HE4: Clinical applications in gynaecological cancer

Stiekema, A.

Link to publication

Creative Commons License (see https://creativecommons.org/use-remix/cc-licenses):
Other

Citation for published version (APA):
A novel diagnostic index combining HE4, CA125 and age may improve triage of women with suspected ovarian cancer

an international multicenter study of women with an ovarian mass

MA Karlsen
EVS Høgdall
IbJ Christensen
C Borgfeldt
G Kalapotharakos
L Zdrazilova-Dubska
J Chovanec
CAR Lok

A Stiekema
I Mutz-Dehbalaei
AN Rosenthal
EK Moore
BA Schodin
WW Sumpaico
K Sundfeldt
B Kristjandsdottir
I Zapardiel

Gynecologic Oncology (138) 2015; 640-6
Abstract

Objective To develop and validate a biomarker-based index to optimize referral and diagnosis of patients with suspected ovarian cancer. Furthermore, to compare this new index with the Risk of Malignancy Index (RMI) and Risk of Ovarian Malignancy Algorithm (ROMA).

Patients and methods A training study, consisting of patients with benign ovarian disease (n=809) and ovarian cancer (n=246), was used to develop the Copenhagen Index (CPH-I) utilizing the variables serum HE4, serum CA125 and patient age. Eight international studies provided the validation population; comprising 1060 patients with benign ovarian masses and 550 patients with ovarian cancer.

Results Overall, 2665 patients were included. CPH-I was highly significant in discriminating benign from malignant ovarian disease. At the defined cut-off=0.070 for CPH-I the sensitivity and specificity were 95.0% and 78.4% respectively in the training cohort and 82.0% and 88.4% in the validation cohort. Comparison of CPH-I, ROMA and RMI demonstrated Area-Under-Curve (AUC) at 0.960, 0.954 and 0.959 respectively in the training study and 0.951, 0.953 and 0.935 respectively in the validation study. Using a sensitivity of 95%, the specificities for CPH-I, ROMA and RMI in the training cohort were 78.4%, 71.7% and 81.5% respectively, and in the validation cohort 67.3%, 70.7% and 69.5% respectively.

Conclusion All three indices perform well at the clinically relevant sensitivity=95%, but CPH-I, unlike RMI and ROMA, is independent of ultrasound and menopausal status, and may provide a simple index to optimize referral of women with suspected ovarian cancer.
Introduction

Correct referral of patients with an ovarian mass of potential malignant nature is a necessity to optimize treatment and prognosis of patients with ovarian cancer (OC). Benign ovarian masses can be managed at local gynaecologic departments, while suspected OC should be treated by a multidisciplinary team at a tertiary center. The treatment of OC patients by trained specialists in gynaecologic oncology has been shown to enhance overall survival. (1-5) In Denmark, as well as in several other countries, the Risk of Malignancy Index (RMI) consisting of serum cancer antigen 125 (CA125) level, specific ultrasound findings and menopausal status is the standard method of detecting patients at high risk of OC. (6) Because the RMI depends on the result of an ultrasound examination, only gynaecologists or other trained ultrasonographers can refer directly to tertiary centers resulting in increased costs and potential diagnostic delay.

Human Epididymis Protein 4 (HE4) performs as well as serum CA125 in differentiating OC from benign ovarian disease and shows an additive effect when combined. (7-9) The Risk Of Malignancy Algorithm (ROMA), based on serum CA125 level, serum HE4 level and menopausal status, has been demonstrated to distinguish OC from benign ovarian masses as well as the ultrasound dependent RMI at a set specificity (SP) (75%).(10) However, divergent results have been reported from other centers in the validation of ROMA. (11-19) All three variables in ROMA are easy to obtain from visit at the general practitioner. It is therefore less time and resource consuming compared to RMI. However, menopausal status can be a clinical challenge due to varying definitions (different age limits, no menstrual bleed for more than one year, serum Follicle-stimulating hormone (FSH) levels etc.). In contrast, age is an easy obtainable and reproducible variable. Furthermore, serum HE4 level steadily increases with age with no sudden increase at the time of the menopause; hence age might improve the performance of an index utilizing serum HE4 level. (20-22)

The primary aim of this study was to develop the Copenhagen Index (CPH-I), an index based on serum CA125, serum HE4 and age, to improve referral of patients with suspected OC to a tertiary center. The secondary aim was to validate the CPH-I in an international multicenter study, and to compare the CPH-I with the two established diagnostic indices ROMA and RMI.
Patients and methods

This study was a prospective, multicenter study with use of archived blood samples. Local ethics committees approved the local protocol for each recruitment center. The REMARK guidelines were followed to the extent possible for diagnostic biomarkers. (23)

2.1 Patients
Patients comprised a training population from a single institution (n=1055) for algorithm development, and a validation population (n=1610), comprising eight studies, seven of which were single institution cohorts, and one was a multi-institution study from six locations (Asia) (Table 1).

2.1.1 Training population:
All patients admitted to the tertiary center at Rigshospitalet, Denmark were included consecutively in case of a potentially malignant pelvic mass as previously described. (13) All diagnoses were histologically confirmed, and OC were compared to benign ovarian masses to develop the CPH-I. In order to minimize bias in referral because of inclusion from a tertiary center, patients referred to Rigshospitalet as their local hospital, as opposed to tertiary referrals, were also analyzed separately.

2.1.2 Validation population:
The CPH-I was validated using datasets from eight studies. All patients with benign ovarian disease or epithelial OC were included from tertiary centers: United Kingdom (UK) (n=98), the Netherlands (24) (n=178), Asia (19) (n=308), Austria (n=179), Czech Republic (n=65), Lund, Sweden (25) (n=292), Gothenburg, Sweden (16) (n=329) and Spain (n=161). The patients from Sweden (both Gothenburg and Lund), Asia, the Netherlands, UK, Czech and Spain were enrolled consecutively, and the Austrian patients were enrolled whenever possible. The Asian patients were enrolled from six different Asian centers (see acknowledgments). The study from Gothenburg excluded solid ovarian masses. Individual studies were selected for use in the validation cohort as they had similar inclusion criteria to the training cohort. Inclusion criteria were: CA125 and HE4 serum levels measured in preoperative blood samples and ethics committee permission to use these samples for research purposes. Furthermore, information regarding histology, stage of OC and, if performed, results from gynaecological ultrasound examination were added. The study from UK included patient who had not all been assessed by a gynecological oncologist. This may explain the low number of cases with OC in the UK study. The patients from the Czech study were highly selected before admitted to this center, which
may explain the low number of patients with benign ovarian disease compared to the other tertiary centers.

2.1.3 All cohorts:
Exclusion criteria were unknown histology, another active cancer, relapse of a previous cancer and histology other than benign ovarian masses or epithelial OC. Due to different study designs, we were not able to include patient cases of borderline ovarian tumours, non-epithelial ovarian cancer and metastases.

2.2 Methods
Blood samples were collected within 14 days prior to surgery, except in the Asian and UK studies, where collection time was not registered. The study from Gothenburg collected blood samples after anesthesia but prior to surgery. Blood samples were handled and preserved within a day after collection. Blood samples were stored at -20°C or colder (Denmark, Austria, Gothenburg, UK and Spain: -80°C, Asia: -70°C to -20°C, the Netherlands: -30°C, Lund: -20°C) until analyzed or analyzed at collection day (Czech). Serum HE4 levels were analyzed on either Abbott’s automated assay (Denmark, the Netherlands, Asia, Austria, Czech Republic and Spain) or on Fujirebio’s manual assay (UK, Gothenburg Sweden and Lund Sweden). CA125 levels were analyzed on Abbott’s automated assay (Denmark, the Netherlands, Asia, Gothenburg Sweden, Austria, Czech Republic and Spain) or on the Roche automated assay (Lund Sweden and UK). Inter-laboratory and inter-assay variability was examined. Values of CPH-I and ROMA were calculated, as well as RMI value, when information of ultrasound findings was accessible (Denmark, UK, the Netherlands, Asia and Austria).

2.3 Statistical analysis
Descriptive statistics are presented by the median and range for continuous variables. Spearman rank correlations were used as a measure of association. Logistic regression was used to develop the multivariate model CPH-I, discriminating patients with benign pelvic masses and OC. The covariates included were age, CA125 and HE4. CA125 and HE4 were log base2 transformed, resulting in odds ratios (OR) for a two-fold difference in the marker level. Age was scored, so OR is for a 10 year difference.

2.3.1 Training study
Univariate and multivariate analyses of CA125, HE4, and age were made to verify the significance of the variables separately and combined in CPH-I. A p-value<0.05 was considered statistically significant.
The coefficients in CPH-I are:

\[ \text{CPH-I} = -14.0647 + 1.0649 \times \log_2(\text{HE4}) + 0.6050 \times \log_2(\text{CA125}) + 0.2672 \times \frac{\text{age}}{10} \]

The predicted probability (PP) is:

\[ \text{PP} = \frac{e^{\text{CPH-I}}}{1 + e^{\text{CPH-I}}} \]

Goodness of fit assessment was done using the Hosmer-Lemeshow test. Internal model validation was done using 300 bootstraps. The development, evaluation and internal model validation of CPH-I was done according to biostatistical recommendations in multivariable models.\(^{(26)}\)

A relevant clinical cut-off was defined from the training study by setting the sensitivity (SN) of CPH-I equal to the SN of RMI at cut-off=200.

2.3.2 Validation study

The CPH-I was validated in eight independent studies. Results are presented by the receiver operator characteristics (ROC), where area under curve (AUC) was used for discrimination between benign ovarian disease and OC. Individual ROC-AUC for each study was calculated, except for the UK population due to a low number of patients with OC (n=2) and the Czech Republic population due to a low number of patients with benign disease (n=3). However, these populations were included in a pooled analysis of all validating studies. Furthermore, pooled analyses of CPH-I and ROMA for studies with RMI data (UK, the Netherlands, Asia and Austria) were made in order to compare with RMI.

Statistical calculations were performed using IBM SPSS (V19.0), SAS (v9.3, SAS Institute, Cary, N.C., USA), and R (R Core Team (2013). R: A language and environment for statistical computing. R statistical Computing. Vienna. Austria. URL. http://www.R-project.org/).
Results

Overall, 2665 patients were included from nine independent studies. The distribution in all studies is presented in table 1. Table 2 demonstrates the median and range for HE4, CA125 and age.

3.1 Training study

The training study included 1055 patients with benign ovarian disease (n=809) or OC (n=246). Early stage OC (FIGO stage I-II) was diagnosed in 60 patients (24%) (table 1).

The association between CA125 and HE4 was r=0.57 (p<0.001). The association between age and CA125 was r=0.30 (p<0.001), and between age and HE4 r=0.57 (p<0.001). Univariate analysis demonstrated a highly significant ability to differentiate benign ovarian disease from OC by HE4, CA125 and age, respectively (all three p-values<0.001). ROC-AUC for each variable are shown in table 3. In multivariable analysis both HE4 and CA125 contributed significantly (p-values<0.001) to the CPH-I model, as well as age (p=0.002). Age demonstrated an odds ratio (OR) of 1.31 (95% confidence interval (CI) 1.10-1.55), and the markers CA125 and HE4 demonstrated an OR of 1.83 (1.58-2.13) and 2.90 (2.21-3.82), respectively. The p-value for the goodness of fit test was 0.76, suggesting an adequate model formulation. Internally validation in the training study by 300 bootstraps yielded only slightly higher standard errors than the asymptotic results (data not shown). The AUC was 0.959 after bootstrapping. Addition of menopausal status to the model did not contribute significantly to the model (p=0.17), and replacement of age with menopausal status resulted in a poorer fit of the model. AUC of CPH-I, ROMA and RMI are demonstrated in table 3 and figure 1. AUC for early stage OC are displayed in table 3.

Subgroup analyses of Danish patients referred to Rigshospitalet as a local hospital (n=415), as opposed to tertiary referrals, comprised 345 patients (83%) with benign ovarian disease and 70 patients (17%) with OC. The AUC for this subgroup was 0.950 for the CPH-I, 0.933 for ROMA and 0.955 for RMI.

To ensure a clinical relevant cut-off of the CPH-I, the SN of RMI at predefined cut-off value=200 was used, which corresponded to SN=95.0% for benign vs. OC, and a SN=87.5% for benign vs. early stage OC in the training study. At a set SN=95% in the training cohort, a cut-off at PP =0.070 and SP at 78.4% were found. The SP for ROMA and RMI at SN=95% was 71.7% and 81.5% respectively. When differentiating early stage OC from benign disease at SN = 87.5%, the SP for the CPH-I, ROMA and RMI was 78.8%, 78.0% and 81.5% respectively.
Table 1
Patient data

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Benign ovarian disease</th>
<th>Ovarian cancer</th>
<th>stage I-II ovarian cancer</th>
<th>Stage III-IV ovarian cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denmark&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1055</td>
<td>809</td>
<td>246</td>
<td>60</td>
<td>186</td>
</tr>
<tr>
<td>UK</td>
<td>98</td>
<td>96</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>178</td>
<td>34</td>
<td>144</td>
<td>34</td>
<td>110</td>
</tr>
<tr>
<td>Asia</td>
<td>308</td>
<td>271</td>
<td>37</td>
<td>10</td>
<td>27</td>
</tr>
<tr>
<td>Austria</td>
<td>179</td>
<td>92</td>
<td>87</td>
<td>33</td>
<td>54</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>65</td>
<td>3</td>
<td>62</td>
<td>18</td>
<td>44</td>
</tr>
<tr>
<td>Lund, Sweden</td>
<td>292</td>
<td>207</td>
<td>85</td>
<td>28</td>
<td>57</td>
</tr>
<tr>
<td>Gothenburg, Sweden</td>
<td>329</td>
<td>215</td>
<td>114</td>
<td>57</td>
<td>57</td>
</tr>
<tr>
<td>Spain</td>
<td>161</td>
<td>142</td>
<td>19</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>2665</strong></td>
<td><strong>1869</strong></td>
<td><strong>796</strong></td>
<td><strong>246</strong></td>
<td><strong>550</strong></td>
</tr>
</tbody>
</table>

<sup>a</sup> Training population
A novel diagnostic index combining HE4, CA125 and age may improve triage of women with suspected ovarian cancer.

Table 2
Baseline data on the variables age, CA125 and HE4

<table>
<thead>
<tr>
<th></th>
<th>Benign cases</th>
<th>Ovarian cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Age median (range)</td>
</tr>
<tr>
<td>Denmark</td>
<td>809</td>
<td>43 (19-90)</td>
</tr>
<tr>
<td>UK</td>
<td>96</td>
<td>41 (18-88)</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>34</td>
<td>56 (34-84)</td>
</tr>
<tr>
<td>Asia</td>
<td>271</td>
<td>38 (19-75)</td>
</tr>
<tr>
<td>Austria</td>
<td>92</td>
<td>49 (15-84)</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>3</td>
<td>44 (42-71)</td>
</tr>
<tr>
<td>Lund, Sweden</td>
<td>207</td>
<td>50 (16-88)</td>
</tr>
<tr>
<td>Gothenburg, Sweden</td>
<td>215</td>
<td>60 (16-88)</td>
</tr>
<tr>
<td>Spain</td>
<td>142</td>
<td>46 (19-95)</td>
</tr>
<tr>
<td>All validating cohorts</td>
<td>1060</td>
<td>48 (15-88)</td>
</tr>
</tbody>
</table>

A HE4 analyses were analyzed using Abbott’s assay
B HE4 analyses were analyzed using Fujirebio’s assay
C Missing CA125 value in one patient
D Missing age value in 11 patients
E Missing age value in four patients
F Missing CA125 value in two patients
### Table 3

ROC-AUC of Copenhagen Index (CPH-I), ROMA and RMI in differentiation of benign ovarian disease and OC

<table>
<thead>
<tr>
<th></th>
<th>Ovarian cancer versus benign ovarian disease</th>
<th>Stage I-II ovarian cancer versus benign ovarian disease</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CA125</td>
<td>HE4</td>
</tr>
<tr>
<td>Denmark</td>
<td>0.931</td>
<td>0.940</td>
</tr>
<tr>
<td>UK a</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>0.902</td>
<td>0.946</td>
</tr>
<tr>
<td>Asia</td>
<td>0.894</td>
<td>0.922</td>
</tr>
<tr>
<td>Austria</td>
<td>0.827</td>
<td>0.937</td>
</tr>
<tr>
<td>Czech Republic b</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lund, Sweden c</td>
<td>0.921</td>
<td>0.932</td>
</tr>
<tr>
<td>Gothenburg, Sweden c</td>
<td>0.868</td>
<td>0.844</td>
</tr>
<tr>
<td>Spain c</td>
<td>0.864</td>
<td>0.917</td>
</tr>
<tr>
<td>All validating cohorts</td>
<td>0.894</td>
<td>0.911</td>
</tr>
<tr>
<td>All validating studies with RMI data</td>
<td>0.951</td>
<td>0.953</td>
</tr>
</tbody>
</table>

A Analyses not possible due to low number of patients with OC (n=2), but data are included in “all international cohorts”

B Analyses not possible due to low number of patients with benign disease (n=3), but data are included in “all international cohorts”

C No RMI data available
3.2 Validation study
The validation study included 1610 patients with benign ovarian disease (n=1060) or OC (n=550). 186 (34%) patients with OC had early stage OC.
A pooled analysis of all international data demonstrated an overall AUC for CA125 of 0.894 and for HE4 of 0.911. The cohort from Gothenburg stood out as the only cohort, where HE4 (AUC=0.844) performed worse than CA125 (AUC=0.868). CPH-I, ROMA and RMI, were compared. However, it should be noted, that there are missing RMI data in the cohorts from Lund, Gothenburg, Czech and Spain (n=847, 52.6% of the validation population). Performances of the three indices in the validation study are demonstrated in table 3 by individual and pooled ROC-AUC values. The pooled analyses are for all cohorts as well as subgroup analysis of the four cohorts including RMI data. ROC-AUC values for early stage OC are furthermore shown in table 3.
Table 4 demonstrates SN and SP at the defined cut-off PP=0.070, SP at SN=95% and SN at SP=75% in individual studies as well as in pooled analyses.

Figure 1
ROC-AUC of Copenhagen Index (CPH-1), ROMA and RMI in differentiating benign ovarian disease from ovarian cancer. Curves are demonstrated for each index separately in both the Danish training cohort and the international validation cohort (pooled)
### Table 4

Sensitivities (SN) and specificities (SP) of Copenhagen Index (CPH-I), ROMA and RMI in differentiation of benign ovarian disease and ovarian cancer

<table>
<thead>
<tr>
<th></th>
<th>CPH-I Cut-off PP (_{\geq 0.070})</th>
<th>SP at set SN = 95% (^{A})</th>
<th>SN at set SP = 75%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SN</td>
<td>SP</td>
<td>CPH-I</td>
</tr>
<tr>
<td>Denmark</td>
<td>95.0</td>
<td>78.4</td>
<td>78.4</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>87.5</td>
<td>82.4</td>
<td>58.8</td>
</tr>
<tr>
<td>Asia</td>
<td>75.7</td>
<td>95.2</td>
<td>68.5</td>
</tr>
<tr>
<td>Austria</td>
<td>78.2</td>
<td>94.6</td>
<td>54.4</td>
</tr>
<tr>
<td>Lund Sweden (^{B})</td>
<td>88.2</td>
<td>88.4</td>
<td>59.9</td>
</tr>
<tr>
<td>Gothenburg Sweden (^{B})</td>
<td>79.8</td>
<td>76.7</td>
<td>34.0</td>
</tr>
<tr>
<td>Spain (^{B})</td>
<td>84.2</td>
<td>92.3</td>
<td>78.2</td>
</tr>
<tr>
<td>All validating cohorts (^{C})</td>
<td>82.0</td>
<td>88.4</td>
<td>55.7</td>
</tr>
<tr>
<td>All validating cohorts with RMI data</td>
<td>95.5</td>
<td>64.2</td>
<td>67.3</td>
</tr>
</tbody>
</table>

\(^{A}\) *Set SN at RMI = 200 in the test cohort*

\(^{B}\) *No RMI data available*

\(^{C}\) *The UK and Czech Republic population were included in the pooled analysis of all international cohorts; but, due to low number of OC patients (n=2) in the UK population, and low number of benign patients (n=3) in the Czech Republic population, individual analyses of these two cohorts were not performed.*

\(^{D}\) *Predicted probability*
Discussion

Numerous efforts have been made to develop an index to differentiate OC from benign ovarian disease. In particular, an index that discriminates early stage OC from benign ovarian disease is desirable due to the poor prognosis when OC has spread beyond the ovaries. The RMI is widely used for diagnostic purpose and has been demonstrated to distinguish OC from benign ovarian masses with a SN of 92% and a SP of 82% (cut-off=200) when used in a tertiary center.(27) Other ultrasound-based prediction models, such as the IOTA Logistic Regression model 2 (LR2) and the Simple Rules (SR), seem promising; and in a meta-analysis evaluating different models for pre-surgical diagnosis of OC, use of an ultrasound-based model is recommended.(28) However, ultrasound is not readily accessible in primary care and medical diagnostic centers and therefore requires referral to a gynaecologist or another experienced sonographer to decide whether the patient should be managed at a local gynaecologic department or at a tertiary center specialised in gynaecologic oncology. In search of an easily applicable method to triage patients with a palpable pelvic mass or suspicious symptoms without the use of ultrasound, several biomarkers have been evaluated, and the tumour marker HE4 has received considerable interest in this area. The HE4 tumour marker level is stable under different handling conditions.(29) Furthermore, no significant differences in HE4 levels have been found in asymptomatic, high risk women with different ethnicities. (21) Therefore, it is expected that a HE4 based index can be useful worldwide.

After demonstration of the additive diagnostic performance of CA125 and HE4, Moore et al developed ROMA.(10) Menopausal status is included in ROMA, but since HE4 increases steadily with age(20), and age is an easily obtainable and reproducible variable compared to menopausal status, it would make sense to substitute menopausal status with age as in CPH-I. A higher AUC of combination of HE4, CA125 and age versus ROMA have been demonstrated in a smaller study differentiating stage I OC and benign ovarian disease, though how the authors included age is not clear.(30) In our training study, the substitution of age with menopausal status in the multivariate analysis resulted in a poorer fit of the model. Beside age, renal function has been demonstrated to affect the serum HE4 level in healthy people.(31) Recently, Kappelmayer et al presented a logistic regression model including serum HE4, serum CA125 and eGFR tested in selected patients. However, addition of eGFR did not enhance the diagnostic accuracy of OC. (32) All three covariates in CPH-I (age, CA125 and HE4) were highly significant in discriminating benign ovarian disease from OC when tested separately and combined. The AUC of HE4 was slightly larger than the AUC of CA125 in both the training cohort and validation cohort, in concordance with earlier studies.(7,15,30) CPH-I performed equivalent to RMI and better than ROMA in the training study, though this was expected, since CPH-I was optimized on these data. More interestingly, in
the international validation population, the CPH-I and ROMA performed equally in discriminating OC from benign disease and early stage OC from benign disease. Caution in comparing RMI to CPH-I and ROMA is necessary due to the lack of RMI data in four cohorts accounting for 52.6\% of the validation cohort. In the pooled AUC, using cohorts with RMI data, CPH-I and ROMA were superior to RMI, both for comparison of OC versus benign ovarian disease and early stage OC versus benign ovarian disease. In this study, the chosen cut-off value for CPH-I was defined at PP=0.070. At this cut-off value, the SN was 95\% in the training study, meaning that 19 out of 20 patients with OC would be correctly referred to a tertiary center. The SN was only 82.0\% in the validating study, though a higher SP was found (SP=88.4\%) compared to the training study (SP=78.4\%). The most relevant cut-off should be considered from local strategies and capacities and with consideration to inter-laboratory and inter-assay differences, though a high SN is desirable due to the improved survival of patients treated by gynecologic oncologists.(1-5) Furthermore, before defining an optimal cut-off for a specific population, positive and negative predictive values should be evaluated taking the prevalence of OC in each specific population into account.
At SN = 95%, CPH-I performed better than ROMA in both the training study and validation study, although in the validation study with RMI data, RMI and ROMA were superior to CPH-I. This might be explained by exclusion of the study from Gothenburg to the latter group. HE4, CPH-I and ROMA perform markedly worse in the Gothenburg population compared to other populations (table 3 and figure 2), which may be explained by the different inclusion criteria (only cystic tumours) or by the older population with benign ovarian disease (table 2).

Our study has some limitations. Firstly, all patients were included by tertiary centers, and inclusion bias is therefore inevitable. However, subgroup analysis of patients admitted to Rigshospitalet, Denmark as their local hospital, resulted in an AUC for CPH-I of 0.959 compared to 0.960 for all patients. This suggests that, although CPH-I was not investigated in a primary care setting, the results can be extrapolated and that CPH-I may be useful in this setting. A possible disadvantage of generalizing a training set from one institution to a validation set from multiple, international institutions is the potential selection bias of patients before inclusion in each separate study and the differences in diagnostic tools used in the different countries. Despite this, the CPH-I performs well in the different included cohorts and can therefore be considered as applicable in a broad variety of clinical settings. A further selection bias is present in our study, since we were not able to include borderline ovarian tumours, non-epithelial OC or metastases due to heterogeneity in inclusion criteria in the contributing studies. We would expect that inclusion of borderline, non-epithelial OC and metastases would improve the performance of a HE4-based index due to the significantly increased serum level of HE4 in OC (31,33), though differentiation of benign ovarian disease from borderline, non-epithelial OC and metastases could be a challenge.

Furthermore, serum HE4 and CA125 levels were measured using two different assays at nine different laboratories, so inter-assay and inter-laboratory variability are substantial with the inter-laboratory variation being larger than inter-assay variation (data not shown). The variation may be due to sampling of specimens, population heterogeneity as well as assay differences. Despite these potential confounders, we found that CPH-I performed well compared to the established RMI and ROMA tests. Performance would likely be improved with greater standardisation of sample collection and assays and adjustment of cut-offs in each individual laboratory. Additionally, the lack of ultrasound data in more than half of the patients in the validation study hampered the comparison of all three indices. This demonstrates the problem with using an ultrasound based index because of reduced accessibility to ultrasound in some countries.

Finally, a limitation is the difficulty in accurate determining menopausal status for use in RMI and ROMA. Some clinics use age >50 years or other age limits, some register women as postmenopausal if they have not had a menstrual bleed for more
than one year, yet others use serum FSH levels as a guide. Definitions of menopausal status are only a hurdle in RMI and ROMA, since age has replaced menopausal status in the CPH-I. Furthermore, CPH-I seems promising as a diagnostic index in settings where ultrasonography is not easily accessible. Overall, CPH-I performed as well as ROMA and RMI in an international validation study, differentiating OC and early stage OC from benign ovarian disease. Age, as included in CPH-I, is objective and easily obtained, in contrast to menopausal status (used in ROMA and RMI) and ultrasound information (used in RMI). Thus CPH-I may have a role in a primary care and diagnostic unit setting to improve triage of women with suspected OC.

**Conflict of interest**
Beth Schodin is employed by Abbott Laboratories and has stocks in the company. Adam Rosenthal has received honoraria from Fujirebio Diagnostics for lecturing at conferences on topics not involving HE4. Karin Sundfeldt was a member of an advisory board at one occasion (June 2012) for Fujirebio Inc.
A novel diagnostic index combining HE4, CA125 and age may improve triage of women with suspected ovarian cancer

**References**


