

Interpretation of Single and Serial Measures of HE4 and CA125 in Asymptomatic Women at High Risk for Ovarian Cancer

Nicole Urban¹, Jason Thorpe¹, Beth Y. Karlan³, Martin W. McIntosh¹, Melanie R. Palomares⁴, Mary B. Daly⁵, Pamela Paley², and Charles W. Drescher^{1,2}

Abstract

Background: Human epididymis protein 4 (HE4) is approved for clinical use with CA125 to predict epithelial ovarian cancer in women with a pelvic mass or in remission after chemotherapy. Previously reported reference ranges for HE4 are inconsistent.

Methods: We report positivity thresholds yielding 90%, 95%, 98%, and 99% specificity for age-defined populations of healthy women for HE4, CA125, and Risk of Ovarian Malignancy Algorithm (ROMA), a weighted average of HE4 and CA125. HE4 and CA125 were measured in 1,780 samples from 778 healthy women aged >25 years with a documented deleterious mutation, or aged >35 years with a significant family history. Effects on marker levels of a woman's age, ethnicity, and epidemiologic characteristics were estimated, as were the population-specific means, variances, and within- and between-woman variances used to generate longitudinal screening algorithms for these markers.

Results: CA125 levels were lower with Black ethnicity ($P = 0.008$). Smoking was associated with higher HE4 ($P = 0.007$) and ROMA ($P < 0.019$). Continuous oral contraceptive use decreased levels of CA125 ($P = 0.041$), and ROMA ($P = 0.12$). CA125 was lower in women age ≥ 55 , and HE4 increased with age ($P < 0.01$), particularly among women age ≥ 55 .

Conclusions: Because of the strong effect of age on HE4, thresholds for HE4 are best defined for women of specific ages. Age-specific population thresholds for HE4 for 95% specificity ranged from 41.4 pmol/L for women age 30 to 82.1 pmol/L for women age 80.

Impact: Incorporation of serial marker values from screening history reduces personalized thresholds for CA125 and HE4 but is inappropriate for ROMA. *Cancer Epidemiol Biomarkers Prev*; 21(11); 2087–94. ©2012 AACR.

Introduction

Human epididymis protein 4 (HE4) is an epithelial ovarian cancer (EOC) serum marker that was identified originally as a candidate early detection marker (1). Since then it has been shown to be potentially useful for remission monitoring (2–4) and has been cleared by the U.S. Food and Drug Administration (FDA) for that use. It has also been shown to perform as well or better than CA125 in distinguishing benign from malignant tumors in wom-

en with a pelvic mass (5–7). The Risk of Ovarian Malignancy Algorithm (ROMA) which includes CA125 as well as HE4 (8) has been cleared by the FDA for evaluation of a pelvic mass. HE4 has also been reported to contribute to identification of malignancy in women with nonspecific abdominal/pelvic symptoms (9). Most recently, it has been reported to provide detectable signal about a year before diagnosis (10, 11), renewing interest in it as an early detection marker. We have recently reported that HE4 outperforms imaging as a second-line screen when rising CA125 is used as a first-line screen (12).

The efficacy of screening for EOC has not been shown. Use of both imaging and CA125 (using a threshold of 35 U/mL for all women) annually in women aged 55 to 74 leads to unnecessary surgery (13) without reducing mortality (14). However, a multimodal strategy using rising CA125 annually to select women for imaging is yielding acceptable positive predictive value (PPV) in an efficacy trial in the United Kingdom where the longitudinal Risk of Ovarian Cancer Algorithm (ROCA) (15) is used to measure rising CA125 (16). The ROCA has not been applied to HE4. An alternative strategy for using novel

Authors' Affiliations: ¹Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington; ²Pacific Gynecology Specialists, Seattle, Washington; ³Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, California; ⁴Division of Clinical Cancer Genetics, City of Hope Comprehensive Cancer Center, Duarte, California; and ⁵Division of Population Science, Fox Chase Cancer Center, Philadelphia, Pennsylvania

Corresponding Author: Nicole Urban, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, M2-B230, P.O. Box 19024, Seattle, WA 98109. Phone: 206-667-6774; Fax: 206-667-7264; E-mail: nurban@fhcrc.org

doi: 10.1158/1055-9965.EPI-12-0616

©2012 American Association for Cancer Research.

markers in a longitudinal algorithm is the Parametric Empirical Bayes (PEB) longitudinal algorithm (17). Use of the PEB algorithm has been reported for CA125 previously (17, 18), and it can be readily adapted for use with HE4 because it is fit using data from healthy women only.

HE4 is in clinical use but reference ranges for HE4 have been reported only very recently and remain uncertain because of inconsistency. Park and colleagues report the 97.5% upper reference limit for HE4 to be 33.2 pmol/L for a young Asian population (19), low compared with the median value of 48 pmol/L in healthy controls reported by the same authors for an Asian hospital population (20). Moore and colleagues report the upper 95th percentile of HE4 to be 89 pmol/L for premenopausal women, 128 pmol/L for postmenopausal women, and 115 pmol/L for all women, lower in pregnancy and rising with age (21). Other investigators have used positivity thresholds for HE4 of 70 pmol/L (22) and 150 pmol/L (23) in clinical validation studies.

Here we report HE4 positivity thresholds yielding 90%, 95%, 98%, and 99% specificity, that account for age, derived from analysis of 1,780 samples from 778 healthy women aged ≥ 25 years with a documented deleterious mutation, or aged ≥ 35 years with a significant family history. We also report parameters characterizing the within- and between-components of variance for HE4 to allow researchers to calculate the thresholds for HE4 when adjusting for a woman's screening history as well as her age using the PEB rule (18). Whenever the total variance of a marker is dominated by between-women differences, the PEB decision rule will maintain overall test specificity but may achieve better sensitivity, longer lead time, and lower thresholds for the majority of women compared with a single threshold (ST) rule (18, 24). Interpretation of serial measures of HE4 using a longitudinal algorithm has not been previously reported. For comparison we also report the PEB-relevant parameters for CA125 and ROMA in this population.

Materials and Methods

Women enrolled in the Novel Markers Trial (NMT) through February 29, 2012 were eligible for the current study if at the time of enrollment they were aged ≥ 25 years and reported having tested positive for a deleterious mutation in *BRCA1* or *BRCA2*, or aged ≥ 35 years with a significant family history and were not diagnosed with EOC during follow-up. The NMT is a 2-arm randomized multi-institutional Phase I screening trial sponsored by the National Cancer Institute (25). The NMT introduces HE4 as either a first- or second-line screen in a multimodal screening strategy that includes CA125, HE4, and transvaginal ultrasound (TVU) (26). Participants completed a baseline questionnaire that provided information about ovarian cancer family history and other risk factors, and contributed up to 5 blood samples about 6 months apart. HE4 and CA125 were measured in 1,780 samples from 778 healthy women on the Abbott Architect automated platform in a Clinical Laboratory Improvement Amend-

ments-approved laboratory using FDA-approved kits. The ROMA Predictive Index (PI) was calculated using previously defined pre- and post-menopause formulae (8). We used $[PI = -12.0 + 2.38 * \ln(HE4) + 0.0626 * \ln(CA125)]$ for women under age 50, and $[PI = -8.09 + 1.04 * \ln(HE4) + 0.732 * \ln(CA125)]$ for the remaining women.

To address interpretation of single and longitudinal measures of markers, we first stratified ages of all women by 4 clinically relevant age strata: age <45 , $45 \leq$ age <55 , $55 \leq$ age <65 , and age ≥ 65 . Natural-log (ln) transformed marker values were used in all analyses and then transformed back to their raw scale for reporting purposes. Because the ROMA PI is a linear combination of log-transformed CA125 and HE4 values (8), no additional transformations were needed and results are reported on its raw scale.

A linear regression model associating marker concentration (Y) to covariates was fit using a generalized estimating equations clustering (27) on participant identifier to adjust for varying numbers of observations per participant to estimate the model $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \beta + \varepsilon$, where ε is a residual normal distribution with mean 0 and residual variance V . The coefficients were used to determine the magnitude of effect for each epidemiologic covariate. The regression predictors included epidemiologic covariates, and main effects of the age strata, age as a continuous variable, and also an interaction between age and age stratum to allow the effect of age to differ among the different age-defined strata. The regression line was used to produce reference ranges for HE4, which has a strong trend with age within each age stratum. Age-defined population thresholds are calculated from the fitted values and residual standard error on the basis of an assumption that the residual errors are approximately normally distributed. For CA125, where levels are constant in the younger and older women, age-defined population thresholds were determined from empirically defined percentiles of the marker within those age ranges.

The marker threshold for an individual woman can be personalized if screening history is known, including age and marker value at each screen. To control for screening history we used the PEB screening rule, taking advantage of its ability to accommodate covariates (17, 18). We assume that each woman's marker levels may systematically differ, on average by a constant amount, from the regression equation above, and marker concentrations will vary around this person-specific regression line with residual variance S , where $S < V$. The intraclass correlation (ICC) is defined here by $ICC = 1 - S/V$, and $0 < ICC < 1$. The ICC is estimated by first computing the regression's residuals, $Z = Y - (\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)$, then calculating the ICC using the ICC1 function of the "multilevel" package of R (28). To approximate thresholds controlling for both age and covariates we use B to represent the ICC of Z and use $B_n = (n)/(1 + (n - 1) \cdot B)$ to represent the ICC of a sample average of " n " independent values of Z , denoted \bar{Z} (17, 18). The PEB threshold is calculated from

the regression equation and the within- and between-woman variance parameters using the following formula

$$\text{Threshold} = \underbrace{\bar{Z}(B_n)}_{\substack{\text{Person specific} \\ \text{deviation}}} + \underbrace{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots}_{\substack{\text{Covariate adjustment} \\ \text{Expected value adjusted for screening history}}} + \underbrace{C_\alpha \sqrt{V(1 - BB_n)}}_{\substack{\text{quantile Residual standard} \\ \text{deviation} \\ \text{Offset}}}$$

where C_α is the alpha quantile, a standardized normal distribution (e.g., 1.64 for 95% specificity). See McIntosh and Urban (17) for derivation and details. Parameters required for calculating the PEB thresholds for HE4 and CA125 are provided in the Results section. With longer screening histories, B_n increases to a maximum of 1 and the intercept term equals a woman's unique individual value level \bar{Z} . When " $n = 0$ " then $B_n = 0$ and the PEB threshold defaults to the population-defined linear regression values. The person-specific residual variance also decreases as screening history accumulates, meaning that the reference range for a woman shrinks in size over time. Theoretical results show that at comparable population-wide specificities, the PEB uniformly improves upon rules that ignore screening history and improves over simpler rules that make decisions on the basis of simple change-from-baseline (17, 18).

Results

Baseline characteristics of NMT participants included in this report are reported in Table 1. The majority of women were aged 45 to 64, Caucasian, non-Hispanic, parous, non-smoking, non-hysterectomized and non-users of hormone replacement therapy (HRT), and most did not have a prior tubal ligation. A deleterious mutation in a *BRCA1* or *BRCA2* gene was reported by 17.7% of women included in this study. A personal history of breast cancer was reported by 17.2% and a family history of breast cancer was reported by 75.3% of women. A family history of ovarian cancer was reported by 43.1%, and Ashkenazi Jewish lineage was reported by 19.5% of women. Current oral contraceptive (OC) use was reported by 5.0% of women; half of these reported continuous OC use resulting in cessation of periods.

Effects of age and age-adjusted effects of other covariates on mean marker levels are reported in Table 2. We first evaluated the effect of age on the marker levels within each age stratum. CA125 levels are lower for women with age ≥ 55 than in age < 45 ($P < 0.001$), but levels are constant and unchanging with respect to age within these categories ($P = 0.49$ for age < 45 and $P = 0.797$ for age ≥ 55). Rapid changes in CA125 were found in the strata defined by $45 \leq \text{age} < 55$, with a decline of 30% over the 10-year period ($P = 0.006$).

For HE4 an increasing trend was found in all age strata but the most dramatic change was found in women after

Table 1. Number and percentage of women and samples included in analyses by age, ethnicity, and personal characteristics

Variable	Participants	Samples contributed
Total count	778 (100.0%)	1780 (100.0%)
Age (mean, SD)	52.3 (10.9)	53.5 (10.9)
Age <45	190 (24.4%)	370 (20.8%)
45 ≤ Age < 55	276 (35.5%)	635 (35.7%)
55 ≤ Age < 65	215 (27.6%)	513 (28.8%)
Age ≥65	97 (12.5%)	262 (14.7%)
Non-White race	78 (10.0%)	159 (8.9%)
Black	10 (1.3%)	18 (1.0%)
Asian	22 (2.8%)	42 (2.4%)
Other	46 (5.9%)	99 (5.6%)
Hispanic	36 (4.6%)	66 (3.7%)
Ashkenazi Jewish	152 (19.5%)	325 (18.3%)
Parous	520 (66.8%)	1,186 (66.6%)
Hysterectomy	67 (8.6%)	156 (8.8%)
Tubal ligation	103 (13.2%)	253 (14.2%)
Current smoker	26 (3.3%)	58 (3.3%)
Current HRT use	101 (13.0%)	252 (14.2%)
Current OC use	39 (5.0%)	91 (5.1%)
Discontinuous OC	21 (2.7%)	48 (2.7%)
Continuous OC	18 (2.3%)	43 (2.4%)
Family history of breast cancer	586 (75.3%)	1340 (75.3%)
Family history of ovarian cancer	335 (43.1%)	792 (44.5%)
Personal history of breast cancer	134 (17.2%)	254 (14.3%)
Deleterious mutation	138 (17.7%)	261 (14.7%)

age 55. Before age 45, Ln HE4 concentrations elevate by an average of 0.0082 per year, or $\exp(0.0082 \times 10) \times 100 = 8.5\%$ per decade on the raw scale ($P = 0.013$), but they elevate by 0.0158 per year faster than the reference group, or $\exp[(0.0158 + 0.0082) \times 10] \times 100 = 27.1\%$ per decade on the raw scale after age 55. Although the slope among women with $45 \leq \text{age} < 55$ was just 2.7%, it did not differ significantly from that of women age < 45 ($P = 0.40$), suggesting that these age categories could potentially be combined when interpreting HE4. An analysis combining these 2 younger periods finds HE4 with a slope of 0.00602, or 6.2% change per decade on the raw scale before age 55. We did not combine these groups in our analyses, however.

For ROMA, which includes both CA125 and HE4, an increasing trend was apparent in the reference group (age < 45 , $P = 0.012$), and the slope was higher among the older age groups though the difference was only significant in women $45 \leq \text{age} < 55$ ($P < 0.005$). However, given the definition of ROMA, especially because its definition is different for women before and after menopause (age 50 here), it is difficult to interpret the reference ranges of ROMA over time. We provide it here for reference.

Table 2. Effect of age, ethnicity, and personal characteristics on marker concentrations

Analysis	Covariate	Sample size	CA125		HE4		ROMA	
			Coefficient	P value	Coefficient	P value	Effect (log scale)	P value
Age-related effects on ln marker concentrations ^a								
Intercept	Age <45 (Reference)	370	2.39859	<0.001	2.99346	<0.0005	-4.72542	<0.0005
	45 ≤ age < 55	635	1.92218	0.007	0.25002	0.412	-6.25026	<0.0005
	Age ≥55	775	-0.35088	0.365	-0.98948	<0.0005	0.21848	0.583
Trends	Age <45 (Reference)	370	0.00524	0.490	0.00820	0.013	0.01985	0.012
	45 ≤ age < 55	635	-0.04155	0.006	-0.00551	0.398	0.13377	<0.0005
	Age >55	775	-0.00224	0.797	0.01579	<0.0005	0.00730	0.420
Covariate effects on percentage change of marker concentrations			% Change	P value	% Change	P value	% Change	P value
Age-adjusted effects								
	Parous	1,186	4.0%	0.355	0.4%	0.867	-0.004	0.935
	Non-White race	159	6.3%	0.371	3.6%	0.357	0.126	0.122
	Black	18	-26.6%	0.008	-3.3%	0.718	-0.203	0.329
	Asian	42	20.4%	0.112	7.7%	0.474	0.317	0.142
	Other	99	7.5%	0.417	2.9%	0.454	0.094	0.200
	Hispanic	66	19.10%	0.097	1.80%	0.683	0.091	0.388
	Ashkenazi Jewish	325	-7.4%	0.163	0.9%	0.702	-0.011	0.850
	Prior Tubal ligation	253	0.4%	0.935	-0.3%	0.928	-0.015	0.79
	Current smoker	58	0.3%	0.977	21.0%	0.007	0.276	0.019
	Current OC use	91	-13.0%	0.100	-1.8%	0.632	-0.115	0.225
	Discontinuous OC	48	-5.4%	0.641	2.4%	0.649	0.04	0.766
	Continuous OC	43	-18.9%	0.041	-5.8%	0.217	-0.265	0.012
	Personal history of breast cancer	584	-3.0%	0.497	3.9%	0.105	0.046	0.335

NOTE: The reference group is women under age 45. Age-related effects on ln(marker levels) are presented in the top half of the table. The bottom half presents the effects of other covariates in terms of percentage change of marker levels. The marker value expected for an individual woman can be calculated using this table using instructions in the table footnote. Sample size refers to blood draws not number of women. To approximate the marker value expected for a woman of a specific age, first approximate the baseline LN marker level for the woman's age as follows. Add the intercept terms (reference group plus the term from the age appropriate category) to the age multiplied by the age-related trend (reference group trend + age appropriate category trend). For example, baseline LN CA 125 level for a woman of age 50 will equal $Y = (2.39859 + 1.92218) + 50 * (0.00524 - 0.04155) = 2.50$, or 12.23 U/mL on the raw scale. To adjust the level on the raw scale for covariates, increase or decrease the calculated raw scale using the approximated effect size for that covariate. For example, a woman of age 50 who is of Black ethnicity will be expected to have CA 125 levels 26.6% lower, or 3.25 U/mL lower.

^aCoefficients and P values obtained from a linear regression using the formula:

$\ln(\text{marker level}) \sim \text{intercept} + \text{age} + I(45 < \text{age} < 55) + \text{age} * I(45 < \text{age} < 55) + I(\text{age} > 55) + \text{age} * I(\text{age} > 55)$, where indicator variables "I (. . .)" are set to 1 when the condition is met and 0 otherwise. Together, the intercept [e.g., "I(45 < age < 55)"] and trend [e.g., "age * I(45 < age < 55)"] terms allow for independent linear trends to be estimated within each age range.

Among all other covariates, concentrations of CA125 were lower in women with Black ethnicity (26.6%; $P = 0.008$) and with continuous OC use (18.9%; $P = 0.041$). Concentrations of HE4 were higher in current smokers (21%; $P = 0.007$). ROMA was higher in smokers ($P < 0.019$) and lower in women reporting continuous OC use ($P = 0.012$).

Table 3 reports age-defined population thresholds. For CA125, where concentrations are constant within the age ranges age <45 and age ≥55, a single reference range is appropriate for all women in those age ranges, and thresholds are determined empirically and without distributional assumptions. However, for HE4, because

it changes so dramatically with age we provide the thresholds at specific ages predicted from the regression equation. However, because the thresholds are linear on the log scale, one can predict reference ranges for any intermediate age by using linear interpolation on the appropriate scale; for example, the 95th percentile for age 65 is $\exp [0.5 \ln (50.8) + 0.5 \ln (64.6)] = 57.3$. Note that only age is accounted for in Table 3. Applying the appropriate percentage reduction using the effect sizes from Table 2 can approximate adjustments for other covariates (e.g., a 21% increase in HE4 for a current smoker). However, because sample sizes for the significant predictors of CA125 and HE4 are small, one must

Table 3. Age-defined population-specific thresholds for HE4, CA125, and the ROMA Predictive Index yielding 90%, 95%, 98%, and 99% specificity

Marker	Age	Mean	90% Specificity	95% Specificity	98% Specificity	99% Specificity
CA125 (U/mL)	Age < 45	13.428	26.306	31.830	39.447	45.513
	45 ≤ age < 55	12.023	23.553	28.500	35.320	40.751
	Age ≥ 55	9.370	18.356	22.211	27.526	31.759
HE4 (pmol/L)	30	25.522	37.216	41.417	46.714	50.617
	40	27.703	40.398	44.957	50.707	54.944
	50	29.312	42.744	47.568	53.652	58.135
	60	31.306	45.651	50.804	57.302	62.089
	70	39.796	58.033	64.583	72.843	78.929
	80	50.590	73.772	82.098	92.599	100.335
ROMA Predictive Index	30	0.016	0.035	0.044	0.056	0.067
	40	0.020	0.043	0.054	0.069	0.081
	50	0.037	0.081	0.101	0.130	0.154
	60	0.056	0.123	0.154	0.197	0.233
	70	0.074	0.162	0.202	0.259	0.306
	80	0.097	0.212	0.265	0.340	0.402

NOTE: For CA 125, the reference range is applicable to all women in the age range indicated. Thresholds for HE4 for a woman whose age is not presented in the table can be calculated by linear interpolation between the $\ln(\text{threshold})$ values.

be cautious as one may anticipate a wide confidence interval around their effects.

For HE4, age-specific population thresholds for 95% specificity ranged from 41.4 pmol/L for women age 30 to 82.1 pmol/L for women age 80. The magnitude of the effect of age on HE4 means and thresholds can be seen in Table 3. The 10-year span from age 30 to 40 increases the 95th percentile for HE4 by only 4 points, but a 10-year span from age 60 to 70 increases HE4 by nearly 14 points. The association with age for HE4 has been previously reported, but not the distinct behavior on either side of 55 years (12, 21, 29). ROMA thresholds similarly increase with age. For CA125, thresholds yielding 95% specificity are 31.8 U/mL, 28.5 U/mL, and 22.2 U/mL for women age <45, 45 ≤ age < 55, and age ≥55, respectively.

Incorporation of serial marker values from screening history potentially reduces thresholds for CA125 and HE4 but was found to be inappropriate for ROMA. Thresholds personalized for both covariates and screening history are obtained using the PEB algorithm for $\ln(\text{HE4})$ and $\ln(\text{CA125})$. We do not include ROMA in the PEB calculation, as it is not recommended for use with healthy women for which longitudinal screening algorithms will be applied, and because applying a longitudinal algorithm to the ROMA as a means to combine markers is not necessarily optimal or superior to applying the PEB algorithms to CA125 and HE4 separately and defining positivity as either marker positive. The PEB parameters V and B , from the equation, are shown in Table 4 [the ICC (or B) is equal to B_n when $n = 1$]. We also provide B_n for scenarios when screening history is between 0 and 6 screens. These values and the results of Table 2 can be used to calculate the

threshold for any screening history for any woman. For HE4, the residual variance V and B is computed from the regression equation represented in the footnote to Table 2. However, for CA125 these values were computed separately using data within each age stratum, and parameters are reported separately for each stratum. Note because CA125 values change so dramatically in the 45 ≤ age < 55 stratum, it is not recommended to use the PEB rule during this time, but we report them here for completeness. The ICC for HE4 is smaller than for CA125 suggesting that its screening history is less informative than that of CA125, but its value exceeds 0.5, suggesting that personalizing thresholds may have a meaningful effect compared with ignoring screening history (18, 24).

The effectiveness of controlling for screening history is shown by the "% reduction in offset" in Table 4, referring to the percentage that the "offset" in the PEB expression, which represents the deviation from the expected value that is tolerated before a screen is declared positive, was reduced. The goal of the PEB rule is to provide a narrower predicted reference for each woman while maintaining overall test specificity. The narrower reference range will facilitate earlier detection of a tumor: a reference range that is half the width can detect tumors when markers elevate half the amount. As seen in Table 4, by the 4th screen the reference range for HE4 is on average 30% narrower when controlling for screening history than when ignoring prior marker values. CA125, because of its higher ICC, can achieve an even greater reduction in reference range size. Note that the ICC compares screening rules that ignore history to those using the same marker accounting for screening history; the ICC does

Table 4. Variance parameters and PEB-specific weights to obtain personalized thresholds by controlling for natural history

Marker	Age group	V	Statistic	Length of screening history (n)						
				0	1	2	3	4	5	6
CA 125	<45	0.31	B_n	0	0.776	0.874	0.912	0.933	0.945	0.954
			Reduction	0	37%	43%	46%	47%	48%	49%
	45–54	0.32	B_n	0	0.709	0.83	0.88	0.907	0.924	0.936
			Reduction	0	30%	36%	39%	40%	41%	42%
	≥55	0.22	B_n	0	0.785	0.88	0.917	0.936	0.948	0.956
			Reduction	0	38%	44%	47%	49%	49%	50%
HE4	All ages	0.09	B_n	0	0.572	0.728	0.8	0.843	0.87	0.889
			Reduction	0	18%	24%	26%	28%	29%	30%

NOTE: V represents the population-wide residual variance of the marker. B_n represents the ICC of the subject's mean screening history. The "Reduction" statistic represents the reduction in the size of the deviation from the woman's expected marker value required to achieve a positive test result compared with a single threshold rule.

not measure the quality of the marker otherwise. As seen in Table 4, the incremental benefit of controlling for screening history diminishes over time. For CA125 with women age ≥ 55 , 2 previous screens can reduce the offset by 43% but a 49% reduction is not obtained until 6 screens.

Discussion

Interpretation of single and serial measures of HE4 and CA125 in women at high risk for ovarian cancer can be summarized in thresholds for positivity at first and subsequent screens. We investigated characteristics that might affect population thresholds for use at first screens. For HE4, we found concentrations to depend on age, with a quite dramatic change over time after age 55. The thresholds reported by us and others [e.g., Moore and colleagues (21) or Park and colleagues (20)] are difficult to compare owing to differences in the ages of the populations, in addition to potential differences in race. We have shown that adequately controlling for effects of age on HE4, given its rapid rise after age 55, will require reference ranges that depend on an individual woman's specific age. This is different from CA125 where all women in broad age categories share the same CA125 mean levels. For CA125, if 95% specificity is desired population thresholds are 31.8 U/mL and 22.2 U/mL for women age <45 and ≥ 55 , respectively. For HE4, it is useful to categorize women more finely because HE4 increases steadily with age. The age-specific population thresholds for HE4 for 95% specificity nearly double over the age range studied, increasing from 41.4 pmol/L for a 30-year-old woman to 82.1 pmol/L for an 80-year-old woman. Although our analysis focused on high-risk women, it is likely to be generalizable to low-risk women because most of the high-risk women in this study were at modestly increased risk, and because prior work suggests that risk does not influence marker levels (30).

Reference ranges for ROMA may provide a means to screen women by combining CA125 and HE4. However,

although ROMA PI was derived using statistically optimal procedures for use in a cross-sectional study (8) that are appropriate for use there, its optimality for use in a longitudinal study has not been shown. It may be inferior to other alternative approaches such as using the markers separately then making decisions on the basis of whether one or both are elevated. In its current form, ROMA is not appropriate for use in early detection among asymptomatic women.

A limitation of this study is that the healthy women included in the NMT were not routinely characterized for the presence or absence of uterine leiomyoma, which may cause elevation in CA125. To address this possibility, we reviewed medical charts as well as imaging reports for all 53 women who had NMT-protocol-indicated TVU to identify any leiomyoma occurring in study participants with elevated markers. Four such women were identified. In all 4, TVU was performed because of transitory elevation in CA125. The presence of a leiomyoma appears to increase within-woman variability in CA125 and to cause occasional elevation above 99% thresholds.

Results of this study are more relevant to research than to clinical application of markers because screening for ovarian cancer is currently not recommended. It is especially important that primary care physicians understand that there is currently no evidence to support screening, especially using imaging which is often their preferred modality (31). The Prostate Lung Colon and Ovary (PLCO) trial in the United States failed to demonstrate a reduction in EOC-specific mortality with annual screening using both CA125 and TVU concurrently (14). At the prevalence screen in the PLCO trial, rates of surgery were 2.6 times higher for TVU-positive only than for CA125-positive only (32). Reports from the United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) suggest that CA125 may outperform TVU in screening for EOC when a longitudinal algorithm is used to measure rising CA125 (16). The sensitivity, specificity,

and PPV for all primary invasive EOCs identified at the prevalence screen in the UKCTOCS were 89.5%, 99.8%, and 35.1%, respectively, for the multimodal strategy using CA125 to select women for TVU, and 75.0%, 98.2%, and 2.8%, respectively, for TVU alone (16). Although results for the CA125 arm are promising, it is not yet known whether the sequential multimodal screening strategy will reduce mortality. The clinical benefit of using HE4 and the PEB algorithm for ovarian cancer screening has also not been documented. Data on clinical outcomes of the NMT cohort are currently being gathered. HE4 and CA125 interpreted by a longitudinal algorithm may have a role to play in EOC screening trials in the future. If research were to identify markers with improved operating characteristics, the population of patients with high risk (mutation carriers and those with significant pedigrees) might one day benefit.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: N. Urban, J.D. Thorpe, B.Y. Karlan, M.R. Palomares, C.W. Drescher

Development of methodology: N. Urban, J.D. Thorpe, B.Y. Karlan, M.W. McIntosh

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): N. Urban, B.Y. Karlan, M.R. Palomares, M.B. Daly, P.J. Paley, C.W. Drescher

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): J.D. Thorpe, B.Y. Karlan, M.W. McIntosh, M.R. Palomares, M.B. Daly, C.W. Drescher

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): J.D. Thorpe, B.Y. Karlan

Writing, review, and/or revision of the manuscript: N. Urban, J.D. Thorpe, B.Y. Karlan, M.W. McIntosh, M.R. Palomares, M.B. Daly, P.J. Paley, C.W. Drescher

Study supervision: N. Urban, B.Y. Karlan, M.R. Palomares

Acknowledgments

The authors gratefully acknowledge helpful comments provided during review of manuscript from Beth Schodin and Barry Dowell of Abbott Diagnostics. The authors also gratefully acknowledged the support for the Translational and Outcomes Research Laboratory from NIH/NCI U01 CA152637 and the Canary Foundation, support for clinical centers from the Marsha Rivkin Center for Ovarian Cancer Research and Canary Foundation, and a grant of no-charge study materials from Abbott Laboratories. The content is solely the responsibility of the authors and does not necessarily represent official views of NIH/NCI, Canary Foundation, Marsha Rivkin Center for Ovarian Cancer Research or Abbott Laboratories.

Grant Support

This work was supported by the Pacific Ovarian Cancer Research Consortium, Award Number P50 CA083636 from the NIH/National Cancer Institute (NCI). NIH/NCI P50 CA083636 and U01 CA152637.

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 May 18, 2012; revised July 24, 2012; accepted August 27, 2012; published OnlineFirst September 7, 2012.

References

- Hellström I, Raycraft J, Hayden-Ledbetter M, Ledbetter J, Schummer M, McIntosh M, et al. The HE4 (WFDC2) protein is a biomarker for ovarian carcinoma. *Cancer Res* 2003;63:3695–700.
- Havrilesky LJ, Whitehead CM, Rubatt JM, Cheek RL, Groelke J, He Q, et al. Evaluation of biomarker panels for early stage ovarian cancer detection and monitoring for disease recurrence. *Gynecol Oncol* 2008;110:374–82.
- Anastasi E, Marchei GG, Viggiani V, Gennarini G, Frati L, Reale MG. HE4: a new potential early biomarker for the recurrence of ovarian cancer. *Tumour Biol* 2010;31:113–9.
- Schummer M, Drescher C, Forrest R, Gough S, Thorpe J, Hellstrom I, et al. Evaluation of ovarian cancer remission markers HE4, MMP7 and Mesothelin by comparison to the established marker CA125. *Gynecol Oncol* 2012;125:65–9.
- Moore RG, McMeekin DS, Brown AK, DiSilvestro P, Miller MC, Allard WJ, et al. A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecol Oncol* 2009;112:40–6.
- Nolen B, Velikokhatnaya L, Marrangoni A, De Geest K, Lomakin A, Bast RC Jr, et al. Serum biomarker panels for the discrimination of benign from malignant cases in patients with an adnexal mass. *Gynecol Oncol* 2010;117:440–5.
- Ruggeri G, Bandiera E, Zanotti L, Belloli S, Ravaggi A, Romani C, et al. HE4 and epithelial ovarian cancer: comparison and clinical evaluation of two immunoassays and a combination algorithm. *Clin Chim Acta* 2011;412:1447–53.
- Moore RG, Jabre-Raughley M, Brown AK, Robison KM, Miller MC, Allard WJ, et al. Comparison of a novel multiple marker assay vs. the Risk of Malignancy Index for the prediction of epithelial ovarian cancer in patients with a pelvic mass. *Am J Obstet Gynecol* 2010;203:228. e1–6.
- Andersen M, Goff B, Lowe K, Scholler N, Bergan L, Drescher C, et al. Use of a Symptom Index, CA125 and HE4 to predict ovarian cancer. *Gynecol Oncol* 2010;116:378–83.
- Anderson GL, McIntosh MW, 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.
- 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* 2011;4:365–74.
- Urban N, Thorpe JD, Bergan L, Forrest R, Kampani A, Scholler N, et al. Potential role of HE4 in multimodal screening for epithelial ovarian cancer. *J Natl Cancer Inst* 2011;103:1630–4.
- 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.
- Buyss SS, Partridge E, Black A, Johnson CJ, 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.
- 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.
- 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.
- McIntosh M, Urban N. A parametric empirical Bayes method for cancer screening using longitudinal observations of a biomarker. *Biostatistics* 2003;4:27–40.
- McIntosh M, Urban N, Karlan B. Generating longitudinal screening algorithms using novel biomarkers for disease. *Cancer Epidemiol Biomarkers Prev* 2002;11:159–66.

19. Park Y, Kim Y, Lee EY, Lee JH, Kim HS. Reference ranges for HE4 and CA125 in a large Asian population by automated assays and diagnostic performances for ovarian cancer. *Int J Cancer* 2012;130:1136–44.
20. Park Y, Lee JH, Hong DJ, Lee EY, Kim HS. Diagnostic performances of HE4 and CA125 for the detection of ovarian cancer from patients with various gynecologic and non-gynecologic diseases. *Clin Biochem* 2011;44:884–8.
21. Moore RG, Miller MC, Eklund EE, Lu KH, Bast RC Jr, Lambert-Messerlian G. Serum levels of the ovarian cancer biomarker HE4 are decreased in pregnancy and increase with age. *Am J Obstet Gynecol* 2012;206:349. e1–7.
22. Jacob F, Meier M, Caduff R, Goldstein D, Pochechueva T, Hacker N, et al. No benefit from combining HE4 and CA125 as ovarian tumor markers in a clinical setting. *Gynecol Oncol* 2011;121:487–91.
23. Chang X, Ye X, Dong L, Cheng H, Cheng Y, Zhu L, et al. Human epididymis protein 4 (HE4) as a serum tumor biomarker in patients with ovarian carcinoma. *Int J Gynecol Cancer* 2011;21:852–8.
24. Sato AH, Anderson GL, Urban N, McIntosh MW. Comparing adaptive and non-adaptive algorithms for cancer early detection with novel biomarkers. *Cancer Biomark* 2006;2:151–62.
25. Urban N. Designing early detection programs for ovarian cancer. *Ann Oncol* 2011;22:viii16–8.
26. A Trial Using Novel Markers to Predict Malignancy in Elevated-Risk Women [updated 2010 Nov 15; cited 2010 Dec 6]. Available from: <http://clinicaltrials.gov/ct2/show/NCT01121640>.
27. Højsgaard S, Halekoh U, Yan J. The R package *geepack* for generalized estimating equations. *J Stat Softw* 2006;15:1–11.
28. Bliese P. Multilevel: multilevel functions. R package version 2.4; 2012 [cited 2012 Jul]. Available from: <http://CRAN.R-project.org/package=multilevel>.
29. Lowe KA, Shah C, Wallace E, Anderson G, Paley P, McIntosh M, et al. Effects of personal characteristics on serum CA125, mesothelin, and HE4 levels in healthy postmenopausal women at high-risk for ovarian cancer. *Cancer Epidemiol Biomarkers Prev* 2008;17:2480–7.
30. Shah CA, Lowe KA, Paley P, Wallace E, Anderson GL, McIntosh MW, et al. Influence of ovarian cancer risk status on the diagnostic performance of the serum biomarkers mesothelin, HE4, and CA125. *Cancer Epidemiol Biomarkers Prev* 2009;18:1365–72.
31. Baldwin LM, Trivers KF, Matthews B, Andrilla CH, Miller JW, Berry DL, et al. Vignette-based study of ovarian cancer screening: Do U.S. physicians report adhering to evidence-based recommendations? *Ann Intern Med* 2012;156:182–94.
32. 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.