Designing early detection programs for ovarian cancer

N. Urban*
Molecular Diagnostics Program, Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, USA

Human epididymis protein 4 (HE4) shows promise as a serum marker that complements CA125 in the early detection of epithelial ovarian cancer, either as a first-line screen or as a second-line screen in a multimodal strategy. Incorporation of symptoms in a screening strategy that includes CA125 and HE4 may warrant further research.

Key words: HE4, CA125, mesothelin, MMP7, symptoms index, screening, ovarian cancer

introduction
Epithelial ovarian cancer (EOC) is a significant cause of mortality both in the United States (US) and worldwide, due largely to the high proportion of cases that present at late stage when survival is poor. Early detection of EOC, particularly the serous subtype, is a promising strategy for reducing mortality. However, the low incidence of ovarian cancer makes the development of an adequately sensitive and specific screening strategy very challenging.

objective
Our goal has been to identify screening tools and evaluate their use together in order to identify a screening strategy that is clinically acceptable and cost-effective for use in both the general and high-risk populations. Our goal is to improve lead time as well as sensitivity while maintaining an acceptable rate of screening-prompted surgical procedures, and to achieve a mortality reduction of 30%.

the novel markers trial

background
Evidence from various studies was used to inform the design of the Novel Markers Trial, a multi-institutional phase I screening trial that is sponsored in part by the NCI-funded Specialized Program of Research Excellence (SPORE) [1]. This two-arm randomized study introduces human epididymis protein 4 (HE4) as either a first- or second-line screen in a multimodal screening strategy that includes CA125, HE4 and transvaginal ultrasound (TVU). Women are eligible for the Novel Markers Trial if they are aged >25 years with a documented deleterious mutation, aged >35 years with a significant family history, or aged >45 years with prior elevation in circulating CA125, HE4, mesothelin (MSLN) or matrix metalloproteinase 7 (MMP7). Mutation carriers and women with a family history are screened every 6 months; the remaining women are screened annually. Surgical consultation is recommended when either the level of CA125 exceeds the woman-specific threshold exceeding the 99th percentile or two of the three screening tests (CA125, HE4 and TVU) are positive.

High sensitivity and lead time motivate procedures for the first-line screen. In arm 1, CA125 and HE4 are used together every 6 or 12 months; the test is considered positive when either one exceeds a threshold consistent with 95% specificity, resulting in a callback for ~10% of women. In arm 2, CA125 is used alone at the same periodicity and specificity, resulting in a callback for 5% of women at each screen. In both arms, a callback is used to test both markers in order to confirm elevation in CA125 and/or HE4 before TVU is ordered. The lead time is increased by using the parametric empirical Bayes (PEB) longitudinal algorithm [2]. The PEB detects a rise in a marker by comparing a woman’s current marker level with a regression-adjusted baseline predicted from her prior values, tailoring positivity thresholds for CA125 and HE4 to the individual woman and lowering positivity thresholds for most women. Early recall is used to monitor marker levels in women who do not have surgery despite elevated markers, usually due to normal imaging.

The Novel Markers Trial screening protocol is modeled on the multimodal arm of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS), in which rising CA125 is used as a first-line screen to select women for TVU. Women with both rising CA125 and abnormal imaging results are referred for surgical consultation. In the single-modality imaging arm of the UKCTOCS, TVU is used annually as a first-line screen without consideration of CA125. The sensitivity, specificity and positive predictive value for all primary invasive EOC identified at the prevalence screen in the UKCTOCS were, respectively, 89.5%, 99.8% and 35.1% for the multimodal strategy, and 75.0%, 98.2% and 2.8% for TVU alone, suggesting uniformly better performance by the CA125-driven multimodal strategy than for annual imaging of all women [3].

*Correspondence to: Dr N. Urban, Molecular Diagnostics Program, Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, Mailstop M2-B230, Seattle, WA 98109, USA. Tel: +1-206-667-6774; Fax: +1-206-667-7264; E-mail: nurban@fhcrc.org

© The Author 2011. Published by Oxford University Press on behalf of the European Society for Medical Oncology. All rights reserved. For permissions, please email: journals.permissions@oup.com
The addition of a second serum marker to the Novel Markers Trial screening strategy was motivated in part by poor performance of TVU in the UKCTOCS, confirmed in a second large efficacy trial in the US. In the Prostate, Lung, Colon and Ovarian (PLCO) Screening Trial, women were screened annually for 6 years by CA125. For the first 4 years they were also screened annually by TVU; if either test was positive, surgical consultation was recommended. CA125 was considered positive if it exceeded a population threshold of 35 U/ml. In the first four rounds of screening, 89 women in the screening arm were diagnosed with EOC, of whom 67% were screen detected. Of these, 32% had abnormalities in both TVU and CA125. Of the remainder, nearly twice as many were positive for CA125 \( n = 27 \) as for TVU \( n = 14 \). The specificity was also better for CA125: among healthy controls, about twice as many imaging tests as CA125 tests were falsely positive (4.6%, 3.4%, 2.9% and 2.9% for TVU versus 1.4%, 1.6%, 1.8% and 1.7% for CA125 in the first, second, third and fourth years, respectively) [4].

HE4 was selected for use in the Novel Markers Trial screening strategy based on clinical and preclinical validation studies showing that HE4 signals ovarian malignancy with high specificity. Recently, HE4 has been shown to predict EOC in women with a pelvic mass [5–7], suggesting a role for HE4 as a second-line screen in the context of early detection. Like CA125, HE4 is approved in the US for use in recurrence monitoring, but is not approved or recommended for use in screening.

**Methods**

We used serial serum samples and data from large cohorts of women participating in a prevention trial and an ovarian cancer screening efficacy trial to perform retrospective validation studies of candidate serum markers and alternative multimodal screening strategies.

We first sought to identify the best serum markers to include in an ovarian cancer screening strategy by evaluating their signal in preclinical disease when the individual is likely to be asymptomatic. We used serial preclinical specimens from the Carotene and Retinol Efficacy Trial (CARET) to evaluate CA125, HE4, MSLN, B7-H4, decoy receptor 3, spondin-2 [8] and MMP7 for their promise as early detection markers. Immunoassays were used to measure these proteins in prediagnostic serum specimens (1–11 samples per participant).

Serial samples were provided 0–18 years before diagnosis by 34 CARET trial participants with ovarian cancer (15 with advanced-stage serous carcinoma) and during a comparable time interval before the reference date from 70 matched control participants. Loess curves were fitted to biomarker levels in cancer patients and control subjects separately to summarize mean levels over time. We also evaluated these circulating proteins as risk markers using a Cox regression model to estimate the hazard ratios associated with each marker.

We then collaborated with others to identify additional promising candidate serum markers. We used clinical ‘phase II’ specimens from 180 ovarian cancer cases and 660 benign disease or general population controls from four ovarian cancer SPORE sites to rank >50 serum markers. Thirty-five best markers, including 7 from our own laboratory, were evaluated in preclinical, proximate ‘phase III’ PLCO specimens from 118 women with ovarian cancer and from 474 matched controls [9]. A marker panel and accompanying decision rule were defined collaboratively, using logistic regression to obtain weights for a composite marker including the best markers from all sites [10]. In addition, we studied marker behavior in PLCO participants to evaluate if serum levels of candidate ovarian cancer biomarkers vary with individual characteristics of healthy women who participate in ovarian cancer screening. Finally, we evaluated the roles of the top markers CA125 and HE4 in multimodal strategies, including imaging and symptoms. We performed retrospective validation of the Novel Markers Trial decision rules using data and samples from the PLCO trial, which uniquely offers preclinical samples obtained prior to diagnosis in women who had both CA125 and TVU at every screen. We also considered the use of symptoms in screening for EOC. Until recently, symptoms of ovarian cancer were thought to develop only after the disease had progressed to an advanced stage, but now it is appreciated that women with early-stage disease often report nonspecific symptoms. A symptom index (SI) yields a decision rule that might be used to identify women at high risk for disease. Women reporting pelvic or abdominal pain, bloating, increased abdominal size, difficulty eating or feeling full quickly >12 times per month and reporting that these symptoms are new in the past 12 months are considered to have a positive SI. When symptoms are solicited rather than reported spontaneously, between 2% and 10% of women report symptoms consistent with a positive SI.

To evaluate the role of symptoms in a screening strategy, a prospective case-control study design including 74 women with ovarian cancer and 137 healthy women was used with logistic regression analysis to evaluate the independent contributions of HE4, CA125 and the SI to predict ovarian cancer status in a multivariate model. The diagnostic performance of various decision rules for combinations of these tests was assessed to evaluate potential use in predicting ovarian cancer [11].

**Results**

In the CARET validation study, the best risk and early detection markers were CA125, HE4, MSLN and MMP7. All markers except DcR3 predict future diagnosis of ovarian cancer. Loess curves demonstrate that mean concentrations of CA125, HE4 and MSLN began to rise in women with cancer relative to control subjects ~3 years before diagnosis, but reach detectable elevations only within the final year before diagnosis if 95% specificity is required for a first-line screen [8].

In the PLCO validation study, the top markers in phase II SPORE specimens included CA125, HE4, transhyretin (TT), CA15.3 and CA72.4 with sensitivity at 95% specificity ranging from 0.73 to 0.40. Except for TT, these markers had similar or better sensitivity when moving to phase III specimens that had been drawn within 6 months of the clinical diagnosis. However, the performance of all markers declined in phase III specimens obtained >6 months prior to diagnosis [9]. The all-site panel, which included CA125, HE4, CA72.4, secretory leukocyte peptidase inhibitor (SLPI) and β-2-microglobulin, performed better than CA125 alone in identifying the cases with a proximate sample drawn >6 months from diagnosis, but no better than...
CA125 alone for cases with a serum sample obtained within 6 months of diagnosis [10].

To evaluate the role of markers in a screening strategy, we considered adding a novel marker to the first- or second-line screen in a multimodal strategy. Using specificity for CA125 consistent with that used in the PLCO trial, we found that HE4 is the best novel marker to include in either a first-line screen with CA125 or a second-line screen with TVU. The SI was also found to contribute to a multimodal strategy in an evaluation using clinical ‘phase II’ data. The SI, HE4 and CA125 all made significant independent contributions to ovarian cancer prediction. A decision rule based on any one of the three tests being positive had a sensitivity of 95% with a specificity of 80%. A rule based on any two of the three tests being positive had a sensitivity of 84% with a specificity of 98.5%. The SI alone had a sensitivity of 64% with a specificity of 88%.

If the SI is used to select women for CA125 and HE4 testing, the specificity is 98.5% and the sensitivity is 58% using the two-of-three-positive decision rule [11].

**conclusions**

HE4 is the best marker other than CA125 identified to date, yielding a lead time of ≥1 year prior to diagnosis in preclinical serial samples from the CARET trial. Serum concentrations of CA125, HE4, MSLN and MMP7 are good risk markers for EOC, and may provide evidence of ovarian cancer 3 years before clinical diagnosis; however, they do not reach a high-specificity positivity threshold until ~1 year prior to clinical diagnosis, when symptoms are likely to be present. Analysis of PLCO proximate samples confirmed that HE4 ranks first among candidate markers to include in a first-line screen with CA125 for early detection.

Analysis of imaging and marker data from PLCO participants suggests that CA125 and HE4 both contribute to a screening strategy to detect ovarian cancer in asymptomatic postmenopausal women. Using both HE4 and CA125 in a first-line screen is more costly than using HE4 only in women with rising CA125, but it may improve the sensitivity. Based on these results, the Novel Markers Trial, a phase I trial of ovarian cancer screening in women at modestly elevated risk as well as women of different ages, introduces the use of HE4 as either a first- or second-line screen in a randomized design.

Additional markers have been identified that might assist in identifying women for testing by CA125, HE4 and/or TVUS. CA125, HE4 and other serum marker levels in healthy, postmenopausal women are significantly associated with women’s personal characteristics, such as age, smoking, body mass index and age of menarche. Incorporation of these covariates in screening algorithms is unnecessary if marker history is used in a longitudinal algorithm such as the PEB. However, failure to match cases and controls on these covariates in phase II clinical validation studies may lead to erroneous conclusions. Understanding the influence of personal factors on the levels of early detection markers in healthy women may also have clinical utility in interpreting serum marker levels in a screening program.

The incorporation of symptoms in a screening strategy that includes CA125 and HE4 may warrant further research. A two-of-three-positive decision rule yields acceptable specificity and higher sensitivity when all three tests are performed than when the SI is used to select women for screening by CA125 and HE4. If positive predictive value is a high priority, testing by CA125 and HE4 prior to imaging may be warranted for women with ovarian cancer symptoms.

**acknowledgements**

This work was supported by the Pacific Ovarian Cancer Research Consortium, Award Number P50 CA083636 from the National Cancer Institute. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Cancer Institute or the National Institute of Health. Support for the Translational and Outcomes Research Laboratory from NCI U01 CA152637 and the Canary Foundation is also gratefully acknowledged.

**disclosures**

None.

**references**