Diagnostic Value of Human Epididymis Protein 4 Compared with Mesothelin for Ovarian Cancer: a Systematic Review and Meta-analysis

Jia-Ying Lin¹&, Jin-Bao Qin²&, Xiao-Yan Li³, Ping Dong¹, Bing-De Yin⁴*

Abstract

Background and Purpose: Ovarian cancer is the leading cause of death among gynecologic cancers because of the lack of effective early detection methods. Accuracies of the human epididymis protein 4 (HE4) and mesothelin in detecting ovarian cancer have never been systematically assessed. The current systematic review aimed to tackle this issue. Methods: MEDLINE, EMBASE, and Cochrane databases were searched (September 1995–November 2011) for studies on the diagnostic performances of HE4 and mesothelin in differentiating ovarian cancer from other benign gynecologic diseases. QUADAS items were used to evaluate the qualities of the studies. Meta-DiSc software was used to handle data from the included studies and to examine heterogeneity. All included studies for diagnostic performance were combined with sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratios (DORs) with 95% confidence intervals (CIs), summary receiver operating characteristic (SROC) curves, and areas under the SROC curves (AUC). Results: A total of 18 studies and 3,865 patients were eligible for the final analysis. The pooled sensitivity estimates for HE4 (74.4%) were significantly higher than those for mesothelin (49.3%). The pooled specificity estimates for mesothelin (94.5%) were higher than those for HE4 (85.8%). The pooled DOR estimates for HE4 (26.22) were higher than those for mesothelin (24.01). The SROC curve for HE4 showed better diagnostic accuracy than that for mesothelin. The PLR and NLR of HE4 were 6.33 (95% CI: 3.58 to 11.18) and 0.27 (95% CI: 0.21 to 0.34), respectively. The PLR and NLR for mesothelin were 11.0 (95% CI: 6.21 to 19.59) and 0.51 (95% CI: 0.42 to 0.62), respectively. The combination of the two tumor markers or their combination with CA-125 increased sensitivity and specificity to different extents. Conclusion: The diagnostic accuracy of HE4 in differentiating ovarian cancer from other benign gynecologic diseases is better than that of soluble mesothelin-related protein. Combinations of two or more tumor markers show more sensitivity and specificity.

Keywords: Ovarian carcinoma - HE4 - mesothelin - meta-analysis

Asian Pacific J Cancer Prev, 13 (11), 5427-5432

Introduction

Epithelial ovarian cancer (EOC), along with primary peritoneal cancer, is the fourth leading cause of cancer death among women, with approximately 15,000 deaths annually in the United States. The poor prognosis of ovarian cancer is primarily due to the fact that the cancer is not detected in a majority of patients until it has spread beyond the ovary (Williams et al., 2007). The detection of a larger fraction of ovarian cancers during the early stages might significantly improve the overall survival rate. Over the last two decades, attempts to develop effective screening methods have focused on ultrasonography and serum markers. Given the low prevalence of ovarian cancer in the general population, an effective strategy must have a high sensitivity (> 75%) during the early stages of the disease and a very high specificity (> 99.6%) to attain a positive predictive value (PPV) of 10%. A PPV of 10% equates to a situation in which only 1 out of 10 surgical interventions leads to the diagnosis of ovarian cancer (Bast, 2004).

To date, the glycoprotein CA-125 is the most thoroughly investigated biomarker for ovarian cancer screening. CA-125 has been used as a tool to differentiate between benign and malignant ovarian masses. CA-125 is not sufficiently sensitive for screening the general population when used as an individual marker. In most studies, CA-125 is elevated in approximately 50% to 60% of stage I diseases at the time of conventional diagnosis (Jacobs & Bast, 1989). In addition, the specificity of CA-125 is compromised because many benign gynecologic and medical conditions and other malignancies can result

¹Department of Otolaryngology Head and Neck Surgery, ²Department of Urology, Shanghai First People’s Hospital, School of Medicine, Shanghai Jiao Tong University, ³Department of Vascular Surgery, the Ninth People’s Hospital, the Affiliated to Shanghai Jiaotong University, School of Medicine, ¹Department of Otolaryngology-Head and Neck Surgery, Shanghai Children’s Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China ⁴Equal contributors *For correspondence: 13818761532@126.com
in elevated serum CA-125 levels. Therefore, new serum markers need to be sought to replace or complement CA-125 to detect a greater fraction of ovarian cancers during the early stages of the disease.

Mesothelin or soluble mesothelin-related protein (SMRP) and HE4 are two of the most promising novel ovarian cancer biomarkers that are currently under evaluation. Mesothelin and HE4 are commonly overexpressed in ovarian cancer tissues and are elevated in the sera of patients with ovarian cancer. SMRP and HE4 have many beneficial characteristics that set them apart from other potential markers currently being studied. The preeminent distinctiveness of SMRP and HE4 can be attributed to their temporal stability (Salceda et al., 2005; Moore et al., 2008), which may assist in the early diagnosis of high-risk patients.

Previous studies on the roles of SMRP and HE4 markers in the differential diagnosis of ovarian cancer from benign gynecologic diseases present conflicting results. Therefore, a comprehensive system review would be useful to synthesize the available information. This study aims to assess the diagnostic performances of SMRP and HE4 in differentiating between ovarian cancer and benign gynecologic diseases using standard meta-analysis techniques.

Materials and Methods

Literature search

A systematic review of original articles analyzing the diagnostic performance of HE4 and SMRP was performed by searching MEDLINE, EMBASE, and Cochrane Databases. Original and review articles published from September 1995 to November 2011 were sought; The search terms were “HE4/ WFDC2”, “mesothelin/SMRP/ MLSN”, “ovarian carcinoma/ovarian cancer/carcinoma of ovary,” sensitivity/specificity/false negative/false positive/diagnosis/detection/accuracy”. We evaluated all associated publications to retrieve the most eligible studies. Moreover, their reference lists were searched manually to find other relevant publications. Both original and review articles were sought because the latter were considered as additional sources of unaccounted original works.

Selection of studies

The eligibility criteria for the meta-analysis of the studies included the following: (1) both the sensitivity and specificity of levels of HE4 and SMRP for the diagnosis of ovarian cancer were provided, or HE4 and SMRP values were provided in a scatter plot form, allowing test results to be extracted for each individual. (2) 50 or more patients were included. (3) The study design included women with ovarian cancer and benign gynecologic diseases, and evaluated the contribution of HE4 and SMRP.

Articles were excluded when data were insufficient to construct a 2x2 table of the test result (serum HE4 and SMRP concentration). The 2x2 tables were constructed independently by two of the authors (JY, L and JB, Q). In the event of disagreement, the judgment of a third author (BD, Y) was decisive.

Data extraction

The final set of English articles was assessed independently by two observers (JY, L and JB, Q). The observers were blinded to publication details, and differences between them were resolved by a consensus. Data retrieved from the reports included the name of the author, publication year, participant characteristics, test method, cut-off value, sensitivity and specificity, and study quality score.

We assessed the quality of the included studies by the criteria selected from the Quality Assessment for Studies of Diagnostic Accuracy checklist for the assessment of diagnostic studies (Whiting et al., 2003): study design (prospective or retrospective); patient selection (consecutive or not); blinding (blind or not to the interpretation of index test); assay method; study size, etc. The numbers of true-positive (TP), false-negative (FN), false-positive (FP) and true-negative (TN) results in the detection of ovarian cancer were extracted on a per-patient or per-lesion basis.

Statistical analysis

The sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), and diagnostic odds ratio (DOR) for each study were calculated, and the results were pooled as per the DerSimonian Liard random effects model (DerSimonian & Laird, 1986). Heterogeneity was assessed using the likelihood ratio $I^2$ index and $X^2$ test. $I^2$ index is a measure of the percentage of total variation across studies due to heterogeneity beyond chance. Values over 50% indicate heterogeneity (Huedo-Medina et al., 2006). In the likelihood ratio $X^2$ test, $p < 0.05$ was considered to indicate apparent heterogeneity. If heterogeneity existed (Dinnes et al., 2005) a random effects model was used in the primary meta-analysis to obtain a summary estimate for sensitivity with 95% confidence interval (CI). The diagnostic performance of the test under study per unit increase in the covariate was used as the accuracy measure. The current work used Moses’ linear model to draw a summary receiver operating characteristic (SROC) curve, which summarized the joint distribution of sensitivity and specificity. The area under the SROC curve (AUC) was calculated. Publication biases were assessed using funnel plots.

Results

Literature search and study design characteristics

The current research yielded 186 primary studies, of which 156 were excluded after reviewing the title and abstract, and 12 after reviewing the full article (Figure 1). Five articles were excluded because only healthy women were included in the control group. Six articles were excluded because their data were insufficient to calculate the true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) values.

One article was excluded because its author published two reports on the same patients and considered only the study with the best quality. A total of 18 studies (Scholler et al., 1999; Hellstrom et al., 2003; McIntosh et al., 2004; Hassan et al., 2006; Badgwell et al., 2007; Moore et al.,
Table 1. Main Characteristics of 18 Studies

<table>
<thead>
<tr>
<th>Author</th>
<th>Location</th>
<th>Temperature(°C)</th>
<th>Number of patients’ serum</th>
<th>Design</th>
<th>Blind</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scholler</td>
<td>USA</td>
<td>4</td>
<td>124</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>McIntosh</td>
<td>USA</td>
<td>ND</td>
<td>95</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Hassan</td>
<td>USA</td>
<td>-80</td>
<td>45</td>
<td>ND</td>
<td>Yes</td>
<td>Consecutive</td>
</tr>
<tr>
<td>Badgwell</td>
<td>USA</td>
<td>-80</td>
<td>254</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Scholler</td>
<td>USA</td>
<td>-80</td>
<td>124</td>
<td>ND</td>
<td>Yes</td>
<td>ND</td>
</tr>
<tr>
<td>Palmer</td>
<td>USA, Sweden</td>
<td>ND</td>
<td>101</td>
<td>Prospective</td>
<td>Yes</td>
<td>Consecutive</td>
</tr>
<tr>
<td>Moore</td>
<td>USA</td>
<td>-80</td>
<td>233</td>
<td>Prospective</td>
<td>Yes</td>
<td>Consecutive</td>
</tr>
<tr>
<td>Shah</td>
<td>USA</td>
<td>-80</td>
<td>267</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Abdel-Azeez</td>
<td>Egypt</td>
<td>-80</td>
<td>65</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Hellstrom</td>
<td>USA</td>
<td>-80</td>
<td>66</td>
<td>Prospective</td>
<td>Yes</td>
<td>Consecutive</td>
</tr>
<tr>
<td>Huhtinen</td>
<td>Finland</td>
<td>-20</td>
<td>148</td>
<td>Retrospective</td>
<td>ND</td>
<td>Consecutive</td>
</tr>
<tr>
<td>Nolen</td>
<td>USA</td>
<td>ND</td>
<td>790</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Gorp</td>
<td>Belgium</td>
<td>-80</td>
<td>65</td>
<td>Prospective</td>
<td>ND</td>
<td>Consecutive</td>
</tr>
<tr>
<td>Jacob</td>
<td>Switzerland</td>
<td>-80</td>
<td>160</td>
<td>Prospective</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Holcomb</td>
<td>ND</td>
<td>229</td>
<td>ND</td>
<td>Retrospective</td>
<td>Yes</td>
<td>Consecutive</td>
</tr>
<tr>
<td>Chang</td>
<td>China</td>
<td>-70</td>
<td>202</td>
<td>ND</td>
<td>Yes</td>
<td>Consecutive</td>
</tr>
<tr>
<td>Montagnana</td>
<td>Italy</td>
<td>-80</td>
<td>104</td>
<td>Retrospective</td>
<td>ND</td>
<td>Consecutive</td>
</tr>
<tr>
<td>Park</td>
<td>Korea</td>
<td>-80</td>
<td>323</td>
<td>Retrospective</td>
<td>ND</td>
<td>Consecutive</td>
</tr>
</tbody>
</table>

Table 2. Test for Heterogeneity and Threshold Effect in the Meta-analysis

<table>
<thead>
<tr>
<th>Marker Combination</th>
<th>Likelihood ratio</th>
<th>I² index(%)</th>
<th>χ²</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE4</td>
<td>43.37</td>
<td>76.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMRP</td>
<td>51.59</td>
<td>84.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA125+HE4</td>
<td>10.29</td>
<td>32</td>
<td>0.173</td>
<td></td>
</tr>
<tr>
<td>CA125+SMRP</td>
<td>26.41</td>
<td>92.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMRP+HE4</td>
<td>6.02</td>
<td>66.8</td>
<td>0.049</td>
<td></td>
</tr>
<tr>
<td>CA125+SMRP+HE4</td>
<td>7.58</td>
<td>73.6</td>
<td>0.023</td>
<td></td>
</tr>
</tbody>
</table>

Diagnostic Value of Human Epididymis Protein 4 Compared to Mesothelin for Ovarian Cancer

As shown in Table 1, the average sample size of the HE4 studies was 217 (range: 65 to 323), whereas that of the SMRP studies was 145 (range: 45 to 267). The studies were published between 1999 and 2011. Ten studies (55.6%) collected samples from consecutive patients. Seven studies (38.9%) reported blinded interpretation of tumor marker assays. Six studies (33.4%) had prospective study designs. Three studies (16.7%) were retrospective. All studies used histopathologic analysis as their reference standard.

Diagnostic accuracies of HE4 and SMRP

Publication bias and heterogeneity To assess possible publication biases, scatter plots were designed using the log DORs of individual studies against their sample size. The plots of these meta-analyses (Figure 2) were symmetric, demonstrating no publication bias. Heterogeneity was observed for most markers, which was confirmed by either the likelihood ratio χ² test or the I² index (Table 2). The random-effects model was used to estimate the overall effect.

Diagnostic accuracy

The SROC curve graphs for the determination of tumor markers that show TP rates versus FP rates from individual studies are shown in Figure 2. Pooled results of the diagnostic accuracy of each tumor marker and marker.
In the present study, HE4 and SMRP showed significant accuracy. It indicates the frequency (expressed as odds) was better than that of HE4 or SMRP alone. SMRP or CA-125 + SMRP for predicting ovarian cancers that the AUCs of HE4 + SMRP, CA125 + SMRP, HE4, and specificity is inappropriate. In addition, HE4 had a higher specificity of 0.945 (range: 0.926 to 0.958) for the diagnosis of ovarian cancers. Hence, SMRP assays are not suitable for screening general patients and may have limited value when utilized as a first-line diagnostic test but may be useful in confirming cancer diagnosis. HE4, a secreted glycoprotein overexpressed by serous and endometrioid epithelial ovarian carcinomas (Drapkin et al., 2005), is believed to be a sensitive and specific screening test that can detect ovarian cancer at a curative stage. Currently, CA-125 remains the most widely used biomarker for ovarian cancer. Medeiros et al. (2009) conducted a meta-analysis to estimate the accuracy of CA-125 assays in the diagnosis of ovarian tumors. The pooled sensitivity and specificity for the diagnosis of borderline tumors or ovarian cancer were 0.80 (95% CI: 0.76 to 0.82) and 0.75 (95% CI: 0.73 to 0.77), respectively. The meta-analysis performed in the current paper showed that HE4 assays had pooled sensitivity and specificity of 74% (95% CI: 0.76 to 0.82) and 0.75 (95% CI: 0.73 to 0.77), respectively. The meta-analysis included in the meta-analysis evaluated the simultaneous determination of two or more serum tumor markers in ovarian cancer diagnosis. The pooled results of the diagnostic accuracy of the combinations of CA-125, HE4, and SMRP are shown in Table 3. The results indicated that some combinations of tumor markers had a higher diagnostic role than one tumor marker alone. CA-125 + HE4 achieved the highest sensitivity in ovarian cancer diagnosis at 0.80 (range: 0.766 to 0.833). The specificity, PLR, NLR, and DOR of CA-125 + HE4 were 0.855 (range: 0.830 to 0.877), 11.42 (range: 3.80 to 34.35), 0.23 (range: 0.20 to 0.28), and 44.88 (range: 25.98 to 77.54), respectively. CA125+SMRP showed the highest specificity at 0.961 (range: 0.928 to 0.982). The sensitivity, PLR, NLR, and DOR of CA125+SMRP were 0.588 (range: 0.507 to 0.665), 12.03 (range: 2.93 to 49.42), 0.42 (range: 0.26 to 0.68), and 43.40 (18.52 to 101.66), respectively.

**Discussion**

EOC is the leading cause of death among gynecologic malignancies in the U.S. because of the lack of sensitive screening methods. Over the past 20 years, many studies have focused on the use of the tumor marker CA-125 and ultrasound imaging in ovarian cancer screening. Recently, the use of novel and multiple biomarkers has been proven to be an accurate method for preoperatively assessing the risk of ovarian cancer in women with ovarian masses. To the best of our knowledge, the present study is the first meta-analysis to estimate the pooled diagnostic accuracy characteristics of serum HE4 and SMRP, both individually and in various combinations with CA-125, in ovarian cancer.

Mesothelin is a 40 kDa cell surface glycoprotein that is highly expressed in pancreatic cancers, ovarian cancers, mesotheliomas, and other cancers (Chang & Pastan, 1996). The measurement of mesothelin in the blood may be useful for diagnosing and monitoring the response of patients to therapy because small amounts of mesothelin can be detected in the blood of patients with mesothelin-positive cancers (Scholler et al., 1999). Serum SMRP assays exhibited a limited diagnostic sensitivity of 0.493 (range: 0.454 to 0.532) because they failed to identify approximately half of the patients but demonstrated a high specificity of 0.945 (range: 0.926 to 0.958) for the diagnosis of ovarian cancers. Hence, SMRP assays are not suitable for screening general patients and may have limited value when utilized as a first-line diagnostic test but may be useful in confirming cancer diagnosis.
of positive test results among patients with the condition of interest compared with patients without the condition (Glas et al., 2003). The value of a DOR ranges from 0 to infinity; higher values indicate better discriminatory test performances (higher accuracy). The current meta-analysis showed that SMRP + HE4 exhibited the highest DOR in detecting ovarian cancer (DOR = 54.86) compared with CA125 + SMRP (DOR = 44.88) or CA125 + SMRP (DOR = 43.40). However, the SROC curve and DOR are not easy to interpret and use in clinical practice, and likelihood ratios are considered to be more clinically meaningful (Deeks, 2001). The current work included both PLR and NLR as the measures of diagnostic accuracy, which indicates high accuracy. Likelihood ratios greater than 10 or less than 0.1 generate large and often conclusive changes from pre-test to post-test probability. Likelihood ratios of 5 to 10 and 0.1 to 0.2 generate moderate but substantial shifts from pre-test to post-test probability (Jaeschke et al., 1994). In the current study, the PLR values of HE4 + CA-125 and CA-125 + SMRP were 11.42 and 12.03, respectively. These values are high enough for clinical purposes. However, the NLR values of HE4 + CA-125 and CA-125 + SMRP were 0.23 and 0.42, respectively. When HE4 + CA-125 or CA-125 + SMRP assay results are negative, the probability that the patient has ovarian cancer is 23% or 42%, which is not low enough to rule out ovarian cancer.

The current meta-analysis has several limitations. First, the exclusion of conference abstracts and unpublished data may have led to publication bias. Second, despite rigorous application of QUADAS criteria, most results showed heterogeneity in the assessment of sensitivity or specificity, and the heterogeneity in the diagnostic performance application of QUADAS criteria could not be clearly explained. This result may be mainly attributed to one or a combination of confounding factors in the literature. These factors include, the unclear age of the enrolled women, different study designs, study population, study quality, test equipment or methods, menstrual state, and different cut-off values of the HE4 or SMRP assays.

The current meta-analysis results showed that the diagnostic accuracy of HE4 to differentiate ovarian cancer from benign gynecologic diseases was better than that of SMRP. The combination of two or more tumor markers seems to be more sensitive and specific. Based on the gathered data, the serum tumor marker CA-125, in combination with HE4 or SMRP, can be used to help predict the presence of malignancies in patients with a pelvic mass.

Acknowledgements

The author(s) declare that they have no competing interests.

References


DOI:http://dx.doi.org/10.7314/APJCP.2012.13.11.5427

Diagnostic Value of Human Epididymis Protein 4 Compared to Mesothelin for Ovarian Cancer


