Published OnlineFirst April 8, 2016; DOI: 10.1158/1078-0432.CCR-16-0316

Cancer Therapy: Preclinical

Clinical Cancer Research

A Prospective Evaluation of Early Detection Biomarkers for Ovarian Cancer in the European EPIC Cohort

Kathryn L. Terry[1,2], Helena Schock[3], Renée T. Fortner[3], Anika Hüsing[3], Raina N. Fichorova[2,4], Hidemi S. Yamamoto[2,4], Allison F. Vitonis[1], Theron Johnson[3], Kim Overvad[5], Anne Tjønneland[6], Marie-Christine Boutron-Ruault[7,8], Sylvie Mesrine[7,8], Gianluca Severi[7,8,9,10], Laure Dossus[11], Sabina Rinaldi[11], Heiner Boeing[12], Vassiliki Benetou[13,14], Pagona Lagiou[13,14], Antonia Trichopoulou[13,14], Vittorio Krogh[15], Elisabetta Kuhn[16], Salvatore Panico[17], H. Bas Bueno-de-Mesquita[18], N. Charlotte Onland-Moret[19], Petra H. Peeters[19], Inger Torhild Gram[20], Elisabete Weiderpass[20,21,22,23], Eric J. Duell[24], Maria-Jose Sanchez[25,26], Eva Ardanaz[26,27,28], Nerea Etxezarreta[26,29], Carmen Navarro[26,30,31], Annika Idahl[32], Eva Lundin[33], Karin Jirström[34,35], Jonas Manjer[34,35], Nicholas J. Wareham[36], Kay-Tee Khaw[37], Karl Smith Byrne[38], Ruth C. Travis[38], Marc J. Gunter[39], Melissa A. Merritt[39], Elio Riboli[39], Daniel W. Cramer[1], and Rudolf Kaaks[2]

Abstract

Purpose: About 60% of ovarian cancers are diagnosed at late stage, when 5-year survival is less than 30% in contrast to 90% for local disease. This has prompted search for early detection biomarkers. For initial testing, specimens taken months or years before ovarian cancer diagnosis are the best source of information to evaluate early detection biomarkers. Here we evaluate the most promising ovarian cancer screening biomarkers in prospectively collected samples from the European Prospective Investigation into Cancer and Nutrition study.

Experimental Design: We measured CA125, HE4, CA72.4, and CA15.3 in 810 invasive epithelial ovarian cancer cases and 1,939 controls. We calculated the sensitivity at 95% and 98% specificity as well as area under the receiver operator curve (C-statistic) for each marker individually and in combination. In addition, we evaluated marker performance by stage at diagnosis and time between blood draw and diagnosis.

Results: We observed the best discrimination between cases and controls within 6 months of diagnosis for CA125 (C-statistic = 0.92), then HE4 (0.84), CA72.4 (0.77), and CA15.3 (0.73). Marker performance declined with longer time between blood draw and diagnosis and for earlier staged disease. However, assessment of discriminatory ability at early stage was limited by small numbers. Combinations of markers performed modestly, but significantly better than any single marker.

Conclusions: CA125 remains the single best marker for the early detection of invasive epithelial ovarian cancer, but can be slightly improved by combining with other markers. Identifying novel markers for ovarian cancer will require studies including larger numbers of early-stage cases. *Clin Cancer Res; 22(18); 4664–75.* *©2016 AACR.*

See related commentary by Skates, p. 4542

[1]Ob/Gyn Epidemiology Center, Brigham and Women's Hospital, Boston, Massachusetts. [2]Obstetrics, Gynecology and Reproductive Biology, Harvard Medical School, Boston, Massachusetts. [3]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. [4]Genital Tract Biology Laboratory, Brigham and Women's Hospital, Boston, Massachusetts. [5]Department of Epidemiology, Aarhus University, Aarhus, Denmark. [6]Danish Cancer Society Research Center, Copenhagen, Denmark. [7]Université Paris-Sud, Centre de recherche en Épidémiologie et Santé des Populations (CESP), Institut Nationale de Santë et de Recherche Médicale (INSERM), Villejuif, France. [8]Gustave Roussy, Villejuif, France. [9]Human Genetics Foundation (HuGeF), Torino, Italy. [10]Cancer Council Victoria and University of Melbourne, Australia. [11]Nutrition and Metabolism Section, International Agency for Research on Cancer, Lyon, France. [12]Department of Epidemiology, German Institute of Human Nutrition, Potsdam-Rehbruecke, Nuthetal, Germany. [13]Helenic Health Foundation Athens, Athens, Greece. [14]WHO Collaborating Center for Nutrition and Health, Unit of Nutritional Epidemiology and Nutrition in Public Health, Department of Hygiene, Epidemiology and Medical Statistics, University of Athens Medical School, Athens, Greece. [15]Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy. [16]Department of Morphology, Surgery and Experimental Medicine and Laboratorio per le Tecnologie delle Terapie Avanzate (LTTA) Centre, University of Ferrara, Ferrara, Italy. [17]Dipartimento di Medicina Clinical e Chirurgia, Federico II University, Naples, Italy. [18]National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands. [19]Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands. [20]Department of Community Medicine, University of Tromsø, The Arctic University of Norway, Tromsø, Norway. [21]Department of Research, Cancer Registry of Norway, Institute of Population-Based Case Research, Oslo, Norway. [22]Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden. [23]Genetic Epidemiology Group, Folkhälsan Research Center, Helsinki, Finland. [24]Unit of Nutrition and Cancer Cancer Epidemiology Research Program, Bellvitge Biomedical Research Institute, Catalan Institute of Oncology, Barcelona, Spain. [25]Escuela Andaluza de Salud Publica, Instituto de Investigación Biosanitaria ibs. Granada, Hospitales Universitarios de Granada, Granada, Spain. [26]Centro de Investigación Biomédica En Red (CIBER), Section Epidemiology and Public Health (CIBERESP), Madrid, Spain. [27]Navarra Public Health Institute, Pamplona, Spain. [28]IdiSNA, Navarra Institute

Translational Relevance

Ovarian cancer is a leading cause of cancer-related death among women. About 60% of ovarian cancers are diagnosed at late stage, when 5-year survival is less than 30% in contrast to 90% for local disease. Biomarkers for early detection are urgently needed to improve survival. Using blood samples of ovarian cancer cases and cancer-free control subjects from the European EPIC study, we examined the prospective diagnostic capacity of CA125, HE4, CA72.4, and CA15.3. All markers were significantly elevated many months before clinical manifestation of ovarian cancer. The best discrimination between cases and controls was within 6 months of diagnosis for CA125 (C-statistic = 0.92) and HE4 (0.84). Marker performance declined with longer time between blood draw and diagnosis and for earlier staged disease. Combining CA125 with HE4 and further markers modestly improved discrimination. Our study confirms CA125 as the single best marker for the early detection of invasive epithelial ovarian cancer.

Introduction

Ovarian cancer generally does not exhibit specific early symptoms. About 60% of ovarian cancers are diagnosed at late stage, which is associated with a 5-year survival of less than 30%, contrasted with more than 90% survival for disease found locally (1). This has prompted extensive research to find early detection biomarkers for ovarian cancer.

Many potential serum biomarkers for ovarian cancer have been indentified (2). Candidate biomarkers are often first identified from preclinical studies using immunohistochemical testing or gene expression profiles of tumor tissue. These are called phase I studies (3). Potential biomarkers are then tested by comparing blood from cases at diagnosis of ovarian cancer with blood from either women with benign disease or healthy controls. This type of study has been described as a phase II study. Markers that have been approved using phase II data include CA125, HE4, and a panel of markers, including prealbumin, apolipoprotein A-1, β_2-microglobulin, and transferrin (4, 5). CA125 has been approved for disease monitoring (6), and HE4 and a panel of markers as tools for distinguishing benign from malignant pelvic masses (4, 5). Phase III data refers to studies based on blood samples from asymptomatic women taken months or years prior to a diagnosis of ovarian cancer, while phase IV data refers to markers tested in a clinical trial in which asymptomatic women are randomized to a screening arm or to usual care.

There have been three randomized trials of screening for ovarian cancer using either CA125 alone or CA125 in combination with transvaginal ultrasound (TVUS). No reduction in ovarian cancer mortality was observed in the Prostate, Lung, Colorectal, and Ovarian cancer (PLCO) screening trial, based on a combination of TVUS and CA125 measurements for 4 years and two additional years of CA125 measurements (7). However, recent results from the UKCTOCs study showed a 15% reduction in mortality for postmenopausal women followed for change in CA125, which was marginally significant (8). One additional randomized trial in Japan (9) showed a nonsignificant increase in early-stage tumors detected in the screening arm but did not follow participants for mortality (9). To date, neither CA125 nor TVUS have been approved or recommended for screening on the basis of the randomized trials.

Although selected specimens from the phase IV studies may be and have been used in the context of discovering and testing new biomarkers, they may not be ideal for this purpose for at least two reasons. First, as CA125 was the primary screening tool, this may lead to preferential selection of CA125-expressing tumors. Second, as annual screening was employed, the natural history of the disease may have been interrupted at early stages and may not provide a true measure of the lead time, that is, the time between early diagnosis with screening and when diagnosis would have occurred in the absence of screening. Therefore, samples collected in asymptomatic women before cancer diagnosis are needed to test new biomarkers. To date, only a few case–control comparisons have been made in prospectively collected specimens obtained in asymptomatic women before clinical cancer diagnosis under usual care (phase III studies; refs. 10–12). Both the Carotene and Retinol Efficacy Trial (CARET) and Women's Health Initiative (WHI) studies were designed as randomized trials to evaluate other disease outcomes but have the advantage of closely monitoring a large group of women with banked blood samples. In the CARET study, a panel of markers including CA125, HE4, mesothelin, B7-H4, DcR3, and spondin-2 were measured on serial samples from 34 women with ovarian cancer and 70 matched controls. Of these, only CA125 and HE4 showed significant differences between cases and controls and had modest discriminatory ability that waned with increasing time between blood draw and diagnosis. Similarly, in the WHI study, CA125 and HE4 were measured in 353 ovarian cancer cases and 1,261 healthy controls and these markers significantly improved a risk prediction algorithm based on epidemiologic factors (11).

for Health Research, Pamplona, Spain. [29]Public Health Division of Gipuzkoa, Regional Government of the Basque Country, Basque, Spain. [30]Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Murcia, Spain. [31]Department of Health and Social Sciences, Universidad de Murcia, Murcia, Spain. [32]Department of Clinical Sciences, Obstetrics and Gynecology Umeå, University of Umeå, Umeå, Sweden. [33]Department of Medical Biosciences, Pathology Umeå, University of Umeå, Umeå, Sweden. [34]Department of Clinical Sciences Lund, Oncology and Pathology, Lund University, Lund, Sweden. [35]Department of Surgery, Skåne University Hospital, Lund University, Malmö, Sweden. [36]Medical Research Council (MRC) Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom. [37]Clinical Gerontology, University of Cambridge School of Clinical Medicine, Cambridge, United Kingdom. [38]Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom. [39]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College of London, London, United Kingdom.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Corresponding Authors: Rudolf Kaaks, Division of Cancer Epidemiology, German Cancer Research Center, Im Neuenheimer Feld 280, Heidelberg 69120, Germany. Phone: 4962-2142-2200; Fax: 4962-2142-2203; E-mail: r.kaaks@dkfz-heidelberg.de; and Kathryn L. Terry, Obstetrics and Gynecology Epidemiology Center Brigham and Women's Hospital and Harvard Medical School, 221 Longwod, Boston, MA 02115. Phone: 617-732-8596; Fax: 617-732-4899; E-mail: kterry@partners.org

doi: 10.1158/1078-0432.CCR-16-0316

Published OnlineFirst April 8, 2016; DOI: 10.1158/1078-0432.CCR-16-0316

Here, we present results from a phase III study using serum samples from the EPIC cohort (European Prospective Investigation into Cancer and Nutrition). We measured CA125, CA15.3, HE4, and CA72.4 in 197 cases of ovarian cancer diagnosed within the first 3 years after blood donation and 724 matched control subjects. For 613 additional ovarian cancer patients diagnosed more than 3 years after blood draw, and for 1,215 additional control subjects, we extended the measurements of CA125 and CA15.3 for examination of a possible longer-term risk diagnostic prediction capacity of these markers and to allow more accurate analyses of possible relationships of these markers with epidemiologic risk factors for ovarian cancer. The objectives of our analyses were: (i) to examine the early detection capacity of our biomarker panel for ovarian cancer diagnoses within comparatively short time intervals (variable lag time strata $\leq$3 years between blood donation and diagnosis); (ii) to examine the capacity of CA125 and CA15.3 to predict ovarian cancer risk over a longer term (>3 years between blood donation and diagnosis); and (iii) to examine whether early diagnostic capacity or longer-term risk prediction by the biomarkers could be improved by integrating further information about a woman's general epidemiologic risk factor profile.

Materials and Methods

The EPIC cohort—background and collection of blood samples

The European Prospective Investigation into Cancer and Nutrition is an ongoing multicenter prospective cohort study designed to investigate the relationship between diet, nutrition, and metabolic factors with cancer. Descriptions of study design, population, and baseline data collection of the cohort have been reported in detail previously (13, 14). In brief, 519,978 participants (366,521 women) were enrolled from 1992 to 2000 in 23 centers in 10 European countries: Denmark, France, Germany, Greece, Italy, Norway, the Netherlands, Spain, Sweden, and the United Kingdom. For women, the age range at recruitment was mostly between 35 and 70 years. At baseline, comprehensive data on diet, lifestyle, reproductive and menstrual factors, current and past use of exogenous hormones [oral contraceptives (OC) and postmenopausal hormone replacement therapy (HRT)] and medical history were collected through standardized questionnaires. In addition, anthropometric measures were obtained.

A total of 385,747 study participants in the EPIC cohort (226,673 women and 159,074 men) also provided a baseline blood sample. In France, the Netherlands, the United Kingdom, Germany, Spain, Italy, and Greece, blood samples were collected according to a standardized protocol (15). From each study participant, about 30 mL of blood was drawn, and serum, plasma, erythrocytes, and buffy coat were aliquoted in 28 plastic straws of 0.5 mL each, which were heat-sealed and stored under liquid nitrogen (−196°C). In Denmark, blood fractions were aliquoted into 1-mL tubes, and stored in the vapor phase in liquid nitrogen containers (−150°C). In the Swedish Center of Umeå, blood samples were divided into 10 aliquots of 1.5-mL each: 6 plasma, 2 buffy-coat, and 2 erythrocytes, which were rapidly frozen at −80°C in standard freezers.

Ascertainment of incident cancer cases

Prospective follow-up for cancer occurrences and histologic confirmation was performed through record linkage with cancer and pathology registries (all countries except France, Germany and Greece) or through active follow-up and systematic verification of self-reports by detailed examination and coding of clinical records. In all countries, vital status was determined by regular linkages with population and mortality registers at the regional or national level.

At the time the current study was initiated, prospective follow-up was complete until the end of 2005 (France) to 2008 (Germany). Within this timeframe, and among those women who had provided a baseline blood sample, a total of 810 incident invasive cases of ovarian cancer had been identified. Case subjects were defined as women who developed incident epithelial invasive ovarian (ICD code: C569), fallopian tube (C570), or peritoneal cancer (C480, C481, C482, C488) after recruitment into the EPIC study according to the International Classification of Diseases for Oncology (ICD) 0–3 and with data on tumor histology.

More detailed information on tumor characteristics [histologic subtype [serous, endometrioid, clear cell, mucinous, not otherwise specified (NOS), grade (well, moderately, or poorly/undifferentiated), and stage (local, regional, metastatic)] were obtained from pathology reports and from cancer registries. For the 810 invasive ovarian cases, complete information on tumor grade was available for 473 patients (58%) and information on tumor stage was available for 712 patients (88%). Well-differentiated tumors were classified as low grade; moderately and poorly/undifferentiated tumors were classified as high grade. We classified cases with local disease (stage I) as low stage and cases with regional (stage II) or metastatic disease (stage III/IV) as high stage.

Design of nested case–control study

For each case subject, up to four control subjects were randomly selected among appropriate risk sets consisting of all female cohort members with a blood sample, alive and free of cancer at the time of diagnosis of the index case. An incidence density sampling protocol was used, such that, in principle, control subjects could include study participants who became a case later in time and each control subject could be sampled more than once—the control subjects actually drawn, however, did not include any of the future cases of ovarian cancer detected so far in the EPIC cohort. Case and control subjects were matched on study recruitment center, age at blood donation ($\pm$6 months), time of the day of blood collection ($\pm$1 hour), fasting status (<3, 3–6, or >6 hours), follow-up time, and menopausal status at blood collection (premenopausal, perimenopausal, postmenopausal), current use of exogenous hormones (OCs, HRT) at the time of blood draw, as well as menstrual cycle phase for premenopausal women (3–5 categories, depending on available data). Cases missing data on phase of menstrual cycle were matched to control subjects whose information on menstrual cycle phase was also missing.

Informed consent and data protection

All participants had given their consent for future analyses of their blood samples and the current study was approved by the IARC Ethics Committee and the Institutional Review Board of Brigham and Women's Hospital (Boston, MA).

Laboratory assays

Measurements of the protein levels of CA125, CA15.3, HE4, and CA72.4 were completed for a total of 197 incident cases of invasive ovarian cancer and 725 matched, cancer-free control subjects. In addition, we measured CA125 and CA15.3 in 613

Published OnlineFirst April 8, 2016; DOI: 10.1158/1078-0432.CCR-16-0316

cases with invasive ovarian cancer who had been diagnosed more than 3 years after blood donation, and 1,214 additional matched control subjects. Laboratory values were missing for CA125 (3 cases, 12 controls), CA15.3 (6 cases, 19 controls), HE4 (2 controls), and CA72.4 (2 controls) due to lack of sufficient sample volume. All measurements were performed in the Genital Tract Biology Laboratory following preanalytic, analytic, and postanalytic SOPs established under the laboratory's accreditation by the College of American Pathologists, using a volume-effective highly sensitive multiplex platform (Meso Scale Discovery, MSD) based on electrochemoluminiscence (ECL) detection. Human CA125 (catalog number K151WC) and Human Prototype CA15.3 (catalog number N45ZA-1) were provided by MSD in singleplex assays. The linearity range for CA125 was 10,000–0.6 U/mL, and for CA15.3 was 12,500–0.19 mU/mL. The HE4 and CA72.4 measurements were done in a custom designed duplex assay. The following reagents were a gift from Fujirebio Diagnostics, Inc.: IgHE4 antigen, which we used to generate a calibration curve with a linear range starting at 3600 pmol/L; anti-HE4 capture IgG1 (2H5 mouse hybridoma, Fujirebio catalog number 414-01S); anti-HE4 detection IgG1 (mouse hybridoma 3D8, Fujirebio catalog number 415-01); TAG72 Defined Antigen, which we used to generate a calibrator curve with a linear range starting at 2,400 U/mL; anti-72.4 capture IgG1 (mouse hybridoma CC49, Fujirebio catalog number 110-005); anti-CA72.4 IgG1 (mouse hybridoma B72.3, Fujirebio catalog number 110-000). The samples were split into batches such that matched case–control sets and samples from the same study center were kept together in the same batches. The samples were tested undiluted in the CA125 singleplex and the HE4/CA72.4 duplex, and they were tested at a 50-fold dilution in the CA15.3 assay. A quality control pool was prepared from serum samples from ovarian cancer patients with within linearity range levels of each protein and split into equal aliquots. To establish interplate variability, one aliquot of this pool was tested at multiple dilutions spanning the linearity range of each assay, three dilutions run in duplicates and two dilutions run in triplicates, providing up to six quality control data points in each assay plate. In addition, blinded, randomly chosen citrated plasma, EDTA plasma, and serum Blood Bank samples were split into aliquots (128 for CA125, 130 for CA15.3, and 104 for HE4/CA72.4) and distributed within and between plates. Coefficients of variation (CV) were calculated as 100 × SD/mean. The unblinded quality control sample pool repeatedly tested on every assay plate showed the following interplate CV and min–max range (mean) of intraplate CV: (i) CA125, 8.4% interplate CV and 0.2%–13.5% (3.4%) intraplate CV; (ii) CA15.3: 15.4% interplate CV and 0.5%–6.1% (2.3%) intraplate CV; (iii) HE4: 8.99% interplate CV and 1.6%–7.6% (3.6%) intraplate CV; (iv) CA72.4: 17.3% interplate CV and 0.9%–13% (5.5%) intraplate CV. Similarly, the blinded aliquots with values within the linearity range of each assay showed the following interplate CVs and min–max (mean) intraplate CVs: 19% and 3%–20% (9%) for CA125, 22% and 3%–5% (4%) for CA15.3, 9% and 4%–10% (6%) for HE4, 16% and 1%–16% (6%) for CA72.4. As the majority of the blinded aliquots for CA72.4 fell below the lower limit of detection, blinded CA72.4 CVs were based on the remaining 13 aliquots, ranging in CA72.4 value from 1.15 to 1.87 U/mL.

Statistical analyses

First, we evaluated the distribution of each biomarker for normality and outliers. As 81% of the samples had CA72.4 values below the lower limit of detection for this assay (1.119 U/mL), we assigned these values to the midpoint between zero and the lower limit of detection for future analyses. Other markers assessed did not have any values below the lower limit of detection. Locally estimated scatterplot smoothing (LOESS) curves were used to describe mean levels of each marker among cancer cases and control subjects at different lag-times until ovarian cancer diagnosis. The discrimination between cases and control subjects was described using ROC) curves, with the AUC, also known as the C-(concordance) statistic, as an overall measure for discrimination capacity. We estimated the diagnostic sensitivities (SE95 and SE98, respectively) of each marker at cut-off points corresponding to 95% and 98% specificity, determined in our full dataset for all control subjects ($N = 1{,}939$ for CA125 and CA15.3; $N = 725$ for HE4 and CA72.4).

The diagnostic sensitivity, specificity, and C-statistics were calculated for risk scores based on the associations between biomarker levels and ovarian cancer risk, overall, and by strata of lag-time between blood donation and cancer diagnosis, conditional logistic regression models were used, accounting for the matched study design. Models were fitted for continuous biomarker measurements after $\log_2$-transformation, to achieve approximate normality of their distributions. Basic analyses focused on single markers. Additional multivariate models were developed to examine the discrimination capacity of multiple markers in combination, and of markers combined with an epidemiologic risk prediction algorithm, including age at menopause, duration of hormone replacement therapy, body mass index, unilateral ovariectomy, duration of oral contraceptive use, and number of full-term pregnancies that we developed previously on the basis of the full EPIC cohort data (16).

To examine how the early detection and/or risk prediction capacities of the biomarkers changed with time between blood draw and clinical cancer diagnosis, all analyses were performed within variable strata of lag-time ($\leq$6 months, $\leq$12 months, 1–2 years, 2–3 years, 3–6 years, >6 years). To examine heterogeneity of diagnostic prediction capacity by tumor stage at diagnosis or by histologic tumor subtypes, likelihood-ratio tests were used comparing the model fit for logistic regression models with and without corresponding interaction terms. For all risk models, the discrimination between cases and control subjects was described using ROC curves.

For multimarker discrimination models, the statistical fit of nested models was compared with likelihood-ratio tests, and bootstrapping methods were used to correct for model overfitting and overoptimism in the estimation of discrimination capacity. In addition, measures of continuous net reclassification improvement were calculated, which represents the percent of case and control subjects correctly reclassified as a result of the added marker (17). Analyses were conducted in SAS (version 9.3, SAS Institute).

Results

Baseline characteristics of ovarian cancer case patients by tumor characteristics are presented in Table 1. Of the 810 case patients examined in this study, 752 (93%) had the ovary classified as primary tumor site, whereas in 33 (4%) the primary site was the fallopian tube and in 25 patients (3%) it was the peritoneum. More than half of the tumors (55%) were of serous histology ($n =$ 445), 12% endometrioid ($n = 96$), 7% mucinous ($n = 58$), 5%

Published OnlineFirst April 8, 2016; DOI: 10.1158/1078-0432.CCR-16-0316

Table 1. Characteristics of cases and controls in the EPIC cohort [median (min–max) or *n* (%)]

	Cases ($n = 810$)	Controls ($n = 1{,}939$)	*P*
Age at blood draw, years	56.4 (29.9–80.7)	56.7 (30.1–79.3)	
Age at blood draw, years			
<50	166 (20%)	405 (21%)	
50–55	187 (23%)	430 (22%)	
55–60	183 (23%)	430 (22%)	
60–65	185 (23%)	445 (23%)	
≥65	89 (11%)	229 (12%)	
Menopausal status			
Pre	132 (16%)	329 (17%)	
Peri[a]	118 (15%)	274 (14%)	
Post	560 (69%)	1 336 (69%)	
BMI	25.1 (17.2–45.4)	25.0 (14.9–50.6)	0.03
Smoking[b]			0.12
Never	432 (55%)	1 103 (58%)	
Former	185 (23%)	435 (23%)	
Current	177 (22%)	368 (19%)	
Parous[b]	617 (83%)	1 585 (89%)	<0.0001
Number of children[b,c]			0.31
1	114 (19%)	277 (18%)	
2	297 (49%)	733 (48%)	
>2	191 (32%)	531 (34%)	
Hysterectomy[b]	69 (11%)	176 (11%)	0.35
Case characteristics			
Age at diagnosis	62.7 (30.6–86.5)	—	
Lag time (years)	6.1 (0–16.0)	—	
Cancer site			
Ovary	752 (93%)	—	
Fallopian tube	33 (4%)	—	
Peritoneum	25 (3%)	—	
Histology			
Serous	445 (55%)	—	
Mucinous	58 (7%)	—	
Endometrioid	96 (12%)	—	
Clear cell	37 (4%)	—	
NOS	136 (17%)	—	
Other	38 (5%)	—	
Cancer grade[b]			
Well differentiated	45 (9%)	—	
Moderately differentiated	164 (35%)	—	
Poorly differentiated/undifferentiated	264 (56%)	—	
Disease spread[b]			
Localized (stage I)	115 (16%)	—	
Regional (stage II)	128 (18%)	—	
Metastatic (stage III/IV)	469 (66%)	—	
Marker[d]			
CA125 (U/mL)	27.6 (26.2–29.1)	20.2 (19.5–20.8)	<0.0001
HE4 (pmol/L)	29.1 (26.9–31.6)	18.9 (18.1–19.7)	<0.0001
CA72.4 (U/mL)[e]	5.5 (4.4–6.9)	2.4 (2.0–2.9)	0.004
CA15.3 (mIU/mL)	624.3 (600.3–649.2)	600.6 (585.6–616.0)	0.10

[a]Defined as women between the ages of 42 and 55 years who have missing or incomplete questionnaire data, reported irregular menstrual cycles in the past 12 months, or had a prior hysterectomy without oophorectomy.

[b]Data were missing on smoking for 16 cases and 33 controls, on parity for 64 cases and 151 controls, on number of children for 79 cases and 195 controls, on hysterectomy for 178 cases and 436 controls, on cancer grade for 337 cases, on the dualistic model for 385 cases, and on disease spread for 98 cases.

[c]Among parous women ($n = 2{,}202$).

[d]Presented as geometric mean (5th–95th percentile); Data were missing on CA125 for 3 cases and 12 controls, on HE4 for 2 controls, on CA72.4 for 2 controls, and on CA15.3 for 6 cases and 19 controls.

[e]Based on 67 (34%) cases and 109 (15%) controls with CA72.4 above the detection limit (CA72.4 was measured in 197 cases and 725 controls).

other (malignant epithelial neoplasms, carcinoma, malignant mixed Müllerian or malignant Brenner tumors; $n = 38$), 4% clear cell ($n = 37$), and 17% not otherwise specified (NOS; $n = 136$). Of the 712 case subjects with information about tumor stage at diagnosis, 115 were classified as stage I, 128 as stage II, and 469 as stage III and higher (stage III/IV). Compared with case patients diagnosed at stage II and higher, there was a relative over-representation of mucinous, clear cell, and endometrioid tumors among the stage I patients, whereas serous tumors were predominantly represented among the patients with cancer in stage II and higher (see Supplementary Table S1). Overall, the median age at cancer diagnosis was 62.7 years (range: 30.6–86.5 years), and varied according to the histologic subtypes (Supplementary Table S1).

Visual inspection of LOESS curves suggests that none of the biomarkers were increased over normal (control) values earlier

Published OnlineFirst April 8, 2016; DOI: 10.1158/1078-0432.CCR-16-0316

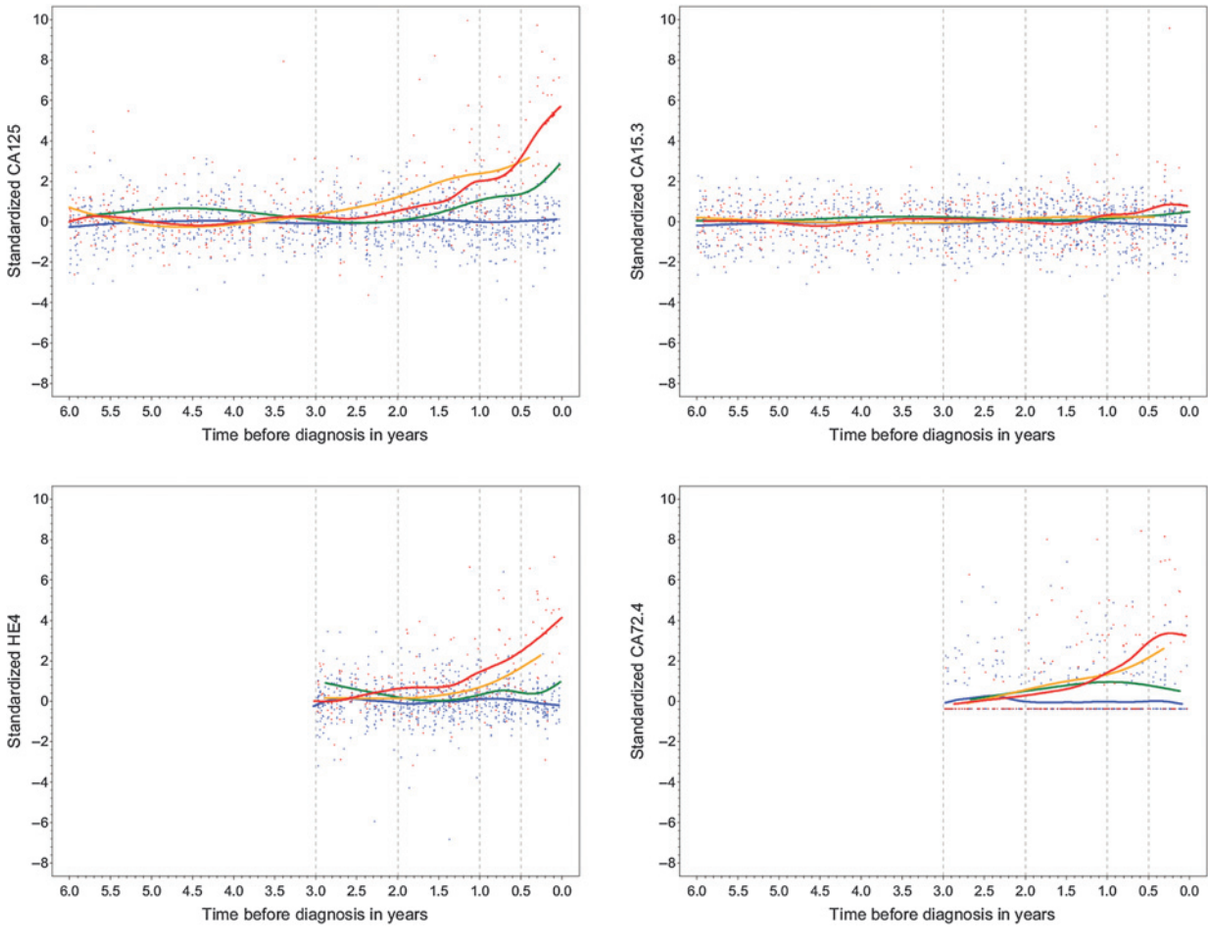

Figure 1.

LOESS curves over time prior to diagnosis. The blue line represents controls, the green line stage I cases, the yellow line stage II cases, and the red line stage III/IV cases. For CA125 and CA15.3, data are shown only for lag-times between blood donation and cancer diagnosis up to 6 years.

than about 2 years prior to diagnosis, and more than 6 months prior to diagnosis increased levels were discernable only for case patients who were diagnosed with ovarian cancer at advanced stage (stage II or III; Fig. 1).

For the predefined variable lag times between blood donation and date of diagnosis, the ability of the early detection markers to discriminate between case patients and control subjects is indicated by C-statistics and estimated sensitivities at specificity cut-off points of 95% (SE95) and 98% (SE98; Table 2). In addition, ROC curves are shown in Fig. 2. For blood samples taken $\leq$6 months prior to diagnosis, the highest C-statistic was observed for CA125 (C = 0.92), followed by HE4 (C = 0.84), CA72.4 (C = 0.77), and CA15.3 (C = 0.73). Correspondingly, within the first 6 months, values for SE95 and SE98 were fairly high for CA125 (0.81 and 0.77, respectively) and HE4 (0.67 and 0.59), and modest for CA72.4 (0.56 and 0.37) and CA15.3 (0.31 and 0.23). For all markers, the capacity to discriminate between future case patients and noncases dropped rapidly with increasing time lags between blood donation and tumor diagnosis (Table 2; Fig. 2). For example, for a time lag between 1 and 2 years, C-statistic values were 0.72 for CA125, 0.65 for HE4, 0.61 for CA72.4, and 0.52 for CA15.3. At time lags between 3 and 6 years, the two markers that were tested for longer-term prediction of ovarian cancer, CA125 and CA15.3 had C-statistics of only 0.55 and 0.53, respectively (Table 2).

Within the first 12 months after blood donation, for all markers except CA15.3 the ability to predict future cancer diagnosis was clearly stronger for advanced tumors (stage II and III/IV) and relatively weak for stage I tumors (Supplementary Fig. S1), and this heterogeneity was statistically significant for CA125 and HE4 ($P_{het} < 0.05$; Table 2). Regarding tumor histology, CA125, HE4, and CA72.4 showed fairly strong discrimination of serous ovarian cancer patients from their matched controls, especially within short lag-times after blood donation (Supplementary Table S2); for the other histologic subtypes, the numbers of patients were too small to obtain reliable estimates.

Among the control subjects, no meaningful correlations between markers were observed ($r = -0.15$–0.18). Among the case subjects only, and especially among those with lag-times since blood donation below 1 or 2 years, moderately strong correlations were observed between CA125, HE4, and CA72.4 (e.g., within 1-year's lag-time: $r = 0.23$–0.74), whereas CA15.3 showed somewhat weaker associations with the other

Published OnlineFirst April 8, 2016; DOI: 10.1158/1078-0432.CCR-16-0316

Table 2. Sensitivity at 95% and 98% specificity and C-statistics by time between blood draw and diagnosis

	# of Sets	Sensitivity at 95% specificity	Sensitivity at 98% specificity	C-statistic (95% CI)	P_{het}[a]
CA125		Cut-off point: 56.64 U/mL	Cut-off point: 77.92 U/mL		
Overall	807	0.14 (0.11–0.18)	0.10 (0.07–0.13)	0.58 (0.56–0.60)	
≤6 months	26	0.81 (0.61–0.92)	0.77 (0.57–0.89)	0.92 (0.86–0.98)	
≤12 months	61	0.59 (0.46–0.71)	0.52 (0.39–0.66)	0.82 (0.76–0.88)	
>1–2 years	75	0.27 (0.18–0.38)	0.20 (0.12–0.31)	0.72 (0.65–0.78)	
>2–3 years	58	0.10 (0.05–0.22)	0.03 (0.01–0.13)	0.56 (0.48–0.64)	
>3–6 years	200	0.08 (0.05–0.13)	0.06 (0.03–0.11)	0.55 (0.50–0.60)	
>6 years	413	0.09 (0.06–0.13)	0.04 (0.02–0.07)	0.54 (0.51–0.57)	
Stage I	115	0.16 (0.1–0.24)	0.10 (0.06–0.18)	0.58 (0.51–0.64)	
≤6 months	7	0.43 (0.14–0.77)	0.29 (0.07–0.68)	0.69 (0.47–0.91)	
≤12 months	16	0.38 (0.18–0.63)	0.31 (0.13–0.57)	0.65 (0.49–0.80)	
>1–2 years	7	0.14 (0.02–0.58)	0.14 (0.02–0.58)	0.55 (0.31–0.80)	
>2–3 years	9	0 (0–0.34)	0 (0–0.34)	0.52 (0.30–0.74)	
>3–6 years	32	0.13 (0.05–0.29)	0.09 (0.03–0.26)	0.60 (0.48–0.72)	
>6 years	51	0.14 (0.07–0.26)	0.06 (0.02–0.17)	0.57 (0.47–0.66)	
Stage II	128	0.14 (0.09–0.22)	0.09 (0.05–0.16)	0.61 (0.55–0.67)	
≤6 months	1				
≤12 months	8	0.75 (0.38–0.94)	0.50 (0.20–0.80)	0.91 (0.79–1.02)	
>1–2 years	10	0.30 (0.10–0.63)	0.30 (0.10–0.63)	0.75 (0.58–0.93)	
>2–3 years	11	0.09 (0.01–0.44)	0 (0–0.28)	0.66 (0.48–0.84)	
>3–6 years	37	0.05 (0.01–0.19)	0.05 (0.01–0.20)	0.53 (0.41–0.64)	
>6 years	62	0.10 (0.04–0.20)	0.05 (0.02–0.14)	0.57 (0.48–0.66)	
Stage III/IV	467	0.14 (0.10–0.18)	0.09 (0.06–0.13)	0.57 (0.54–0.60)	0.26
≤6 months	15	0.93 (0.65–0.99)	0.93 (0.64–0.99)	0.96 (0.92–1.01)	0.09
≤12 months	31	0.68 (0.49–0.82)	0.65 (0.46–0.80)	0.87 (0.80–0.94)	0.045
>1–2 years	50	0.26 (0.15–0.40)	0.16 (0.08–0.30)	0.72 (0.64–0.80)	0.77
>2–3 years	36	0.14 (0.06–0.30)	0.06 (0.01–0.20)	0.57 (0.46–0.67)	0.20
>3–6 years	106	0.07 (0.03–0.13)	0.04 (0.01–0.10)	0.52 (0.45–0.59)	0.68
>6 years	244	0.07 (0.05–0.12)	0.04 (0.02–0.07)	0.53 (0.48–0.57)	0.34
HE4		Cut-off point: 39.41 pmol/L	Cut-off point: 54.01 pmol/L		
Overall	197	0.24 (0.17–0.32)	0.18 (0.11–0.27)	0.67 (0.63–0.71)	
≤6 months	27	0.67 (0.46–0.82)	0.59 (0.39–0.77)	0.84 (0.76–0.92)	
≤12 months	62	0.48 (0.35–0.62)	0.39 (0.25–0.54)	0.79 (0.73–0.85)	
>1–2 years	76	0.17 (0.10–0.28)	0.14 (0.07–0.26)	0.65 (0.58–0.72)	
>2–3 years	59	0.07 (0.02–0.17)	0 (0–0.06)	0.56 (0.48–0.65)	
Stage I	32	0.13 (0.05–0.30)	0.09 (0.03–0.27)	0.59 (0.48–0.70)	
≤6 months	7	0.14 (0.02–0.59)	0.14 (0.02–0.59)	0.51 (0.27–0.76)	
≤12 months	16	0.19 (0.06–0.45)	0.19 (0.06–0.46)	0.57 (0.41–0.73)	
>1–2 years	7	0 (0–0.41)	0 (0–0.41)	0.52 (0.28–0.77)	
>2–3 years	9	0.11 (0.02–0.51)	0 (0–0.34)	0.71 (0.52–0.90)	
Stage II	29	0.10 (0.03–0.28)	0.07 (0.02–0.25)	0.68 (0.57–0.79)	
≤6 months	1				
≤12 months	8	0.25 (0.06–0.63)	0.13 (0.02–0.55)	0.72 (0.52–0.92)	
>1–2 years	10	0.10 (0.01–0.47)	0.10 (0.01–0.48)	0.66 (0.46–0.85)	
>2–3 years	11	0 (0–0.28)	0 (0–0.28)	0.67 (0.49–0.85)	
Stage III/IV	119	0.31 (0.22–0.42)	0.23 (0.14–0.34)	0.69 (0.64–0.75)	0.34
≤6 months	16	0.88 (0.61–0.97)	0.75 (0.48–0.91)	0.97 (0.93–1.01)	0.007
≤12 months	32	0.72 (0.53–0.85)	0.56 (0.37–0.74)	0.92 (0.87–0.97)	0.01
>1–2 years	50	0.22 (0.12–0.37)	0.18 (0.09–0.33)	0.67 (0.58–0.75)	0.07
>2–3 years	37	0.08 (0.03–0.23)	0 (0–0.09)	0.53 (0.42–0.63)	0.09
CA72.4		Cut-off point: 2.46 U/mL	Cut-off point: 5.96 U/mL		
Overall	197	0.23 (0.17–0.32)	0.13 (0.08–0.21)	0.61 (0.56–0.65)	
≤6 months	27	0.56 (0.36–0.74)	0.37 (0.20–0.58)	0.77 (0.68–0.87)	
≤12 months	62	0.45 (0.32–0.59)	0.26 (0.15–0.40)	0.72 (0.65–0.79)	
>1–2 years	76	0.20 (0.12–0.32)	0.11 (0.05–0.21)	0.61 (0.54–0.68)	
>2–3 years	59	0.05 (0.02–0.15)	0.02 (0–0.12)	0.53 (0.44–0.61)	
Stage I	32	0.22 (0.10–0.40)	0.06 (0.01–0.23)	0.58 (0.47–0.69)	
≤6 months	7	0.14 (0.02–0.59)	0 (0–0.41)	0.61 (0.37–0.85)	
≤12 months	16	0.25 (0.09–0.52)	0.13 (0.03–0.4)	0.61 (0.46–0.77)	
>1–2 years	7	0.14 (0.02–0.59)	0 (0–0.41)	0.57 (0.32–0.81)	
>2–3 years	9	0.22 (0.05–0.58)	0 (0–0.34)	0.52 (0.30–0.74)	
Stage II	29	0.21 (0.09–0.40)	0.14 (0.05–0.33)	0.59 (0.47–0.71)	
≤6 months	1				
≤12 months	8	0.38 (0.12–0.72)	0.25 (0.06–0.63)	0.69 (0.48–0.90)	
>1–2 years	10	0.30 (0.10–0.63)	0.20 (0.05–0.55)	0.67 (0.48–0.86)	
>2–3 years	11	0 (0–0.28)	0 (0–0.28)	0.56 (0.37–0.76)	

(Continued on the following page)

Published OnlineFirst April 8, 2016; DOI: 10.1158/1078-0432.CCR-16-0316

Table 2. Sensitivity at 95% and 98% specificity and C-statistics by time between blood draw and diagnosis (Cont'd)

	# of Sets	Sensitivity at 95% specificity	Sensitivity at 98% specificity	C-statistic (95% CI)	P_{het}[a]
Stage III/IV	119	0.24 (0.16–0.34)	0.14 (0.08–0.24)	0.61 (0.56–0.67)	0.40
≤6 months	16	0.69 (0.43–0.87)	0.50 (0.26–0.74)	0.80 (0.68–0.92)	0.29
≤12 months	32	0.56 (0.38–0.73)	0.34 (0.19–0.54)	0.76 (0.67–0.85)	0.80
>1–2 years	50	0.18 (0.09–0.32)	0.10 (0.04–0.23)	0.61 (0.52–0.70)	0.35
>2–3 years	37	0.03 (0–0.17)	0.03 (0–0.18)	0.52 (0.41–0.62)	0.39
CA15.3		Cut-off point: 1,372 mIU/mL	Cut-off point: 1,610 mIU/mL		
Overall	804	0.07 (0.05–0.09)	0.04 (0.02–0.06)	0.51 (0.49–0.54)	
≤6 months	26	0.31 (0.16–0.51)	0.23 (0.11–0.43)	0.73 (0.62–0.84)	
≤12 months	61	0.16 (0.09–0.28)	0.13 (0.06–0.25)	0.58 (0.50–0.66)	
>1–2 years	74	0.14 (0.07–0.24)	0.08 (0.04–0.17)	0.52 (0.45–0.60)	
>2–3 years	58	0 (0–0.06)	0 (0–0.06)	0.54 (0.45–0.62)	
>3–6 years	200	0.04 (0.02–0.08)	0.02 (0–0.05)	0.53 (0.48–0.58)	
>6 years	411	0.06 (0.04–0.09)	0.03 (0.02–0.05)	0.51 (0.48–0.55)	
Stage I	114	0.07 (0.03–0.14)	0.04 (0.02–0.11)	0.52 (0.46–0.59)	
≤6 months	7	0.29 (0.07–0.68)	0 (0–0.41)	0.78 (0.59–0.97)	
≤12 months	16	0.19 (0.06–0.45)	0.06 (0.01–0.34)	0.66 (0.51–0.81)	
>1–2 years	6	0.17 (0.02–0.63)	0.17 (0.02–0.63)	0.67 (0.42–0.91)	
>2–3 years	9	0 (0–0.34)	0 (0–0.34)	0.56 (0.34–0.77)	
>3–6 years	32	0.03 (0–0.19)	0 (0–0.11)	0.61 (0.49–0.73)	
>6 years	51	0.06 (0.02–0.17)	0.06 (0.02–0.17)	0.55 (0.46–0.65)	
Stage II	128	0.06 (0.03–0.12)	0.02 (0–0.06)	0.54 (0.49–0.6)	
≤6 months	1				
≤12 months	8	0 (0–0.37)	0 (0–0.37)	0.52 (0.29–0.75)	
>1–2 years	10	0.30 (0.10–0.63)	0.1 (0.01–0.47)	0.59 (0.39–0.79)	
>2–3 years	11	0 (0–0.28)	0 (0–0.28)	0.50 (0.30–0.69)	
>3–6 years	37	0 (0–0.09)	0 (0–0.09)	0.56 (0.44–0.67)	
>6 years	62	0.08 (0.03–0.18)	0.02 (0–0.11)	0.55 (0.46–0.64)	
Stage III/IV	465	0.06 (0.04–0.09)	0.04 (0.02–0.07)	0.50 (0.46–0.53)	0.37
≤6 months	15	0.27 (0.10–0.54)	0.27 (0.10–0.54)	0.71 (0.57–0.86)	0.51
≤12 months	31	0.16 (0.07–0.34)	0.16 (0.07–0.34)	0.55 (0.43–0.66)	0.64
>1–2 years	50	0.10 (0.04–0.22)	0.08 (0.03–0.20)	0.50 (0.41–0.59)	0.16
>2–3 years	36	0 (0–0.1)	0 (0–0.1)	0.56 (0.45–0.67)	0.32
>3–6 years	106	0.04 (0.01–0.10)	0.02 (0–0.07)	0.50 (0.44–0.57)	0.16
>6 years	242	0.06 (0.04–0.10)	0.03 (0.01–0.06)	0.53 (0.48–0.57)	0.22

[a]Heterogeneity of discrimination capacity by tumor stage was examined with likelihood-ratio tests comparing the model fit for logistic regression models with and without corresponding interaction terms.

markers (within 1-year's lag-time: $r = 0.11–0.24$; Supplementary Table S3).

In a stepwise forward selection strategy, focusing on variable lag-time strata within the first 3 years after blood donation, the overall model fit for a logistic risk model improved statistically significantly with successive additions of CA125, HE4, CA72.4, and CA15.3 as prediagnostic predictors of future ovarian cancer diagnosis, although the statistical significance for CA15.3 was lowest and largely restricted to lag-times less than 6 months (Table 3). However, the overall improvements in the overall discrimination, assessed by C-statistic (Table 3) or NRI (Supplementary Table S4), were small compared with a model based on any of the markers CA125, HE4, or CA72.4 alone.

Finally, we examined whether the overall discrimination between case patients and control subjects could be improved by combining the biomarkers with an epidemiologic stratification algorithm that was developed previously using the full EPIC cohort data (16). Because some of the key epidemiologic risk variables included in the algorithm (age, menopausal status, use of OCs or HRT), overlapped with some of the matching factors for the present nested case–control study, the risk model showed a lower discrimination (C = 0.56) in our case–control set as compared with our previous full cohort analysis (corrected for over-optimism, C = 0.64). For lag-times below 2 years, combining the risk model with the biomarkers did not improve overall discrimination as compared with each of the biomarkers alone (results not shown). In contrast, for lag times greater than 3 years, the longer-term prediction of future ovarian cancer diagnosis was moderately but significantly improved when CA125 was added to the model (C = 0.57 vs. C = 0.55), whereas adding CA15.3 showed no improvement.

Discussion

In our evaluation of four potential ovarian cancer screening biomarkers measured in prospectively collected samples from women with ovarian cancer and matched controls in the EPIC cohort, we observed the best sensitivity and specificity for CA125, followed by HE4, CA72.4, and finally CA15.3. The ability of these biomarkers to distinguish cases from controls declined with increasing time between blood draw and diagnosis, as well as with earlier stage at diagnosis. These observations suggest that, generally, these markers are best at identifying advanced disease close to diagnosis, but their ability to detect early disease that is amenable to interventions that can improve survival may be limited. Addition of a previously established risk prediction model did not improve the performance of markers in women who went on to develop clinically manifest ovarian cancer less than three years in advance of diagnosis. In contrast, adding CA125 (but not CA15.3) to the risk prediction model did slightly improve the longer-term prediction of ovarian cancer occurrence over a time interval of about 3 to 6 years after blood donation.

Published OnlineFirst April 8, 2016; DOI: 10.1158/1078-0432.CCR-16-0316

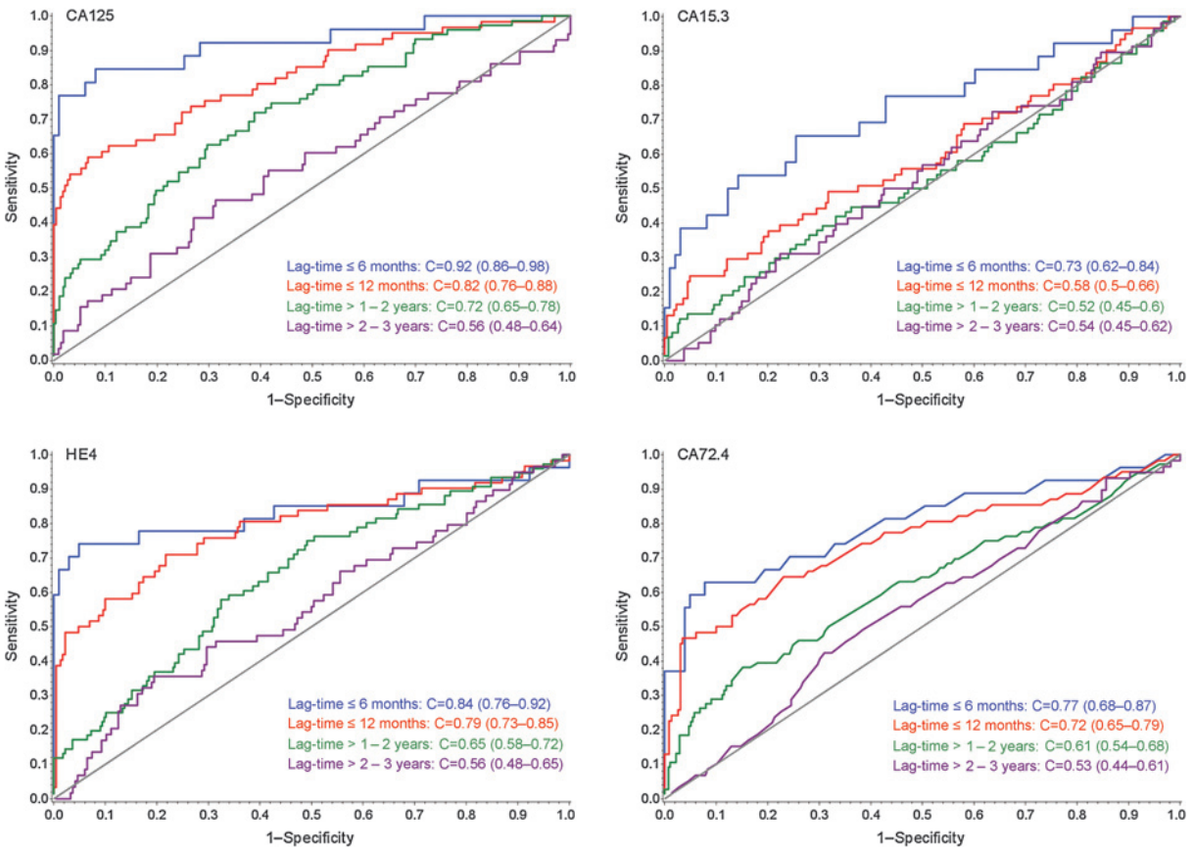

Figure 2.
ROC curves and C-statistics for diagnosis ≤6 months, ≤12 months, 1–2 years, and 2–3 years after blood collection.

Our results are consistent with those from large randomized trials and prospective assessments of these markers in other populations. All four of these biomarkers were included in an ancillary study of 49 biomarkers previously evaluated in phase II studies, and all four were also among the best 35 that were subsequently examined within the prospective PLCO cohort (18). Results from the PLCO study were similar to what we observed in EPIC, with the best performance in cases diagnosed 6 months or less after blood draw (depending on the time between blood draw and diagnosis), C-statistics ranged from 0.83 to 0.96 for CA125, 0.78 to 0.88 for HE4, 0.80 for CA72.4, and 0.72 for CA15.3. As in our current study, the discriminatory ability of these markers in the PLCO cohort declined rapidly for samples collected more than 6 months prior to diagnosis. An

Table 3. Stepwise forward selection of early detection markers in combined prediagnostic models for variable lag-times since blood donation[a]

		Model 1: CA125		Model 2: CA125 + HE4		Model 3: CA125 + HE4 + CA72.4		Model 4: CA125 + HE4 + CA72.4 + CA15.3	
	# Sets	C-Statistic (95% CI)	P^b	C-Statistic (95% CI)	P^c	C-Statistic (95% CI)	P^d	C-Statistic (95% CI)	P^e
All women	197	0.70 (0.66–0.74)	<0.001	0.71 (0.67–0.75)	<0.001	0.71 (0.67–0.75)	0.002	0.71 (0.67–0.75)	0.05
Corrected C[f]		—		*0.71*		*0.71*		*0.71*	
≤6 months	27	0.92 (0.86–0.98)	<0.001	0.89 (0.83–0.96)	0.58	0.90 (0.83–0.96)	0.36	0.92 (0.86–0.97)	0.06
Corrected C[f]		—		*0.89*		*0.90*		*0.92*	
≤1 year	62	0.82 (0.76–0.88)	<0.001	0.82 (0.77–0.88)	0.005	0.84 (0.78–0.89)	0.002	0.84 (0.78–0.89)	0.55
Corrected C[f]		—		*0.81*		*0.83*		*0.83*	
1–2 years	76	0.72 (0.65–0.78)	<0.001	0.73 (0.66–0.79)	0.004	0.73 (0.67–0.80)	0.03	0.73 (0.67–0.80)	0.08
Corrected C[f]		—		*0.72*		*0.72*		*0.72*	
2–3 years	59	0.56 (0.48–0.64)	0.42	0.58 (0.50–0.67)	0.18	0.58 (0.50–0.66)	0.57	0.58 (0.50–0.66)	0.35
Corrected C[f]		—		*0.53*		*0.53*		*0.53*	

[a]Conditional logistic regression model with $\log_2$-transformed markers as continuous variables.
[b]Likelihood ratio test, comparing Model 1 with empty model.
[c]Likelihood ratio test, comparing Model 2 with Model 1.
[d]Likelihood ratio test, comparing Model 3 with Model 2.
[e]Likelihood ratio test, comparing Model 4 with Model 3.
[f]C-Statistic in multivariate models corrected for over optimism obtained by bootstrap sampling.

Published OnlineFirst April 8, 2016; DOI: 10.1158/1078-0432.CCR-16-0316

important difference between PLCO and our study, however, is that PLCO participants had been annually screened for ovarian cancer by CA125 plus ultrasonography, thereby reducing the occurrence of further ovarian cancer diagnoses over time periods longer than 12 months. In another prospective study, Anderson and colleagues measured CA125 and HE4 among other markers in a total of 34 incident cases of ovarian cancer who had provided serial blood samples up to 18 years prior to diagnosis and noted a similar decline in marker performance over time with an AUC at <2 years of 0.74 and 0.71 and from 2 to 4 years of 0.68 and 0.67 for CA125 and HE4, respectively (10).

The ultimate goal of screening is to identify cancer at a stage at which medical intervention has the highest chances of providing a cure or prolonging survival. None of the four markers that we tested showed a clear capacity for predicting disease that was diagnosed in stage I, even for follow-up times of less than 6 months. This finding, however, may have been confounded by tumor histology, as among patients diagnosed with stage I disease, there was an over-representation of mucinous and clear cell tumors, which are generally slowly growing tumor subtypes and diagnosed at earlier stages. In contrast, within less than 12 months of prospective follow-up especially, CA125 and HE4 showed substantial discriminatory capacity for tumors that had been subsequently diagnosed at a more advanced stage, and the discrimination of these markers was also significantly stronger for the more aggressive type II tumors as compared with type I tumors. A major limitation of prospective studies such as EPIC and PLCO is that they do not inform about a patient's tumor stage at the time of blood donation; hence, no direct information is available whether those women whose tumor might have been diagnosed 6–12 months before clinical diagnosis through CA125 or HE4 screening would have actually shown a sufficient shift towards an earlier tumor stage to allow speculations about significant improvements in survival.

Despite having the best performance among various candidate markers considered for ovarian cancer screening in various studies (10, 18, 19), annual CA125 measurement (combined with transvaginal ultrasound) in the PLCO randomized trial showed no mortality benefit (7). This lack of benefit is likely related to insufficient sensitivity of a single CA125 measurement for detecting the more aggressive forms of ovarian cancer in an early stage of disease, as suggested by the findings in both PLCO and our study. However, several studies have shown that the use of serial measurements over time can improve the diagnostic performance for ovarian cancer detection (20–23). In the world's largest ongoing, randomized screening trial for ovarian cancer, the United Kingdom Collaborative Trial on Ovarian Cancer Screening (UKCTOCS), consideration of change in CA125 over time using the ROCA algorithm improved marker performance from C-statistic = 0.87 for a single CA125 to C = 0.92 and doubled the number of screen-detected invasive epithelial ovarian cancers compared with CA125 screening with a fixed cutoff (22). Furthermore, recent mortality results from the UKCTOCs revealed a 15% reduction in mortality for women screened using the ROCA algorithm among incident cases ($P = 0.02$; ref. 8).

Results from various phase II and other clinical studies have suggested that combinations of multiple biomarkers may be better at distinguishing malignant from benign tumors than CA125 alone. For example, improved discrimination has been documented for the combination of CA125 and HE4, the two strongest discriminating biomarkers in our analyses, as compared either marker alone (12, 24–31). While our analyses confirm that biomarker combinations improve prediction of future ovarian cancer diagnosis, the absolute gain in classification appeared to be small in our data, and a similar observation was made in the PLCO cohort (19). Thus, the addition of biomarkers can improve the discriminatory ability of CA125 but current biomarkers may not improve performance to the degree required for population screening.

Analyses in the PLCO study as well as ours, show substantial discrepancy between the often promising findings from phase II discovery studies based on clinical case–control comparisons and their lack of replication in prospective evaluations based on prediagnostic blood samples. This observation has triggered recommendations that greater care should be taken in selecting the appropriate sample set for screening biomarker discovery. In particular, it was recommended that prospective cohort studies should be used for new biomarker discovery rather than simply validation of known candidate biomarkers (32–34). One advantage of such an approach would be that it ensures rigorous internal validity for the evaluation of systematic differences between case and control subjects. Another possible advantage of the prospective design is that by focusing on blood samples collected months prior to cancer diagnosis, one would avoid a bias toward markers exclusively associated with advanced disease (32). While attractive from a methodologic perspective, however, the use of prospective cohorts for biomarker discovery may have several limitations in practice. In our study, among 366,521 women mostly aged 35–70 years at blood donation, there was an annual incidence of about 35 ovarian cancer cases. Thus, assuming an early detection time window of 6–18 months prior to diagnosis (excluding the first 6 months of follow-up to reduce the presence of advanced disease), studies for marker discovery would be based on a very limited, yet etiologically diverse sample set. This basic observation illustrates that even very large prospective cohorts may not have a sufficient number of cases for biomarker discovery studies focusing on early-stage disease. Moreover, as already noted, the tumor grade and stage at the time of blood sampling would remain unknown. Therefore, uncertainty will remain as to whether those patients whose tumor would have had elevated biomarkers 6–18 months prior to diagnosis (and hence potentially detected) would actually benefit from detection at that timepoint. In light of these limitations, we believe that, as a complement to prospective cohort studies, bio-banking initiatives in large clinical networks will remain needed for the collection of samples especially from well-characterized early-stage patients to allow large-scale comparisons with samples from cancer-free individuals.

In summary, CA125 and HE4 continue to hold potential for ovarian cancer screening but lack sensitivity and specificity needed to detect early-stage disease. New biorepositories of early-stage disease and matched controls are needed to identify novel markers that focus on the disease timepoint where intervention can make the biggest improvement in mortality and morbidity.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: K.L. Terry, A. Hüsing, T. Johnson, K. Overvad, A. Tjønneland, G. Severi, H. Boeing, H.B. Bueno-de-Mesquita, E. Weiderpass, M.-J. Sanchez, E. Ardanaz, C. Navarro, D.W. Cramer, R. Kaaks

Development of methodology: K.L. Terry, T. Johnson, P.H. Peeters, E. Weiderpass, E. Ardanaz
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): H.S. Yamamoto, K. Overvad, A. Tjønneland, M.-C. Boutron-Ruault, G. Severi, H. Boeing, V. Benetou, P. Lagiou, A. Trichopoulou, V. Krogh, S. Panico, H.B. Bueno-de-Mesquita, P.H. Peeters, E. Weiderpass, M.-J. Sanchez, E. Ardanaz, N. Etxezarreta, C. Navarro, A. Idahl, E. Lundin, K. Jirström, J. Manjer, N.J. Wareham, K.-T. Khaw, R.C. Travis, M.J. Gunter, R. Kaaks
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): K.L. Terry, H. Schock, R.T. Fortner, A. Hüsing, H.S. Yamamoto, A.F. Vitonis, V. Benetou, P.H. Peeters, I.T. Gram, E. Weiderpass, E. Ardanaz, R. Kaaks
Writing, review, and/or revision of the manuscript: K.L. Terry, H. Schock, R.T. Fortner, A. Hüsing, H.S. Yamamoto, K. Overvad, A. Tjønneland, M.-C. Boutron-Ruault, S. Mesrine, G. Severi, L. Dossus, S. Rinaldi, H. Boeing, V. Benetou, P. Lagiou, A. Trichopoulou, V. Krogh, E. Kuhn, S. Panico, H.B. Bueno-de-Mesquita, N.C. Onland-Moret, P.H. Peeters, I.T. Gram, E. Weiderpass, E.J. Duell, M.-J. Sanchez, E. Ardanaz, N. Etxezarreta, C. Navarro, A. Idahl, E. Lundin, J. Manjer, N.J. Wareham, K.-T. Khaw, K.S. Byrne, R.C. Travis, M.J. Gunter, M.A. Merritt, E. Riboli, D.W. Cramer, R. Kaaks
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): H. Schock, H.S. Yamamoto, A.F. Vitonis, T. Johnson, K. Overvad, H.B. Bueno-de-Mesquita, P.H. Peeters, E. Weiderpass, E. Ardanaz, E. Lundin, J. Manjer, K.-T. Khaw, R. Kaaks
Study supervision: E. Weiderpass, E. Ardanaz, R. Kaaks
Other (Development and performance of laboratory assays): R.N. Fichorova
Other (Principal Investigator on the grant, co-senior author): D.W. Cramer

Acknowledgments

The authors thank Fujirebio Diagnostics for supplying the reagents for the HE4 and CA72.4 assays.

Grant Support

This work was supported by the NIH (R01CA158119). The coordination of the European Prospective Investigation into Cancer and Nutrition is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM; France); Deutsche Krebshilfe, Deutsches Krebsforschungszentrum and Federal Ministry of Education and Research (Germany); Hellenic Health Foundation (Greece) and the Stavros Niarchos Foundation; Italian Association for Research on Cancer (AIRC) and National Research Council (Italy); Dutch Ministry of Public Health, Welfare, and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (the Netherlands); ERC-2009-AdG 232997 and Nordforsk, Nordic Centre of Excellence programme on Food, Nutrition, and Health (Norway); Health Research Fund (FIS), Regional Governments of Andalucía, Asturias, Basque Country, Murcia (no. 6236) and Navarra, ISCIII RETIC (RD06/0020; Spain); Swedish Cancer Society, Swedish Scientific Council, and Regional Government of Skåne and Västerbotten (Sweden); Cancer Research UK (grant no. 14136, to K.-T. Khaw; grant no. C570/A16491, to R.C. Travis), Medical Research Council (United Kingdom; grant no. G1000143).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 5, 2016; revised March 7, 2016; accepted March 21, 2016; published OnlineFirst April 8, 2016.

References

1. American Cancer Society. Cancer Facts & Figures. Atlanta, GA: American Cancer Society; 2015.
2. Terry KL, Sluss PM, Skates SJ, Mok SC, Ye B, Vitonis AF, et al. Blood and urine markers for ovarian cancer: a comprehensive review. Dis Markers 2004;20:53–70.
3. Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, et al. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst 2001;93:1054–61.
4. U.S. Food and Drug Administration. 510(K) Substantial equivalence determination decision summary. Available from: http://www.accessdata.fda.gov/cdrh_docs/reviews/K081754.pdf.
5. U.S. Food and Drug Administration. HE4 EIA. Available from: http://www.fda.gov/ohrms/dockets/ac/08/briefing/220028-4403b1–02-Proposed%20HE4%20EIA%20Package%20Insert%20%20.pdf.
6. U.S. Food and Drug Administration. 510 (k) Substantial equivalence determination decision summary. Available from: http://www.accessdata.fda.gov/cdrh_docs/reviews/K042731.pdf.
7. Buys SS, Partridge E, Black A, Johnson CC, Lamerato L, Isaacs C, et al. Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. JAMA 2011;305:2295–303.
8. Jacobs IJ, Menon U, Ryan A, Gentry-Maharaj A, Burnell M, Kalsi JK, et al. Ovarian cancer screening and mortalityin the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. Lancet 2016;387:945–56.
9. Kobayashi H, Yamada Y, Sado T, Sakata M, Yoshida S, Kawaguchi R, et al. A randomized study of screening for ovarian cancer: a multicenter study in Japan. Int J Gynecol Cancer 2008;18:414–20.
10. Anderson GL, McIntosh M, Wu L, Barnett M, Goodman G, Thorpe JD, et al. Assessing lead time of selected ovarian cancer biomarkers: a nested case-control study. J Natl Cancer Inst 2010;102:26–38.
11. Urban N, Hawley S, Janes H, Karlan BY, Berg CD, Drescher CW, et al. Identifying post-menopausal women at elevated risk for epithelial ovarian cancer. Gynecol Oncol 2015;139:253–60.
12. Urban N, Thorpe JD, Bergan LA, Forrest RM, Kampani AV, Scholler N, et al. Potential role of HE4 in multimodal screening for epithelial ovarian cancer. J Natl Cancer Inst 2011;103:1630–4.
13. Bingham S, Riboli E. Diet and cancer–the European Prospective Investigation into Cancer and Nutrition. Nat Rev Cancer 2004;4:206–15.
14. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr 2002;5:1113–24.
15. Hainaut P, Vozar B, Rinaldi S, Riboli E, Caboux E. The European Prospective Investigation into Cancer and Nutrition biobank. Methods Mol Biol 2011;675:179–91.
16. Li K, Husing A, Fortner RT, Tjønneland A, Hansen L, Dossus L, et al. An epidemiologic risk prediction model for ovarian cancer in Europe: the EPIC study. Br J Cancer 2015;112:1257–65.
17. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med 2011;30:11–21.
18. Cramer DW, Bast RC Jr, Berg CD, Diamandis EP, Godwin AK, Hartge P, et al. Ovarian cancer biomarker performance in prostate, lung, colorectal, and ovarian cancer screening trial specimens. Cancer Prev Res (Phila) 2011;4:365–74.
19. Zhu CS, Pinsky PF, Cramer DW, Ransohoff DF, Hartge P, Pfeiffer RM, et al. A framework for evaluating biomarkers for early detection: validation of biomarker panels for ovarian cancer. Cancer Prev Res (Phila) 2011;4:375–83.
20. Drescher CW, Shah C, Thorpe J, O'Briant K, Anderson GL, Berg CD, et al. Longitudinal screening algorithm that incorporates change over time in CA125 levels identifies ovarian cancer earlier than a single-threshold rule. J Clin Oncol 2013;31:387–92.
21. Lu KH, Skates S, Hernandez MA, Bedi D, Bevers T, Leeds L, et al. A 2-stage ovarian cancer screening strategy using the Risk of Ovarian Cancer Algorithm (ROCA) identifies early-stage incident cancers and demonstrates high positive predictive value. Cancer 2013;119:3454–61.
22. Menon U, Ryan A, Kalsi J, Gentry-Maharaj A, Dawnay A, Habib M, et al. Risk algorithm using serial biomarker measurements doubles the number of screen-detected cancers compared with a single-threshold rule in the United Kingdom Collaborative Trial of Ovarian Cancer Screening. J Clin Oncol 2015;33:2062–71.
23. Skates SJ, Menon U, MacDonald N, Rosenthal AN, Oram DH, Knapp RC, et al. Calculation of the risk of ovarian cancer from serial CA-125 values for

preclinical detection in postmenopausal women. J Clin Oncol 2003;21: 206s–10s.
24. Azzam AZ, Hashad DI, Kamel NA. Evaluation of HE4 as an extrabiomarker to CA125 to improve detection of ovarian carcinoma: is it time for a step forward? Arch Gynecol Obstet 2013;288:167–72.
25. Ghasemi N, Ghobadzadeh S, Zahraei M, Mohammadpour H, Bahrami S, Ganje MB, et al. HE4 combined with CA125: favorable screening tool for ovarian cancer. Med Oncol 2014;31:808.
26. Karlsen MA, Hogdall EV, Christensen IJ, Borgfeldt C, Kalapotharakos G, Zdrazilova-Dubska L, et al. A novel diagnostic index combining HE4, CA125 and age may improve triage of women with suspected ovarian cancer - An international multicenter study in women with an ovarian mass. Gynecol Oncol 2015;138:640–6.
27. Karlsen MA, Sandhu N, Hogdall C, Christensen IJ, Nedergaard L, Lundvall L, et al. Evaluation of HE4, CA125, risk of ovarian malignancy algorithm (ROMA) and risk of malignancy index (RMI) as diagnostic tools of epithelial ovarian cancer in patients with a pelvic mass. Gynecol Oncol 2012;127:379–83.
28. Kristjansdottir B, Levan K, Partheen K, Sundfeldt K. Diagnostic performance of the biomarkers HE4 and CA125 in type I and type II epithelial ovarian cancer. Gynecol Oncol 2013;131:52–8.
29. Moore RG, Brown AK, Miller MC, Skates S, Allard WJ, Verch T, et al. The use of multiple novel tumor biomarkers for the detection of ovarian carcinoma in patients with a pelvic mass. Gynecol Oncol 2008;108: 402–8.
30. Van Gorp T, Cadron I, Despierre E, Daemen A, Leunen K, Amant F, et al. HE4 and CA125 as a diagnostic test in ovarian cancer: prospective validation of the Risk of Ovarian Malignancy Algorithm. Br J Cancer 2011;104:863–70.
31. Yurkovetsky Z, Skates S, Lomakin A, Nolen B, Pulsipher T, Modugno F, et al. Development of a multimarker assay for early detection of ovarian cancer. J Clin Oncol 2010;28:2159–66.
32. Jacobs I, Menon U. The sine qua non of discovering novel biomarkers for early detection of ovarian cancer: carefully selected preclinical samples. Cancer Prev Res 2011;4:299–302.
33. Pepe MS, Li CI, Feng Z. Improving the quality of biomarker discovery research: the right samples and enough of them. Cancer Epidemiol Biomarkers Prev 2015;24:944–50.
34. Schully SD, Carrick DM, Mechanic LE, Srivastava S, Anderson GL, Baron JA, et al. Leveraging biospecimen resources for discovery or validation of markers for early cancer detection. J Natl Cancer Inst 2015;107:djv012.

Published OnlineFirst April 8, 2016; DOI: 10.1158/1078-0432.CCR-16-0316

Clinical Cancer Research

A Prospective Evaluation of Early Detection Biomarkers for Ovarian Cancer in the European EPIC Cohort

Kathryn L. Terry, Helena Schock, Renée T. Fortner, et al.

Clin Cancer Res 2016;22:4664-4675. Published OnlineFirst April 8, 2016.