ORIGINAL REPORT

Longitudinal Screening Algorithm That Incorporates Change Over Time in CA125 Levels Identifies Ovarian Cancer Earlier Than a Single-Threshold Rule

Charles W. Drescher, Chirag Shah, Jason Thorpe, Kathy O'Briant, Garnet L. Anderson, Christine D. Berg, Nicole Urban, and Martin W. McIntosh

Charles W. Drescher, Jason Thorpe, Kathy O'Briant, Garnet L. Anderson, Nicole Urban, and Martin W. McIntosh, Fred Hutchinson Cancer Research Center; Chirag Shah, Pacific Gynecology Specialists, Seattle, WA; and Christine D. Berg, National Cancer Institute, Bethesda, MD.

Corresponding author: Charles Drescher, MD, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, M2-B230, Seattle, WA 98104; e-mail: cdresche@fhcrc.org.

A B S T R A C T

Purpose
Longitudinal algorithms incorporate change over time in biomarker levels to individualize screening decision rules. Compared with a single-threshold (ST) rule, smaller deviations from baseline biomarker levels are required to signal disease. We demonstrated improvement in ovarian cancer early detection by using a longitudinal algorithm to monitor annual CA125 levels.

Patients and Methods
We retrospectively evaluated serial preclinical serum CA125 values measured annually in 44 incident ovarian cancer cases identified from participants in the PLCO (Prostate Lung Colorectal and Ovarian) Cancer Screening Trial to determine how frequently and to what extent the parametric empirical Bayes (PEB) longitudinal screening algorithm identifies ovarian cancer earlier than an ST rule.

Results
The PEB algorithm detected ovarian cancer earlier than an ST rule in a substantial proportion of cases. At 99% specificity, which corresponded to the ST-rule CA125 cutoff $\geq$ 35 U/mL that was used in the PLCO trial, 20% of cases were identified earlier by using the PEB algorithm. Among these cases, the PEB signaled abnormal CA125 values, on average, 10 months earlier and at a CA125 concentration 42% lower (20 U/mL) than the ST-rule cutoff. The proportion of cases detected earlier by the PEB algorithm and the earliness of detection increased as the specificity of the screening rule was reduced.

Conclusion
The PEB longitudinal algorithm identifies ovarian cancer earlier and at lower biomarker concentrations than an ST screening algorithm adjusted to the same specificity. Longitudinal biomarker assessment by using the PEB algorithm may have application for screening other solid tumors in which biomarkers are available.

J Clin Oncol 31:387-392. © 2012 by American Society of Clinical Oncology

INTRODUCTION

Blood-based biomarkers are of particular interest as cancer screening tests as a result of their convenience, low cost, and quantitative measure. The incorporation of biomarkers into a screening program requires algorithms or decision rules to identify, from among all participants, individuals who are likely to have disease. The most commonly used algorithm is a single-threshold (ST) rule, which turns positive when the biomarker concentration exceeds a common population wide threshold. The threshold may be adjusted to account for patient-specific characteristics that are known to influence the level of a biomarker (eg, the impact of menopausal status on the CA125 level) but does not depend on screening history. Longitudinal algorithms use information available from the screening history of an individual to personalize screening decisions. Compared with an ST rule, smaller deviations from baseline biomarker levels are required to identify a tumor.[1,2]

The incorporation of a longitudinal algorithm into a cancer screening program should be compelled by evidence that the longitudinal decision rule improves screening performance. The magnitude of improvement is likely to depend on many factors including the behavior of the marker in healthy individuals and those with disease, screening periodicity, and disease natural history in addition to characteristics of the algorithm itself. Although several longitudinal algorithms have been described,[1-4]

little information has been reported regarding their impact on screening performance. One critical measure is how often and to what extent a longitudinal algorithm identifies cancer earlier than an ST rule set at an identical specificity.

The parametric empirical Bayes (PEB) longitudinal algorithm uses PEB statistical theory to model the trajectory of markers in asymptomatic healthy individuals over time. The estimated trajectories are used to generate person-specific positivity thresholds that depend on the screening history of each individual.[1,2] Any specificity threshold can be selected. Compared with an ST rule, the PEB lowers the positivity threshold for most individuals and spreads false-positive tests evenly across the screened population. The impact of the PEB algorithm is greatest for markers that are more similar within compared with between healthy individuals.[5,6] The PEB algorithm is closely related to methods currently used to establish "biologic passports" to monitor athletes for performance-enhancing drugs.[7-10]

The aim of this study was to determine whether monitoring annual CA125 biomarker levels by using the PEB algorithm improves ovarian cancer detection compared with an ST rule. By using serial CA125 levels from participants in the intervention arm of the US PLCO (Prostate Lung Colorectal and Ovarian) Cancer Screening Trial, we retrospectively screened selected cases and controls by using both algorithms configured to an identical specificity. The date of the first positive screen for each algorithm was recorded. The frequency of discordant decisions among cases and controls across a range of screening specificities is measured, and we evaluated how much earlier (or later) the PEB rule identified cancer than the ST rule.

We found that, for all specificities tested, the PEB algorithm identified a meaningful fraction of ovarian cancers earlier than the ST rule, and CA125 concentrations at the time of a PEB-positive screen were much lower than threshold values for the corresponding ST rule. Our findings have important implications for designing ovarian cancer screening strategies by using CA125 and possibly other markers.

PATIENTS AND METHODS

PLCO Cancer Screening Trial Data

Details of the PLCO study design have been reported previously.[11-13] Briefly, the PLCO is a multicenter controlled screening trial that randomly assigned women age 55 to 74 years to an intervention arm (n = 39,105) that included ovarian cancer screening by using concurrent CA125 and transvaginal ultrasound (TVU) or a control arm (n = 39,111) that consisted of standard clinical care.[12] Enrollment began in November 1993 and concluded in July 2001. Participants assigned to the ovarian cancer screening arm were scheduled to undergo four annual screens. The screening arm was later modified to include two additional years of screening by using CA125 alone (6 years total). CA125 levels were measured at a central reference laboratory by using the Fujirebio CA-125 II radioimmunoassay (Malvern, PA). Women with CA125 levels that exceeded 35 units/mL or an abnormal TVU were referred to their primary care physicians for additional evaluation. PLCO study guidelines did not define a specific follow-up for abnormal tests, but data on all follow-up testing and interventions were collected. Incident cancer cases were identified by annual study questionnaires and through linkage to population-based cancer registries. Participants were followed for a maximum of 13 years from the date of enrollment. The final results showed that there was no ovarian cancer–specific mortality reduction from screening with CA125 and TVU compared with controls.[13]

A total of 221 cases of invasive ovarian, primary peritoneal, or fallopian tube cancer have been diagnosed in the intervention arm. We obtained serial preclinical CA125 values for the 71 incident cases diagnosed before 2006 and

controls matched to cases by age (5-year intervals) and calendar year at the time of the blood draw closest to the diagnosis of the case. We restricted our analysis to the 44 incident cases and 701 controls with at least two consecutive blood samples and whose proximate blood draw falls within a year of diagnosis or the completion of screening. These criteria ensured that study participants provided a minimum number of samples for estimating patient-specific thresholds by using the PEB algorithm and avoided bias associated with missing or censored data.

Calculation of PEB Algorithm

Preclinical natural log (CA125) values from the cancer free women were used to fit the PEB algorithm. The population mean (μ), variance (V), and intraclass correlation (ICC; B) of ln (CA125) were estimated by using a linear random-effects model consistent with previously described natural history models of CA125 in cancer-free women[1,3,6,14,15]: Specifically, X_{ij} = ln (CA125) for person i at screen j = 1 . . . n follows the statistical model $X_{ij}|\mu_i \sim N(\mu_i, \sigma^2)$ with the person means μ_i varying in the population by $\mu_i \sim N(\bar{\mu}, \tau^2)$. This model implies that a single (n = 1) measurement has a population mean $\bar{\mu}$ variance

$$V = \sigma^2 + \tau^2 \quad (1)$$

and ICC

$$B_1 = \tau^2/(\tau^2 + \sigma^2) \quad (2)$$

The ICC measures the degree of similarity in biomarker levels between individuals compared with within individuals.[16] The multilevel package[17] (see ICC1) of R statistical software[18] was used to estimate B_1. At each screen, a PEB-deviation z score was calculated for each woman by comparing her current ln (CA125) value (Y) to a function that includes the sample mean of her $n \geq 0$ previous ln (CA125) values denoted $\bar{X}_i$, the ICC of that sample mean

$$B_n = \tau^2/(\tau^2 + \sigma^2/n) \quad (3)$$

and the population mean and variance. B_n can be written conveniently as

$$B_n = (n \times B)/(n \times B + (1 - B)) \quad (4)$$

The PEB Z score is positive whenever

$$Z = \frac{Y - [\mu \times (1 - B_n) + \bar{X}_i \times B_n]}{\sqrt{V \times (1 - B_1 \times B_n)}} \geq \text{cutoff} \quad (5)$$

where the cutoff, selected to control test specificity, can be estimated empirically from training data or by using percentiles of the standard normal distribution.

The PEB provides a family of screening rules, depending on the size of the screening history. At the first screen, when screen n = 0 (no screening history), B_n = 0, and the PEB rule equates to

Table 1. Comparison of the Performance of PEB and ST Rules at Various Positivity Thresholds by Using Blood Samples Collected From 701 Control Participants

Screening-Rule Threshold		Actual Specificity (%)	Total No. of FP Screens	No. of Unique Women Experiencing an FP Screen	
ST CA125 (U/mL)	PEB z Score			ST	PEB
35	2.8	99.1	35	21	29
30	2.23	98.2	68	38	59
25	1.69	95.8	161	64	134
20	1.25	90.4	368	135	287

Abbreviations: FP, false-positive screen; PEB, parametric empirical Bayes; ST, single threshold.

Table 2. Comparison of the Performance of the PEB and ST Rules at Various Positivity Thresholds by Using Blood Samples Collected From Case Participants (n = 44)

Screening-Rule Threshold		Actual Specificity (%)	No. of Cases Detected by Either the PEB or ST Rule That Were Detected Earlier, Later, or at the Same Screen by the PEB Rule Compared With the ST Rule			Difference in Years Between PEB- and ST-Positive Screens or Clinical Diagnosis Among Discordant Cases*				Average CA125 Value at Time of First Positive-PEB Screen
ST CA125 (U/mL)	PEB z Score		Earlier	Same	Later	Mean	Median	Max	Min	
35	2.8	99.1	9†	23	0	0.9	0.93	1.29	0.12	20.1
30	2.23	98.2	12†	23	0	0.93	0.93	2.1	0.12	18.9
25	1.69	95.8	16†	19	0	1.06	1.09	2.1	0.12	17.4
20	1.25	90.4	15†	22	1	1.59	1.31	3.42	−0.08‡	16

Abbreviations: Max, maximum; Min, minimum; PEB, parametric empirical Bayes; ST, single threshold.
*Discordant case was defined as a case identified earlier in time or missed by one of the decision rules.
†All tests were significant(McNemar test). P values from top to bottom were < .008, < .001, < .001, and < .002, respectively
‡Negative value reflects a case when the ST rule turned positive before to PEB.

$$z = (Y - \bar{\mu})/\sqrt{V} \quad (6)$$

which is equal to the ST rule. When *n* is large then $B_n \approx 1$ and

$$z \approx (Y - \bar{X}_i)/\sigma \quad (7)$$

which depends only on the individual screening history of a woman and her within-person variance (eg, an ST rule personalized to her natural history). Under most conditions, the PEB rule is a compromise between these two extremes.

Retrospective Screening

PEB parameters μ, V, and B for the PLCO population were estimated by using CA125 values from 3,817 screens performed among the 701 PLCO control participants. The estimated values $\bar{\mu} = 2.33$, $V = 0.232$, and $B_1 = 0.81$ were comparable with other populations.[3,6,14] The PEB and ST rules adjusted to the same specificity were retrospectively applied to serial CA125 values from the 44 cases and 701 control participants. We recorded the date of first positive test for the ST and PEB rules and whether the screening rules were concordant (ie, both rules identified a case at the same time or both rules failed to detect the
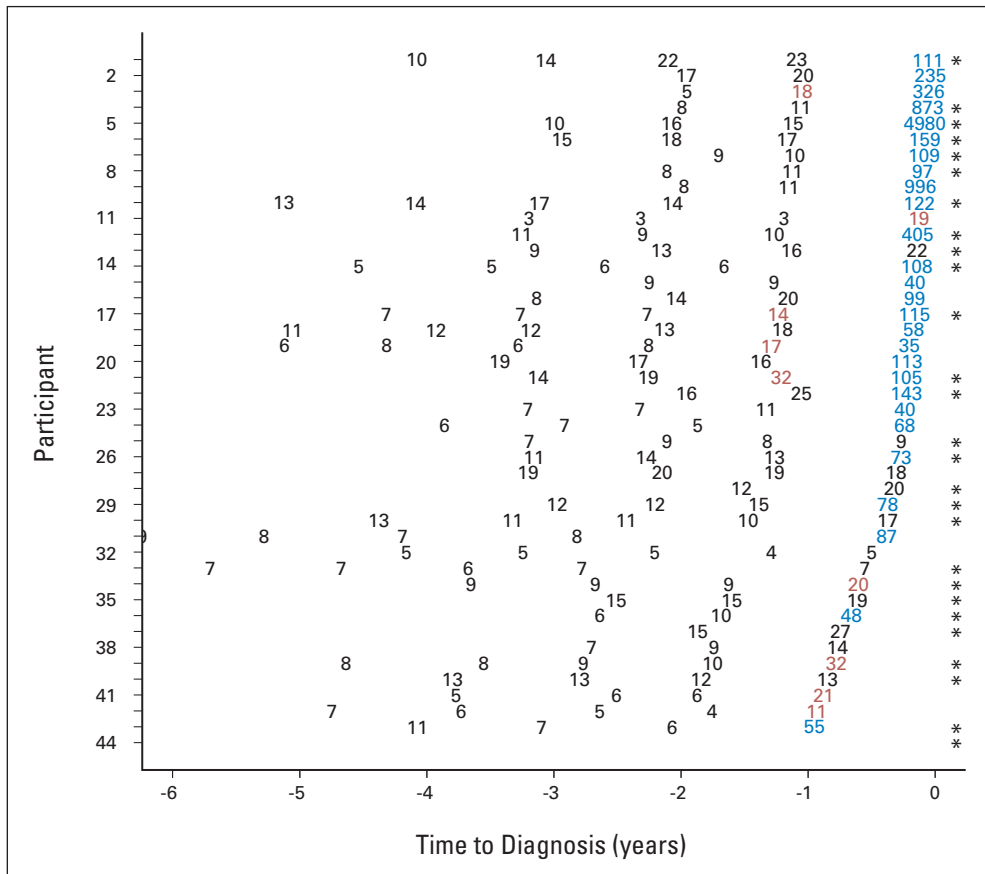

Fig 1. Summary of screening histories for the 44 incident PLCO (Prostate Lung Colorectal and Ovarian) Cancer Screening Trial ovarian cancer patient cases having at least two consecutive screens and whose proximate screen falls within a year of diagnosis. Numerical values represent CA125 concentrations. Participants were ordered by duration between the proximate screen and clinical diagnosis (horizontal axis). Concordant (blue) positive and (black) negative screens for parametric empirical Bayes (PEB) and single-threshold (ST) rules at 99% specificity and (red) discordant screen (PEB positive and ST negative) at 99% specificity used in the PLCO trial are shown. There were no PEB-negative and ST-positive screens. (*) Ovarian cancer of serous histology.

case) or discordant (ie, one rule identified a case earlier in time or detected a case missed by the other rule). The McNemar test was used to determine whether the PEB or ST performed better in cases with discordant decisions. These analyses were performed across a range of screening specificities.

RESULTS

A summary of the comparison of PEB and ST screening algorithms applied to blood samples from the 701 control participants at positivity thresholds corresponding to screening specificities from 90% to 99% are provided in Table 1. For each ST-rule cutoff, a corresponding PEB z score was calculated such that the number of false-positive tests identified among the cohort was identical with the number identified by the ST rule. These empirically derived PEB thresholds closely matched the theoretical values predicted by the quintiles of a normal distribution. Specifically, the reported PEB z scores of 2.80, 2.23, 1.69, and 1.61 corresponded to predicted specificities of 99.7%, 98.7%, 95.4%, and 89.5%, which closely matched the actual specificities measured in the cohort (99.1%, 98.2%, 95.8%, and 90.4%, respectively). This result suggested a high degree of accuracy of the statistical model for the PLCO data. At the PLCO study-protocol threshold of 35 U/mL, the ST rule identified 35 false-positive screens in 21 individual women, with an average of 1.66 positive tests per woman. Several women experienced multiple false-positive screens including five women with three false-positive screens and one woman with six false-positive screens (data not shown). In comparison, the PEB rule allocated the 35 false-positive tests among 29 unique women, with an average of 1.2 positive screens per woman with no woman experiencing more than two false-positive tests.

Performances of ST and PEB rules were separately evaluated in the 44 case samples (Table 2). The reduction of the positivity threshold (and hence the specificity) increased the number of cases detected by each algorithm. The PEB rule identified more cases at each of the specificity levels tested, and the proportion of cases detected earlier by the PEB increased as specificity was reduced. For example, 20% and 34% of cases were identified earlier by the PEB rule at the 99.1% and 90.4% specificity thresholds, respectively. Among discordant cases the PEB rule detected disease, on average, from 0.9 to 1.59 years earlier than the ST rule depending on the specificity level tested. The average CA125 value at the time of first positive PEB screen varied from 20.1 U/mL at 99.1% specificity to 16.0 U/mL at 90.4% specificity.

Detailed screening histories and screen-test classifications at the 99% specificity threshold (which corresponded to the PLCO study–defined cutoff of 35 U/mL) for the 44 cases are provided in Figure 1. The PEB and ST rules signaled cancer in 32 cases (72%) and 27 cases (61%), respectively. In nine cases (20%), the PEB rule signaled cancer, on average, 0.9 years earlier than the ST rule and, at a CA125 concentration of 20.1 U/mL, 42% lower than the ST-rule cutoff. Eighteen cases (56%) detected by the PEB rule were serous cancers, which was comparable with the proportion of serous cancers detected by the ST rule (16 of 27; 59%) and in the study population overall (27 of 44; 61%).

Paired CA125 values from the first and second screen and PEB- and ST-rule cutoffs at 99% specificity for the 44 cases are shown in Figure 2. The ST rule was positive for all screens that fell above the horizontal line (first or second screen $\geq$ 35 U/mL). At the first screen, at which no screening history was available, the PEB and ST rule
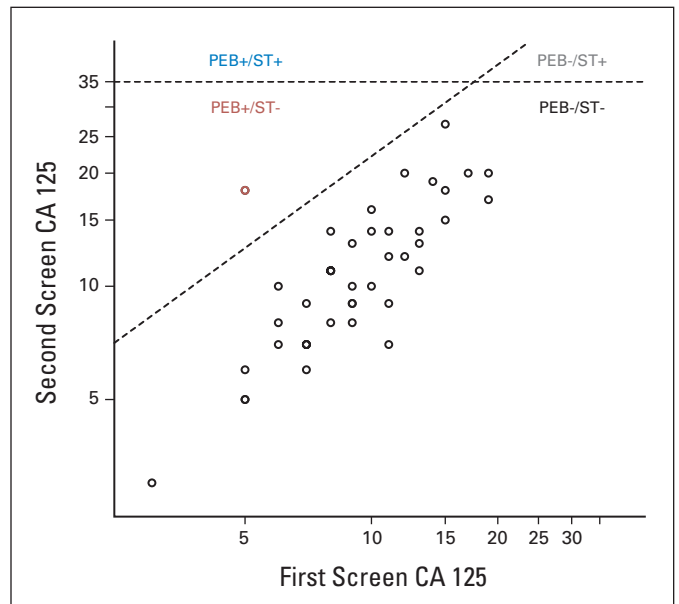

Fig 2. Representation of the single-threshold (ST) and parametric empirical Bayes (PEB) decision rules for all 44 PLCO cases at the second screen. Horizontal line $Y = 35$ represents the ST rule for 99% specificity at the second screen. The diagonal line defined by $Y = \bar{\mu} \times (1 - B_1) + X_i \times B_1 + z\sqrt{V \times (1 - B_1^2)}$ represents the 99% specificity PEB-rule threshold for the second screen (one previous screen) that was based on the ln (CA 125) level at the first screen indicated by X_i. The red circle represents case 3 in Figure 1 identified as abnormal by the PEB rule at the second screen with raw CA 125 = 18, which was below the ST-rule threshold.

positivity thresholds were identical. The diagonal line represents the PEB threshold for the second screen as a function of the CA125 level at screen 1. A discordant decision is represented in red. This woman was positive according to the PEB rule at screen 2 although her CA125 level was well below the ST rule threshold. Although not represented in the data, a woman may be positive according to the ST rule and negative according to the PEB algorithm. For example, a woman with a screen 1 CA125 level that exceeds the intersection of the two dotted lines (approximately 20 U/mL) will have a screen 2 PEB threshold that is higher than the cutoff for the ST rule. This effect is observed for a case listed in Table 1 which was detected earlier by the ST rule than the PEB algorithm at 90% specificity.

Discussion

Our results demonstrated that, for ovarian cancer screening by using CA125, the PEB longitudinal algorithm detected ovarian cancer earlier than an ST rule in a statistically significant and meaningful proportion of cases. This observation was consistent across a range of specificities relevant for a first-line test in a multimodal screening program. The proportion of cases detected earlier by the PEB algorithm increased as the specificity was reduced.

It could not be determined from our data that the earlier detection attributable to the PEB rule would necessarily lead to an earlier diagnosis or more importantly to diagnosis at an early stage or to greater mortality reduction. Even at the highest specificity tested, a confirmatory second test would be required before diagnostic surgery to maintain an acceptable positive predictive value (PPV). The average CA125 value at the time of a PEB-positive test is low. In light of the poor sensitivity of TVS observed in the PLCO trial,[19] there is concern

that many tumors would be too small to be detectable by TVU at these low biomarker levels. We recently reported that HE4, which is an ovarian cancer biomarker newly approved by the US Food and Drug Administration, has greater sensitivity than TVU for confirming cancer in patients with rising CA125[20] at a comparable specificity. It is possible that a sequential screening strategy using two markers may perform better than one that relies on currently available imaging as a second-line test.

The PLCO trial did not demonstrate a mortality reduction with ovarian cancer screening by using an ST rule.[13] The PEB algorithm detected ovarian cancer, on average, 0.9 to 1.59 years earlier than an ST rule depending on the specificity level tested. Results from the ongoing UKCTOCS (United Kingdom Collaborative Trial of Ovarian Cancer Screening), which uses the risk of ovarian cancer (ROCA) longitudinal algorithm, should provide insight as to whether earlier detection in the range noted leads to a stage shift and mortality reduction.

The PEB algorithm can be easily fit to any biomarker in which information about the behavior of the marker in unaffected individuals from a target population is available. The benefits of the PEB are greatest for markers that have a high ICC.[5] The ICC is mathematically equivalent to the correlation in marker levels between two time points. As demonstrated in Figure 2, the size of the discordant region representing PEB-positive and ST-negative screens was related to the strength of the correlation between markers levels at an earlier and later screen. Previous work has suggested that substantial gains are likely to require an ICC 0.5; however, for ovary cancer markers, an ICC of this magnitude is not uncommon.[5,21]

We used the PEB algorithm in this study because it is computationally simple, easy to implement, and nonproprietary. ROCA is an alternative longitudinal algorithm that uses a computationally complex change-point algorithm to identify women whose marker levels appear to follow a multiplicative exponential growth model and combines this information with other risk factors such as age to estimate the risk of having ovarian cancer for a woman at the time of the screen.[3,15] A key difference between the PEB and ROCA algorithms is the dynamic manner in which the ROCA accommodates time in the statistical model. Specifically, an equivalent increase in the marker level that takes place over 6 or 12 months is given the same significance by the PEB rule, whereas the more-rapid increase is given greater significance in the ROCA. However this property is likely to have little value in a screening protocol with fixed screening intervals. Unlike the PEB, the ROCA may be useful for markers that do not have a large ICC.

The ROCA has been shown to have a higher PPV than an ST rule when combined with TVU in a multimodal screening program.[22] However, more-direct comparisons of the performance characteristics of the ROCA and ST rule are complicated by the fact that ROCA triages women into different categories of follow-up. Women at highest risk are referred for immediate TVU, women at intermediate risk are recalled for repeat CA125 testing in 12 weeks, and normal risk women return to annual screening. The improvement in PPV associated with the ROCA is likely derived in part from repeat testing of intermediate-risk women and not a property unique to the risk calculation. Both the PEB and ST rules can be adapted to accommodate early recall by defining a lower threshold that identifies an intermediate-risk group. However, even with this adjustment, a direct comparison between the PEB rule and ROCA is not possible at this time. Previous studies that compared the ROCA and ST rule have emphasized an improvement in PPV associated with ROCA and have not reported on differences in lead time.

The PEB algorithm warrants additional testing in randomized ovarian cancer–screening trials. The algorithm is computationally simple, easy to implement, and readily adaptable to multiple and novel markers and screening programs that incorporate early recall. Indeed, the PEB algorithm is currently being evaluated in a phase I screening trial (National Institutes of Health/National Cancer Institute; Award No. P50 CA083636). The PEB may prove useful for screening for other cancers for which biomarkers are available.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

AUTHOR CONTRIBUTIONS

REFERENCES

1. McIntosh MW, Urban N: A parametric empirical Bayes method for cancer screening using longitudinal observations of a biomarker. Biostatistics 4:27-40, 2003

2. McIntosh MW, Urban N, Karlan B: Generating longitudinal screening algorithms using novel biomarkers for disease. Cancer Epidemiol Biomarkers Prev 11:159-166, 2002

3. Skates S, Pauler D, Jacobs I: Screening based on the risk of cancer calculation from Bayesian hierarchical change point and mixture models of longitudinal markers. J Amer Stat Assoc 96:429-439, 2001

4. Lin H, Turnbull B, Mcculloc CE, et al: Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer. J Amer Stat Assoc 97:53-65, 2002

5. Sato AH, Anderson GL, Urban N, et al: Comparing adaptive and non-adaptive algorithms for cancer early detection with novel biomarkers. Cancer Biomark 2:151-162, 2006

6. Crump C, McIntosh MW, Urban N, et al: Ovarian cancer tumor marker behavior in asymptomatic healthy women: Implications for screening. Cancer Epidemiol Biomarkers Prev 9:1107-1111, 2000

7. Sottas PE, Baume N, Saudan C, et al: Bayesian detection of abnormal values in longitudinal biomarkers with an application to T/E ratio. Biostatistics 8:285-296, 2007

8. Sottas PE, Saudan C, Schweizer C, et al: From population- to subject-based limits of T/E ratio to detect testosterone abuse in elite sports. Forensic Sci Int174:166-172, 2008

9. Van Renterghem P, Van Eenoo P, Sottas PE, et al: A pilot study on subject-based comprehensive steroid profiling: Novel biomarkers to detect testosterone misuse in sports.Clin Endocrinol (Oxf)

10.1111/j.1365-2265.2011.03992.x [epub ahead of print on February 2, 2011]

10. Sottas PE, Robinson N, Saugy M, Niggli O: A forensic approach to the interpretation of blood doping markers. Law, Probability and Risk 7:191-210, 2008

11. Gohagan J, Marcus P, Fagerstrom R, et al: Baseline findings of a randomized feasibility trial of lung cancer screening with spiral CT scan vs chest radiograph: The Lung Screening Study of the National Cancer Institute. Chest 126:114-121, 2004

12. Gohagan JK, Prorok PC, Hayes RB, et al: The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: History, organization, and status. Control Clin Trials 21:251S-272S, 2000 (6 suppl)

13. Buys SS, Partridge E, Black A, et al: Effect of screening on ovarian cancer mortality: The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. JAMA 305:2295-2303, 2011

14. Pauler DK, Menon U, McIntosh M, et al: Factors influencing serum CA125II levels in healthy postmenopausal women. Cancer Epidemiol Biomarkers Prev 10:489-493, 2001

15. Skates SJ, Xu FJ, Yu YH, et al: Toward an optimal algorithm for ovarian cancer screening with longitudinal tumor markers. Cancer 76:2004-2010, 1995 (10 suppl)

16. Bliese PD: Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis, in KJKSWK (ed): Multilevel Theory, Research, and Methods in Organizations. San Francisco, CA, Jossey-Bass, 2000, pp 349-381

17. Bliese P: Multilevel: Multilevel functions. R package 2.3 (ed 2008)

18. Team RDC. A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing, 2009

19. 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 (Phila) 4:375-383, 2011

20. Urban N, Thorpe JD, Bergan LA, et al: Potential role of HE4 in multimodal screening for epithelial ovarian cancer. J Natl Cancer Inst 103:1630-1634, 2011

21. Cramer DW, Bast RC Jr, Berg CD, et al: Ovarian cancer biomarker performance in prostate, lung, colorectal, and ovarian cancer screening trial specimens. Cancer Prev Res (Phila) 4:365-374, 2011

22. Menon U, Skates SJ, Lewis S, et al: Prospective study using the risk of ovarian cancer algorithm to screen for ovarian cancer. J Clin Oncol 23:7919-7926, 2005

■ ■ ■