Bast et al. in 1983 on 888 healthy men and women led to the cutoff level of 35 U/mL (35 kU/L) being widely adopted as the upper boundary of the reference range for serum CA125. It is acknowledged that CA125 levels can vary by menstrual cycle, pregnancy, hysterectomy, and a number of benign conditions. Bon et al. reported in 1996 that healthy postmenopausal women tend to have values less than 20 kU/L.

One additional limitation of this test is that patients with OC may not have an increased CA125 level. Recognizing the limited single-test use of CA125 as a tumor marker, investigators have employed a
multimarker algorithm to increase testing sensitivity. In 1990, Jacobs et al. described the use of age, ultrasound score, menopausal status, and CA125 values to develop a risk of malignancy index (RMI) for distinguishing among 142 patients with benign (101) and malignant (42) masses. The sensitivity of the RMI was 85% and the specificity was 97%, making this a better algorithm for ruling out (TN) than for ruling in (TP) cancer for those patients.

In 2012 through 2014, four meta-analyses of the literature were performed to determine the diagnostic utility of using CA125 for diagnosing OC (Table I), while noting this is not an FDA-approved use for the test. As shown below, these CA125 statistics are the biomarker gold standard against which other biomarkers are compared. The summaries also point to the role possibly played by differences in populations tested and cutoffs utilized for differentiating TP from FP.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Number of Studies</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang et al. (2014)</td>
<td>23</td>
<td>0.79 (0.74-0.84)</td>
<td>0.82 (0.77-0.87)</td>
</tr>
<tr>
<td>Zhen et al. (2014)</td>
<td>25</td>
<td>0.74 (0.72-0.76)</td>
<td>0.83 (0.81-0.84)</td>
</tr>
<tr>
<td>Ferraro et al. (2013)</td>
<td>13</td>
<td>0.79 (0.77-0.82)</td>
<td>0.78 (0.76-0.80)</td>
</tr>
<tr>
<td>Yu et al. (2012)</td>
<td>10</td>
<td>0.66 (0.62-0.70)</td>
<td>0.87 (0.85-0.89)</td>
</tr>
</tbody>
</table>

* Summarized from Serum Biomarker Human Epididymis Protein 4 (HE4) Policy #: 445.27

The National Comprehensive Cancer Network 2015 policy states that CA125 is “not used alone to diagnose ovarian cancer… It may also be done during and after treatments to check treatment results.” It is interesting, however, to find evidence that some physicians consider the CA125 test an effective screening method for women at low risk for OC. Baldwin et al. used a cross-sectional survey of 1088 physicians providing women’s primary care, and 33% believed that OC screening was effective, and routinely offered/ordered it, despite evidence to the contrary.

CA125 has been combined with multiple other tests in an effort to develop the best algorithm to show its utility for diagnosis. In 2009, the FDA approved the use of the OVA1 test, which uses blood levels of CA125, transthyretin, apolipoprotein A1, β2 microglobulin, and transferrin to calculate a risk factor from 1 to 10. The Risk of Ovarian Cancer Algorithm (ROCA), developed by the MD Anderson Cancer Center, uses CA125 testing performed sequentially over time (with follow-up ultrasound if needed) to place women into one of three OC risk categories: low, intermediate, and high. Studies done by Menon et al. over 11 years with 4051 women showed 99.9% specificity and 60% PPV, with no cases of invasive OC missed. A British study used ROCA with 46,237 women, 50 years or older, from June 2002 to July 2014. The ROC curve from these studies showed that the ROCA score was more useful diagnostically than the individual CA125 levels alone.
The Search for the Best Tumor Marker for Ovarian Cancer: HE4

In 2008, the FDA approved a second biomarker for use in monitoring progression or recurrence of OC, but not for diagnosis or to assess risk of disease outcomes. First described in 1991 by Kirchhoff et al., using cDNA cloning studies, human epididymis gene product 4 (HE4) is similar to CA125 in not being specific for OC, since it can also be elevated in benign conditions as well as in breast, pancreatic, and endometrial cancers.

A summary of meta-analyses of studies using HE4 alone or with CA125 is presented in Table II. When combined with the meta-analyses of CA125 alone (Table I), the pooled data of the 12 studies suggest that HE4, whether alone or in combination with CA125, did not improve the overall specificity or sensitivity for diagnosis. While four of six analyses (bold in Table II) showed that HE4 alone had higher specificity than CA125 and one (Yu et al.) showed higher sensitivity, the evidence was deemed insufficient to use HE4 alone or in combination with CA125 to achieve better diagnostic performance than CA125 alone.
Table II. Meta-Analyses of Studies on HE4 and the Combination of HE4 and CA125 for Diagnosing Ovarian Cancer. a

<table>
<thead>
<tr>
<th>Authors</th>
<th>Number of Studies</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE4 Alone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maceda et al. (2014)</td>
<td>36</td>
<td>0.78 (0.77 - 0.79)</td>
<td>0.86 (0.85 - 0.87)</td>
</tr>
<tr>
<td>Yang et al. (2013)</td>
<td>31</td>
<td>0.73 (0.71 - 0.75)</td>
<td>0.89 (0.88 - 0.90)</td>
</tr>
<tr>
<td>Wang et al. (2014)</td>
<td>28</td>
<td>0.76 (0.72 - 0.80)</td>
<td>0.93 (0.90 - 0.96)</td>
</tr>
<tr>
<td>Zhen et al. (2014)</td>
<td>25</td>
<td>0.74 (0.72 - 0.76)</td>
<td>0.90 (0.89 - 0.91)</td>
</tr>
<tr>
<td>Ferraro et al. (2013)</td>
<td>14</td>
<td>0.79 (0.76 - 0.81)</td>
<td>0.93 (0.92 - 0.94)</td>
</tr>
<tr>
<td>Yu et al. (2012)</td>
<td>12</td>
<td>0.80 (0.77 - 0.83)</td>
<td>0.92 (0.90 - 0.93)</td>
</tr>
<tr>
<td>Combination HE4 and CA125</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhen et al. (2014)</td>
<td>9</td>
<td>0.90 (0.87 - 0.92)</td>
<td>0.85 (0.82 - 0.87)</td>
</tr>
<tr>
<td>Ferraro et al. (2013)</td>
<td>4</td>
<td>0.82 (0.78 - 0.86)</td>
<td>0.76 (0.72 - 0.80)</td>
</tr>
</tbody>
</table>

a Summarized from Serum Biomarker Human Epididymis Protein 4 (HE4) Policy #: 445.27 See text for explanation of bold type.

However, additional algorithms using both HE4 and CA125 have been developed that show promise for diagnostic purposes. In September 2011, the FDA approved the Risk of Ovarian Malignancy Algorithm (ROMA) to determine if an ovarian mass, in premenopausal or postmenopausal women older than 18 years, places the patient in low or high risk for surgical discovery of malignancy. It was noted that the ROMA “is still not intended as a screening or stand-alone diagnostic assay.” Wang et al. did a meta-analysis in 2014 on 32 studies, which compared CA125, HE4, and ROMA for diagnosing OC. ROC plots for the three tests showed similar AUC for each (HE4, 0.89; CA125, 0.87; and ROMA, 0.91), indicating a similar stand-alone test discriminatory ability in the studies. However, HE4 showed better overall specificity (HE4, 93.6%; CA125, 82.1%; ROMA, 82.4%), especially in the premenopausal group (HE4, 93.8%; CA125, 76.3%; ROMA, 85.1%). For postmenopausal women the opposite was true: when comparing premenopausal with postmenopausal data CA125 and ROMA had better sensitivity (AUC: CA125-pre, 0.85 vs. post, 0.92; ROMA-pre, 0.86 vs. post 0.93). The authors concluded that HE4 may be better for ruling out OC in premenopausal women, whereas CA125 and ROMA may be more informative for diagnosing those who are postmenopausal. In any event, there is a need for more high-quality randomized clinical trials to confirm all of these impressions.

One additional algorithm was proposed by Moore et al. in 2014, which incorporated ROMA after an initial cancer risk assessment (ICRA) by a generalist clinician to classify patients as low-likelihood or high-likelihood for having cancer. This study enrolled 461 women with benign tumors (375), epithelial OC (48), low-malignancy-potential tumors (18), and non-ovarian malignancies (20). The addition of ROMA to the initial clinical impression for OC increased sensitivity from 85.4% to 93.8% and decreased specificity from 84.3% to 67.2%, but increased the NPV from 97.8% to 98.8%. For all cancers, ROMA improved sensitivity of ICRA from 77.9% to 89.7%, identifying 13 additional cancers missed by ICRA alone.
EDUCATIONAL COMMENTARY - HE4 AND CA125—BIOMARKERS FOR OVARIAN CANCER (cont.)

Summary

The challenges that remain in early OC diagnosis include (1) the wide variety of cancer subtypes, each of which can have its own unique fingerprint of cell surface markers, affected intracellular signaling pathways, and levels of proteins expressed into body fluids; (2) the variability of biomarker blood levels, even in healthy women, based on age and menstrual status; (3) the small numbers of patients included in most clinical trials, reducing the power of obtaining statistics that are interpretable as “significant”; (4) the lack of analytic specificity of any biomarker discovered to date for only OC; (5) the lack of sufficient studies to establish a cutoff for HE4 and ROMA to differentiate “cancer-free” from “cancer”; and (6) that prevalence is not always known before interpreting any PPV/NPV/ROC data. The utility of algorithms involving multiple assays seems to have potential for increasing diagnostic usefulness.

References


10. What are the key statistics about ovarian cancer? American Cancer Society website. 

11. What are the risk factors for ovarian cancer? American Cancer Society website. 


15. How is ovarian cancer diagnosed? American Cancer Society website. 


© ASCP 2016

Use the following table to answer questions 1 and 2 in the exercise that accompanies this commentary:

<table>
<thead>
<tr>
<th></th>
<th>No Cancer</th>
<th>Cancer</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test positive</td>
<td>94</td>
<td>40</td>
<td>134</td>
</tr>
<tr>
<td>Test negative</td>
<td>400</td>
<td>3</td>
<td>403</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>494</strong></td>
<td><strong>43</strong></td>
<td><strong>537</strong></td>
</tr>
</tbody>
</table>