

Perspective

Perspective on Cramer et al., p. 365, and Zhu et al., p. 375

Challenges Related to Developing Serum-Based Biomarkers for Early Ovarian Cancer Detection

Phuong L. Mai¹, Nicolas Wentzensen², and Mark H. Greene¹

Abstract

In this issue of the journal, Cramer and colleagues and Zhu and colleagues report carefully designed phase 3 assessments of candidate ovarian cancer screening biomarkers. The main conclusion is that CA-125 remains the "best of a bad lot"; the new candidates have fallen short of expectations. We review factors impeding the development of an effective ovarian cancer screening strategy, highlight the requirements related to validating proposed screening biomarkers, and emphasize the risks from premature clinical applications of unvalidated tests, all underscoring the need for new research strategies. *Cancer Prev Res*; 4(3); 303–6. ©2011 AACR.

Introduction

Ovarian cancer is the second most common gynecologic malignancy in the United States, where it caused approximately 13,850 deaths in 2010 (1). An effective screening strategy has long been sought for this disease, which typically presents at an advanced stage and brings death to the majority of affected women. Numerous studies have been conducted to investigate candidate screening biomarkers for women at an average ovarian cancer risk. The majority of these studies have focused on CA-125, a large transmembrane glycoprotein first described in ovarian cancer cell lines in 1981 (2). The gene encoding the CA-125 antigen, *MUC16*, was cloned in 2001, but the physiologic function of this protein and its role in ovarian carcinogenesis and metastasis remain poorly understood (3). CA-125 is expressed in many tissues (4), and serum CA-125 levels are elevated in the settings of several cancers and benign conditions.

Early population-based studies were too small to provide conclusive results about the value of CA-125 testing for ovarian cancer early detection (5, 6). The combination of serum CA-125 and transvaginal ultrasonography (TVU) is currently being evaluated in large, randomized, population-based trials in the United States (both tests concurrently) and the United Kingdom (CA-125, followed by TVU only when the CA-125 level is abnormal). Data from the first screening round in the U.S. trial suggest that each of these 2 screening modalities has a low positive predictive

value (PPV; 3.7% for abnormal CA-125, 1.0% for abnormal TVU), which increases to 23.5% when both tests are abnormal (7). Mortality data, the golden metric by which screening trials are ultimately judged, are expected soon for this trial. Of interest, the strategy of using CA-125 with TVU indicated only for subjects with abnormal biomarker levels showed encouraging PPVs for ovarian cancer at the prevalence screen, but data on serial annual screening and mortality are not yet available (8).

Over the years, several studies investigating serum biomarkers other than CA-125 for early detection of ovarian cancer have shown promising results early on but very few markers have been evaluated in prospective studies to prove their value as potentially useful screening tests (9–11). Some studies reporting enthusiastically on ovarian cancer screening markers have been criticized as underpowered or methodologically flawed (12). Other approaches to cancer screening such as direct examination for changes in the target organ (e.g., mammography, cervical Pap smears, sigmoidoscopy) have been more successful because they can increase both sensitivity (due to direct visualization of the target organ or its changes) and specificity (not measuring factors that can be influenced by other sources in the body).

Therefore, reports from studies funded by the Early Detection Research Network (EDRN) using prospectively diagnosed ovarian cancer data from the Prostate, Lung, Colorectal, and Ovarian Cancer (PLCO) screening trial have been eagerly awaited. The authors of the 2 reports in this issue of the journal are to be commended for having designed and conducted scientifically solid phase 3 studies (Table 1; ref. 13), which were nested in a large, randomized screening trial and will serve as the standard against which future analyses of this kind should be judged (14, 15). It is frustrating that none of the 28 ovarian cancer serum biomarkers selected for in-depth analysis in prediagnostic serum specimens from PLCO ovarian cancer cases and controls were shown, when evaluated singly, to have test

Authors' Affiliations: ¹Clinical Genetics Branch and ²Hormone and Reproductive Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, Maryland

Corresponding Author: Mark H. Greene, 6120 Executive Boulevard, Room EPS 7032, Rockville, MD 20852. Phone: 301-594-7641; Fax: 1-301-496-1854; E-mail: greenem@mail.nih.gov.

doi: 10.1158/1940-6207.CAPR-11-0053

©2011 American Association for Cancer Research.

Table 1. Phases of development of early cancer detection biomarkers

Phases	Purpose
Phase 1: Preclinical exploratory studies	Identification of potentially discriminating biomarkers. Usually involves comparing tumor tissue with normal tissue. Exploratory data analysis is an integral part of this phase.
Phase 2: Clinical assay development for clinical disease	Optimization of the assay (reproducibility and specimen source) used to measure the biomarker identified in phase 1. Determination of the performance characteristics of the biomarker assay to distinguish cases from noncases. Identification of factors that are associated with biomarker levels. Note: The cases and controls selected for this phase should ideally be representative of population to be screened.
Phase 3: Retrospective longitudinal repository studies	Determination of the capacity, as a function of time before clinical diagnosis, of a biomarker to detect subclinical disease, using specimens obtained prior to clinical diagnosis for cases. Identification of covariates that can modify the abilities of the biomarker to discriminate between those with and without subclinical disease. Selection of biomarkers or panels of biomarkers that seem to be most promising. Establishment of the criteria for a positive screening test and the screening interval, if appropriate, to be used in phase 4.
Phase 4: Prospective screening studies	Determination of the operating characteristics of the biomarker-based screening test to detect asymptomatic cancer at an early stage of development, a point at which initiation of treatment is more likely to result in an improved outcome. Assessment of feasibility of a large-scale screening program and compliance. Collection of preliminary data on the effects of screening on costs and mortality due to the cancer being screened.
Phase 5: Cancer control studies	Determining whether screening results in a reduction in disease morbidity and mortality in large, randomized, controlled, clinical trials in target populations. Obtaining data on cost-effectiveness of the screening program.

performance characteristics that were equal, let alone superior, to CA-125 levels. Furthermore, when these biomarkers were evaluated in multianalyte panels, based on predefined models, combinations of biomarkers did not improve test performance measures compared with CA-125 alone.

Why has it been so difficult to develop an effective serum biomarker-based ovarian cancer screening strategy? In the following sections, we present some of the requirements for a successful screening biomarker candidate, that is, one that can reduce mortality at an acceptable cost. Unfortunately, some of these requirements are very difficult to achieve in ovarian cancer screening.

Early Enough Cancer Detection that Intervention Is Likely to Alter Disease Outcome

The window between when early detection can improve outcome and when it becomes too late for effective intervention is often narrow. A test with apparently adequate performance characteristics for detection might not result in clinically meaningful changes in disease outcome if the cancer is not detected at a sufficiently early stage (16). In addition, the window of

meaningful early detection must be sufficiently wide to permit a reasonable screening interval. Screening intervals must be short when there is only a brief duration between first test positivity and the end of an opportunity for successful interventions. Some models have shown that screening intervals of less than 1 year might be required to achieve substantial reductions in mortality for ovarian cancer (17). The early phase of development of a new test generally employs blood samples acquired at the time of a clinical cancer diagnosis, and the cases ascertained in this fashion might include cancers that are biologically more advanced than would be ideal for successful intervention. To the extent that advanced disease is included in the analysis, the performance characteristics of the test might be misleading. On the other hand, early detection of indolent disease might result in overdiagnosis, treatment of clinically insignificant cases, and no net improvement in disease-specific mortality. Screening preferentially detects slow-growing, more benign tumors with longer progression times that are less likely to be fatal without screening, resulting in an overly positive assessment of screening benefit. Overdiagnosis of indolent disease can increase intervention-related morbidity and mortality, with little to no survival benefit.

Downloaded from <http://aacrjournals.org/cancerpreventionresearch/article-pdf/4/3/303/2337383/303.pdf> by guest on 29 May 2022

Sensitive Enough to Detect the Target Cancer at an Asymptomatic Stage and Specific Enough to Avoid a Significant False-Positive Rate

Sensitivity and specificity are determined by the distribution of a biomarker in cases and controls and are maximized when the distribution between cases and controls is very different. The requirement for a sufficiently large difference in average test levels between cases and controls for effective early detection is often difficult to achieve because oftentimes only larger, later-stage cancers would release readily detectable levels of a particular biomarker molecule. With regard to specificity, CA-125 and the other serum biomarkers investigated to date are not exclusively associated with ovarian cancer (10); elevated levels may be associated with other cancers and nonovarian diseases.

As shown in the PLCO studies, several biomarkers can be measured simultaneously (in "panels"), with the results based on combining presumably independent information derived from each of the different markers rather than considering each marker individually. Although biomarker panels can potentially increase performance, for example, by combining several highly specific markers that have low sensitivity individually, the multimarker panels included in the study by Cramer and colleagues (15) did not live up to that theoretical potential. Risk modeling based on serial CA-125 measurements over time comprises another novel strategy aimed at improving screening test performance. Results from the prevalence screen in a general population study, based on the risk of ovarian cancer algorithm (ROCA), showed a promising PPV of 43% for the ROCA arm of the trial, which remains in follow-up (8).

Common Enough Target Tumor in the Screening Population for a Highly Sensitive, Specific Test to Achieve an Adequate PPV

A validated biomarker must result in test-positive individuals having a sufficiently high probability of occult cancer to warrant an intervention that might mitigate disease morbidity and mortality (adequate PPV). Likewise, individuals testing negative for the biomarker must be reasonably certain that an intervention is not required [adequate negative predictive value (NPV)]. The prevalence of disease determines the PPV and NPV for a biomarker with a given sensitivity and specificity. Ovarian cancer is a rare disease, with an estimated prevalence among postmenopausal women of approximately 1 in 2,500. At this prevalence, with a sensitivity of 75%, a screening test must have specificity of more than 99.6% to achieve a PPV of 10% or more. Although the tolerable PPV threshold depends on available follow-up test(s) and disease natural history, 10% (or 10 operations for each detected cancer) has historically been viewed as the lowest acceptable PPV for ovarian cancer screening. A screening test with a high false-positive rate is particularly problematic in ovarian cancer screening because a definitive workup would require

bilateral salpingo-oophorectomy, an invasive intervention with potentially significant morbidity. Note that a screening test that is inappropriate for the general population might be very beneficial in a high-risk population, such as women with *BRCA1/2* mutations, because of its higher ovarian cancer prevalence and hence a higher PPV for the test.

Understanding Enough of the Cancer's Natural History and Carcinogenesis Basis Can Help Determine Whether Screening Is Likely to Improve Survival

The ideal screening program relies on a test that identifies disease or indicates risk at a time when an intervention can effectively interrupt the natural history of disease. Over time, the test levels associated with either risk of developing disease or the disease itself become increasingly different between cases and unaffected individuals but the effectiveness of interventions tend to diminish. Unfortunately, ovarian cancer is an etiologically heterogeneous group of diseases (18) and precursors to the most aggressive cancers have not been identified. Moreover, the natural history of ovarian cancer is poorly understood, and many questions, such as the cell of origin of ovarian cancer, its site of initiation, and the duration between initiation of and incurable disease, remain unanswered. With the sobering findings of the PLCO biomarker studies in hand, we need to go back to the drawing board to identify other more appropriate and more promising screening biomarkers.

Applying a Rational, Systematic Approach to Developing and Validating Screening Biomarkers

A structured, systematic approach to developing and validating new biomarkers is essential. A 5-phase framework has been proposed by the EDNR (Table 1). As shown in the 2 articles published in this issue of the journal (14, 15), candidate biomarkers identified in earlier-phase studies frequently are not validated by later-phase studies. Furthermore, although the identification of novel, seemingly promising biomarkers in early-phase studies often leads to initial enthusiasm, a thorough validation is necessary to avoid premature acceptance of their clinical utility. Equally important, if performance characteristics from early-phase studies indicate that the biomarker will most likely not be successful in the specific setting of interest, evaluation in a large, costly trial needs to be avoided.

The premature proposals to introduce 2 new biomarker-based tests for ovarian cancer screening into clinical practice have provided invaluable object lessons. One a blood test comprising a 6-analyte panel (19) and the other a proteomic assay (20) were both reported to have remarkably favorable PPV in initial reports, but these parameters were estimated from cross-sectional data without properly taking population-specific disease prevalence into account

(21). Unfortunately, the ability to distinguish clinically detected cases from controls may have little relevance for the ultimate performance characteristics of tests involving prediagnostic serum in detecting asymptomatic, prospectively diagnosed ovarian cancers. In the prospective evaluation reported in this issue, the 6-analyte panel did not live up to its expectations (14, 15). Neither of these proposed assays has been recommended for clinical practice. On the basis of current knowledge, it is difficult to envision a scenario in which a new ovarian cancer biomarker would be proposed for clinical application without first having been studied in the manner described by Zhu and colleagues (14) and Cramer and colleagues (15), followed by further prospective studies and randomized trials (Table 1).

Conclusions

Faced with these complicated realities, the medical community and the public must remain appropriately skeptical when a new serum-based, ovarian cancer biomarker screening test is proposed and must examine the evidence carefully, using the criteria discussed earlier. The pressure on the scientific community from providers and at-risk women alike to develop such a test is as great as it is understandable. Until a validated screening strategy for ovarian cancer in the general population is in hand, however, we believe that no test is preferable to an unproven test, given the potential

harms summarized earlier. At least theoretically, inappropriate interventions could paradoxically increase mortality among women being screened rather than improving life expectancy and quality of life, the goal for which we all strive. As discouraging as the results published in this issue of the journal might be regarding the current state of biomarker-based ovarian cancer screening, we have learned that the process for identifying and selecting new candidate biomarkers for further development has not yielded promising candidates and no lesson could be more important. Simply continuing to do more discovery of the kind illustrated here would seem to be an inefficient use of increasingly scarce research resources. We urgently need novel, meticulously evaluated research ideas if we are to solve the dilemma of ovarian cancer screening.

Disclosure of Potential Conflicts of Interest

The authors report no conflicts of interest.

Grant Support

The research of P.L. Mai, N. Wentzensen, and M.H. Greene is supported by the Intramural Research Program of the National Cancer Institute, NIH.

Received January 20, 2011; revised January 24, 2011; accepted January 25, 2011; published online March 3, 2011.

References

- Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. *CA Cancer J Clin* 2010;60:277–300.
- Bast RC Jr, Feeney M, Lazarus H, Nadler LM, Colvin RB, Knapp RC. Reactivity of a monoclonal antibody with human ovarian carcinoma. *J Clin Invest* 1981;68:1331–7.
- Bouanene H, Miled A. Conflicting views on the molecular structure of the cancer antigen CA125/MUC16. *Dis Markers* 2010;28:385–94.
- Karam AK, Karlan BY. Ovarian cancer: the duplicity of CA125 measurement. *Nat Rev Clin Oncol* 2010;7:335–9.
- Einhorn N, Sjøvall K, Knapp RC, Hall P, Scully RE, Bast RC Jr, et al. Prospective evaluation of serum CA 125 levels for early detection of ovarian cancer. *Obstet Gynecol* 1992;80:14–8.
- Helzlsouer KJ, Bush TL, Alberg AJ, Bass KM, Zacur H, Comstock GW. Prospective study of serum CA-125 levels as markers of ovarian cancer. *JAMA* 1993;269:1123–6.
- Buys SS, Partridge E, Greene MH, Prorok PC, Reding D, Riley TL, et al. Ovarian cancer screening in the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial: findings from the initial screen of a randomized trial. *Am J Obstet Gynecol* 2005;193:1630–9.
- Menon U, Gentry-Maharaj A, Hallett R, Ryan A, Burnell M, Sharma A, et al. Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *Lancet Oncol* 2009;10:327–40.
- 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.
- Husseinzadeh N. Status of tumor markers in epithelial ovarian cancer has there been any progress? A review. *Gynecol Oncol* 2011;120:152–7.
- Dutta S, Wang FQ, Phalen A, Fishman DA. Biomarkers for ovarian cancer detection and therapy. *Cancer Biol Ther* 2010;9:668–77.
- Diamandis EP. Cancer biomarkers: can we turn recent failures into success? *J Natl Cancer Inst* 2010;102:1462–7.
- 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.
- 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* 2011;4:375–383.
- Cramer DW, Bast RC Jr, Berg CD, Godwin AK, Hartge P, Lokshin AE, et al. Ovarian cancer biomarker performance in Prostate, Lung, Colorectal, and Ovarian Cancer screening trial specimens. *Cancer Prev Res* 2011;4:365–374.
- Clarke-Pearson DL. Screening for ovarian cancer. *New Engl J Med* 2009;361:170–7.
- Havrilesky LJ, Sanders GD, Kulasingam S, Myers ER. Reducing ovarian cancer mortality through screening: Is it possible, and can we afford it? *Gynecol Oncol* 2008;111:179–87.
- Kurman RJ, Shih Ie M. The origin and pathogenesis of epithelial ovarian cancer: a proposed unifying theory. *Am J Surg Pathol* 2010;34:433–43.
- Visintin I, Feng Z, Longton G, Ward DC, Alvero AB, Lai Y, et al. Diagnostic markers for early detection of ovarian cancer. *Clin Cancer Res* 2008;14:1065–72.
- Petricoin EF III, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002;359:572–7.
- The importance of test positive predictive value in ovarian cancer screening. *Clin Cancer Res* 2008;14:7574.