

Perspective

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

The *Sine Qua Non* of Discovering Novel Biomarkers for Early Detection of Ovarian Cancer: Carefully Selected Preclinical Samples

Ian Jacobs, and Usha Menon

Abstract

This perspective discusses reports by Cramer and colleagues (beginning on page 365 in this issue of the journal) and Zhu and colleagues (beginning on page 375), which provide the first systematic and reliable comparison of a large number of candidate biomarkers for the early detection of ovarian cancer in a sample set well-suited for this purpose. This research has important implications for the future design of cancer biomarker studies. *Cancer Prev Res*; 4(3); 299–302. ©2011 AACR.

There has been a proliferation of candidate biomarkers for ovarian cancer during the last decade, driven by recognition of the need for progress in managing this disease and by technological developments including mass spectrometry and multiplex analytic approaches. Excitement generated by publication of new markers and marker panels and related media coverage has been matched by controversy over their true potential for early detection and by disappointment and frustration when anticipated or claimed breakthroughs have not been validated. Despite intensive efforts to find new, better biomarkers for the early detection of ovarian cancer, the gold standard remains CA 125, a marker discovered using monoclonal antibody technology almost 30 years ago.

It is possible that the failure to improve upon CA 125 is a genuine indication that current technologies are unable to identify biomarkers with greater or complementary potential for early detection and diagnosis. This seems unlikely, however, given the power of available technologies and our vast knowledge of the molecular changes and biology of cancer. An alternative explanation is that the available technologies, powerful as they are, have not yet been applied appropriately either in discovery or in validation.

Against this background, new data published in this issue of the journal by Cramer and colleagues (1) and Zhu and colleagues (2) are of great importance and provide valuable insight into the current status of this field. First, the articles represent a major collaborative effort involving 9 of the leading research centers for ovarian cancer research in

the United States, an achievement in itself for which the authors should be congratulated. Second, the research team has clearly distinguished between samples obtained at the time of clinical diagnosis (phase II specimens) and samples collected and banked prior to clinical diagnosis (phase III specimens). Third, the studies utilized phase III specimens and related data accumulated within the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer study. Fourth, careful attention to the design of the analyses included a systematic blinded approach allowing comparison between biomarkers, biomarker panels, and centers. The result is the first systematic and reliable comparison of a large number of candidate biomarkers for early detection of ovarian cancer in a sample set well-suited for this purpose.

The study of Cramer and colleagues (1) involves a comparison of the sensitivity and specificity of 35 biomarkers for ovarian, primary peritoneal, and fallopian tube cancers in phase II and phase III sample sets. The results were reported at a fixed specificity of 95%, and CA 125 proved to have the highest sensitivity in phase II samples overall (73%), early-stage phase II samples (56%), and phase III samples taken within 6 months of diagnosis (86%). A number of other markers (HE4, transthyretin, CA 15.3, CA 72.4) had sensitivities ranging from 40% to 60% for all phase II samples, but none had a sensitivity of more than 40% for early-stage phase II samples and only HE4 (at 73%) had a sensitivity of more than 50% for phase II samples from within 6 months of diagnosis. Sensitivity in phase III samples decreased significantly at increasing intervals before diagnosis for CA 125 (86% within 6 months, 33% >6–12 months, 12% >12–18 months) and for HE4 (73% within 6 months, 23% >6–12 months, 18% >12–18 months). The most important finding in this study was the lack of evidence that any of the newer biomarkers improve upon the performance of CA 125 in either the phase II or phase III setting.

The study of Zhu and colleagues (2) involved 5 previously described panels of 6 to 8 biomarkers each (28 total

Authors' Affiliation: Gynaecological Oncology, Institute for Women's Health, University College London, United Kingdom

Corresponding Author: Ian Jacobs, Gynaecological Oncology, UCL EGA Institute for Women's Health, 149 Tottenham Court Road, London W1T 7DN, United Kingdom. Phone: 020-7380-9747; Fax: 020-7380-9748. E-mail: i.jacobs@ucl.ac.uk

doi: 10.1158/1940-6207.CAPR-11-0048

©2011 American Association for Cancer Research.

biomarkers, with some overlap between panels) with a phase III case-control set of 118 ovarian cancers and 951 controls. Cancer samples were obtained within 365 days prior to diagnosis in 57% of cases and from 366 up to 2,898 days before diagnosis in the rest. Three analytic approaches were used sequentially, from validation blinding all samples, through splitting into a training and validation set, to a discovery set without validation. Surprisingly and disappointingly, none of the panels or analytic approaches revealed an improvement compared with CA 125 alone, which was analyzed at a cutoff of 35 U/mL and had a sensitivity/specificity of 63.1%/98.5% in the validation-only set and 72.4%/97.9% in the training and validation set, with an area under the curve in the validation set of 0.89. Only one panel, which included CA 125 along with B7-H4, CA15-3, CA72-4, and HE4, had results comparable with those of CA 125 alone. The Yale algorithm panel (CA125, IGF2, leptin, MIF, OPN, and prolactin) performed particularly poorly compared with previously published data for the same algorithm in phase II samples, with a sensitivity of just 34.3% at a specificity of 96.8% in the validation-only set. The Pittsburgh panel (CA125, CA72-4, EGFR, eotaxin, HE4, MMP3, prolactin, and VCAM1) was equally disappointing, with sensitivity/specificity of only 37.9%/89.8% in the validation-only set.

Lessons can be learned from these carefully conducted studies and from the experience over the last two decades in sample selection for discovery of biomarkers relevant to screening/early detection and their validation. A key message is selection of the appropriate sample set for screening biomarker discovery. Several markers with high ranks in early-stage phase II samples, such as CA19.9, apolipoprotein A1, and prolactin, did not perform well in phase III specimens (1). The intended use of a diagnostic marker is to aid assessment of whether symptomatic, clinically presenting patients have a particular condition, and a screening marker is intended to assess in advance of symptoms whether disease is present in apparently healthy individuals. The standard approach of biomarker discovery to date has been to undertake studies for both purposes in clinical samples obtained from symptomatic patients and often from patients with advanced stage disease. Even an approach utilizing phase II samples from patients with clinically diagnosed early-stage ovarian cancer is dubious. There is increasing evidence that clinically diagnosed early-stage ovarian cancer is often low-grade serous/endometrioid/mucinous carcinoma, which behaves indolently and thus contrasts with advanced stage and highly aggressive high-grade serous/undifferentiated/malignant mixed mesodermal ovarian cancers. The results reported in this issue of the journal (1, 2) highlight the need to use sample sets that precede a cancer diagnosis by more than 6 to 12 months (phase III sample sets) for discovery of screening biomarkers. There have been very few reports in the literature to date on the use of this approach, and one can be reasonably optimistic that discovery with well-characterized phase III samples will yield novel screening markers during the next few years.

The second message is that samples for biomarker validation must be chosen carefully and appropriately. It has become increasingly clear that systematic bias introduced by differences in cases and controls has led to the exaggerated reports of cancer biomarker performance that have appeared over the past decade. The current studies highlight this point and provide strong support for future biomarker study designs that involve nested case-control studies within clinically relevant prospective cohort studies, in which specimens have been banked before outcome ascertainment. The design proposed in 2008 by Pepe and colleagues for prospective specimen collection, retrospective blinded evaluation (PRoBE) eliminates many common biases since specimens are collected and handled in a "blinded" manner, prior to diagnosis (3). Given the large biobanks built in the course of ongoing screening [PLCO, ref. 4; United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS), ref. 5] and prevention [European Prospective Investigation into Cancer and Nutrition (EPIC); refs. 6, 7] trials and those with high-quality biospecimen repositories resulting from specific biobanking efforts (e.g., UKBiobank; ref. 8), there should now be sufficient high-quality samples to facilitate optimal study designs. Every effort must be made to access samples from these banks for future biomarker studies. It is also imperative that the custodians of the banks make every effort to facilitate such collaborations if the full potential of these banks to help reduce cancer mortality through early detection is to be realized. This goal will often require an innovative open-access approach involving both commercial and academic partners.

Another important issue is the heterogeneity of ovarian cancer. The use of all histologic types of ovarian cancer to identify early detection markers is based upon the paradigm that ovarian cancer is a single disease originating in the ovary and spreading to the pelvis, abdomen, and distant sites. Recent morphologic and molecular genetic studies have questioned these assumptions, with important implications both for our understanding of the origin of ovarian cancer and for early detection research. These studies distinguish 2 groups of epithelial ovarian tumors—type I and type II. Type I comprises low-grade serous, low-grade endometrioid, clear cell, mucinous, and transitional carcinomas, which exhibit a shared lineage with the corresponding benign cystic neoplasm, often through an intermediate (borderline tumor) step. In contrast, type II tumors are highly aggressive, evolve rapidly, and in most cases have *TP53* mutations. They include high-grade serous, undifferentiated, and malignant mixed mesodermal cancers. As type II tumors account for most of the mortality associated with ovarian cancer, it is crucial for the impact of screening on mortality that histologic type is taken into account in choosing sample sets for discovery. In the current reports of Cramer and colleagues and Zhu and colleagues, borderline tumors were excluded and primary invasive ovarian, fallopian, and peritoneal cancers were included. Although the majority of patients in these

studies had primary invasive epithelial ovarian cancer, a few nonepithelial (granulosa, squamous) and unknown histology cancers were included. Ideally, a histologic review of all cancers in the cohort with reclassification into type I and type II cancers should be undertaken and along with a focus on initial discovery of early detection biomarkers in type II cancers (9).

There are suggestions that type II cancers originate in other pelvic organs (tube, endometrium) and involve the ovary secondarily (10). This possibility implies that early detection of type II tumors might not result in a stage shift but in detection only of low-volume disease. Residual disease is nevertheless one of the most powerful prognostic markers in ovarian cancer, and low-volume disease may still translate into better outcomes. It is likely that this explains the significant survival benefit (median survival of women with index cancers in the screened group 72.9 months vs. 41.8 months in the control group) noted in the absence of stage shift in the first U.K. randomized controlled trial of ovarian cancer screening (11).

Screening for type II cancers will require sensitive and specific biomarkers that are expressed during ovarian carcinogenesis and reach the peripheral circulation early enough at levels distinguishable from those in healthy individuals and thus permit interventions that can alter the natural history of the disease. Many investigators are skeptical that this objective is achievable with current screening strategies, which are based on CA 125 and pelvic ultrasonography. After 4 rounds of screening in the PLCO trial, 72% of screen-detected ovarian cancer cases were late stage (III/IV; ref. 12). Neither the PLCO trial nor the studies reported here by Cramer and colleagues and Zhu and colleagues, however, utilized serial changeover time in individuals' biomarker profiles. A longitudinal time-series approach to CA 125 analysis has now been shown to increase CA 125 performance in 3 separate screening trials (13–15). In each of these studies, CA 125 was interpreted according to the Risk of Ovarian Cancer (ROC) algorithm, which uses a change-point model to estimate risk based on age-specific ovarian cancer incidence and serial CA 125 levels in serum (16). It is likely that an improved screening performance achieved in UKCTOCS (14) compared with that in the ovarian arm of the PLCO was due to the use of the ROC algorithm in UKCTOCS (along with a more rigorously defined and managed screening protocol; ref. 17). The sensitivity for preclinical primary invasive epithelial ovarian cancer, using the ROC algorithm to interpret CA 125 in UKCTOCS, was 89.5% in the CA 125 (multimodal) group at prevalence screening (12) compared with 67.4% after 4 rounds of screening in PLCO (12, 17). More important, 47.1% of patients in UKCTOCS had early-stage

disease at incidence screening compared with 22% at incidence screening (17) and 28% after 4 rounds of screening (12) in PLCO. In the multimodal arm of UKCTOCS, 9 of the 34 women with screen-detected primary invasive epithelial cancer initially had a normal CA 125 result, but they underwent repeat testing for an estimated "intermediate" risk based on the ROC algorithm. During ongoing incidence screening in UKCTOCS, it has become increasingly clear that low-volume, high-grade serous ovarian cancer can be detected with serial CA 125 monitoring interpreted by the ROC algorithm in the absence of ultrasound abnormalities (unpublished data).

Whether or not optimal use of CA 125 in screening for ovarian cancer can lead to a reduction in ovarian cancer mortality will be unclear until the PLCO and UKCTOCS studies report their mortality findings. It is already clear, however, that CA 125 can achieve a significant diagnostic lead time over clinical diagnosis. Indeed, a limitation of the reports by Cramer and colleagues and Zhu and colleagues is that their baseline is influenced by intervention based on CA 125 measurement in the PLCO trial. Women with elevated CA 125 levels tended to be diagnosed within 1 year of sample collection, and samples from these cases were excluded from the more than 12-month sample set, which must be interpreted in this context. For this reason, biobanks such as PLCO and UKCTOCS are ironically of greater value for estimating the potential lead time in nonovarian cancers and diseases which were not targeted by the screening intervention.

It is now possible to link the power of novel technologies for biomarker discovery with carefully selected phase III samples from biobanks such as UKCTOCS and PLCO. Of equal importance, there is increasing acceptance of the need for rigorous attention to study design, especially in the choice of sample sets. The new approach integrating these advances offers hope for identification of a novel generation of biomarkers which can achieve sufficient lead time to alter the natural history of ovarian and other cancers and reduce their mortality. The reports by Cramer and colleagues and Zhu and colleagues provide a foundation for this work.

Disclosure of Potential Conflicts of Interest

Both authors collaborate on studies of ovarian cancer biomarkers and screening with some of the authors of the primary companion articles. Both authors have a financial interest in Abcodia, Ltd., a company formed to develop academic and commercial development of biomarkers for screening and risk prediction. I. Jacobs is a consultant to Becton Dickinson in the field of ovarian cancer.

Received January 10, 2011; revised January 22, 2011; accepted January 24, 2011; published online March 3, 2011.

References

1. Cramer DW, Bast RC, Berg CD, et al. Ovarian cancer biomarker performance in prostate, lung, colorectal, and ovarian cancer screening trial specimens. *Cancer Prev Res* 2011;4:365–374.
2. Zhu CS, Pinsky PF, Cramer DW, et al. A Framework for evaluating biomarkers for early detection: validation of biomarker panels for ovarian cancer. *Cancer Prev Res* 2011;4:375–383.

3. Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst* 2008;100:1432–8.
4. Hayes RB, Sigurdson A, Moore L, Peters U, Huang WY, Pinsky P, et al. Methods for etiologic and early marker investigations in the PLCO trial. *Mutat Res* 2005;592:147–54.
5. Menon U, Gentry-Maharaj A, Ryan A, Sharma A, Burnell M, Hallett R, et al. Recruitment to multicentre trials—lessons from UKCTOCS: descriptive study. *BMJ* 2008;337:a2079.
6. Riboli E. Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). *Ann Oncol* 1992;3:783–91.
7. Beulens JW, Monninkhof EM, Verschuren WM, Van Der Schouw YT, Smit J, Ocke MC, et al. Cohort profile: the EPIC-NL study. *Int J Epidemiol* 2010;39:1170–8.
8. Elliott P, Peakman TC. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int J Epidemiol* 2008;37:234–44.
9. 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.
10. Folkins AK, Jarboe EA, Roh MH, Crum CP. Precursors to pelvic serous carcinoma and their clinical implications. *Gynecol Oncol* 2009;113:391–6.
11. Jacobs IJ, Skates SJ, MacDonald N, Menon U, Rosenthal AN, Davies AP, et al. Screening for ovarian cancer: a pilot randomised controlled trial. *Lancet* 1999;353:1207–10.
12. Partridge E, Kreimer AR, Greenlee RT, Williams C, Xu JL, Church TR, et al. Results from four rounds of ovarian cancer screening in a randomized trial. *Obstet Gynecol* 2009;113:775–82.
13. Menon U, Skates SJ, Lewis S, Rosenthal AN, Rufford B, Sibley K, et al. Prospective study using the risk of ovarian cancer algorithm to screen for ovarian cancer. *J Clin Oncol* 2005;23:7919–26.
14. 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.
15. K. H. Lu SS, T. B. Bevers TB, W. Newland W, R. G. Moore RG, L. Leeds L, S. Harris S, et al. A prospective U.S. ovarian cancer screening study using the risk of ovarian cancer algorithm (ROCA). *J Clin Oncol* 2010; 28 Suppl:15s:abstr 5003.
16. Skates SJ PD, Jacobs IJ. Screening based on the risk of cancer calculation from Bayesian hierarchical change point and mixture models of longitudinal markers. *J Am Stat Assoc* 2001;96:429–35.
17. 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.