Sources of Variability in Normal CA 125 Levels[1]

Carolyn Westhoff,[2] Bruce Levin, Gillian Ladd, and John O'Connor

Department of Obstetrics and Gynecology [C. W., G. L.], Divisions of Epidemiology [C. W.] and Biostatistics [B. L.] and G. Sergievsky Center [B. L.], School of Public Health, and Core Laboratory, Irving Center for Clinical Research [J. O.], Columbia University College of Physicians & Surgeons, New York, New York 10032, and the New York State Psychiatric Institute, New York, New York 10032 [B. L.]

Abstract

Two hundred eighty-eight serum specimens from 144 asymptomatic menopausal women were assayed two times each for CA 125. The specimens were randomly assigned to sixteen CA 125 kits. Our purpose was to quantitate the biological variability and the assay variability that occurs at low levels of this tumor marker. All of the subjects had CA 125 values less than 35 units/ml. The total error from all sources was about 13% of the observed values. These findings suggest that single, low CA 125 values are reliable indicators of a woman's true CA 125 value. Clinical trials of ovarian cancer screening that incorporate CA 125 measurements should evaluate the performance of this test at thresholds lower than the usual 35 units/ml.

Introduction

The CA 125 kit (Centocor, Inc.) is marketed for quantifying residual disease in women with invasive ovarian cancer, but the possible use of this test for the early identification of preclinical disease has excited much interest. Most reports to date indicate that only 50% of women who have Stage I ovarian cancer will have values of CA 125 that exceed 35 units/ml (1). Thus, the use of this test for screening or early diagnosis will have very limited sensitivity at the usual positive threshold of 35 units/ml.

In a recent study we evaluated CA 125 levels among menopausal women who were free of gynecological complaints and who had normal ovarian volumes measured by transvaginal sonography. In that study women had lower levels of CA 125 (mean, 5.6 units/ml) with less variability (SD, 3.5 units/ml) than has been reported elsewhere (2). Our results indicate that asymptomatic menopausal women usually have such low levels of CA 125 that, in the evaluation of this marker as a screening test, the positive threshold could be lowered in order to increase test sensitivity without an unacceptable loss of specificity.

Received 10/29/91.
[1] Supported in part by National Cancer Institute Grant 47827, the Milbank Memorial Fund, and Columbia University GCRC Grant RR00645.
[2] To whom requests for reprints should be addressed, at 630 W. 168 Street, New York, NY 10032.

The present study reanalyzed the specimens from a subset of these normal women in order to determine the sources and magnitude of error in measurements below 35 units/ml. The manufacturer recommends that values of CA 125 between 26 and 52 must be repeated to correctly assign a positive or negative result according to a 35 units/ml threshold. We are interested in the validity of much lower values that might be of interest for screening normal women. The sources of variability considered here are biological variability (*i.e.*, within-woman) and assay variability (*i.e.*, within- and between-kit).

Materials and Methods

The study population included 258 asymptomatic women with no spontaneous menstrual period in the preceding 4 months. Participants included women who were attending a general medical clinic or a screening mammography appointment and women who were members of the hospital or university staff. No women were recruited during visits for gynecological care. The population was about 62% white, 20% black, and 13% Hispanic. Their age range was 43 to 74 years. CA 125 levels were not associated with age or ethnic group (2). Only one of the original subjects had a CA 125 level > 35 units/ml; she was not included in the following sample. One hundred forty-four subjects were randomly selected for the following analyses from those original subjects who had sufficient quantities of stored samples. Their blood had been drawn on two occasions a median of 12 days apart. Specimens were refrigerated immediately; serum was separated by centrifugation within 24 h, and aliquots were stored frozen until analysis. CA 125 levels were determined by radioimmunoassay; duplicate measurements were always made on each serum sample within a single assay. Values for duplicates that differed from their mean cpm by more than 15% were rejected according to the manufacturer's instructions.

Analyses were organized as follows. Sixteen CA 125 kits were run with 36 unknowns in duplicate per kit; each kit included both first and second specimens from nine women, the first specimen only from nine other women, and the second specimen only from nine other women. All specimens for a single kit were prepared on the same day and were assayed in a single run. All assays were performed by a single technician. All kits included reagents from the same lot. In total, 288 specimens from 144 women were randomly assigned to be assayed in these kits. There were a total of 1152 observations (2 replicates × 2 specimens × 144 women × 2 assays each). In addition to these unknowns, each assay included duplicates of low and high control specimens provided with the kit and duplicates of nine standard specimens. Five of the standards were provided with the kit, but because all of the unknowns had values less than 35 units/ml (in the previous study), additional standards

were used in each kit (following Pittaway, Ref. 3) in order to improve the standard curve in the relevant range. The 1:1 dilution with albumin of the lowest standard provided by the manufacturer consistently yielded cpm that were greater than the undiluted standard; therefore, this diluted standard was not included in generating the standard curve. Because the 200- and 500-units/ml standards were outside of the range of our data, they were also excluded from the construction of the standard curve. The standard curve was constructed to connect the value for the lowest standard to the origin with a straight line. This assumption of linearity at the low end of the range, although not specifically justified by previous analyses, has been made by the manufacturer and by all previous investigators. The standard curve between the lowest standard and the 80-units/ml standard was constructed using least-squares linear regression on the standards in this range. These regression lines had an excellent fit to the data for all 16 kits, with each r^2 greater than 0.97.

The statistical analysis included only those 104 subjects who had complete and satisfactory data for every planned observation. Subjects were excluded only if at least one result from the set was missing due to insufficient specimen or laboratory error, or if the values for the replicates differed by more than 15% from their mean. Because the CA 125 values, even in this truncated population, had a right-skewed distribution, all statistical analyses were performed using log-transformed values. The transformed values are approximately normally distributed.

The variability in these measurements of CA 125 was quantified using the following model:

$$\log x_{ijkl} = \alpha_i + \beta_{ij} + \gamma_k + \epsilon_{ijkl}$$

where x_{ijkl} is the observed value of CA 125 for woman i with specimen j, kit k, and replicate l; α_i is the true average log CA 125 value for woman i; β_{ij} is the true deviation from α_i for specimen j; γ_k is the systematic effect of kit k for all specimens; ϵ_{ijkl} is the replicate error; and

$$\delta_i = \beta_{i2} - \beta_{i1}$$

which is the true change in CA 125 for woman i between specimens. The model specification is completed with the following assumptions of constant variance: σ_δ^2 (variance of δ_i) is the within-woman variance of change between specimens; σ_k^2 (variance of γ_k) is the between-kit variance; and σ_ϵ^2 (variance of ϵ_{ijkl}) is the replication error within specimen and kit. These parameters and their SEs were estimated by normal theory maximum likelihood. The assumption of constant within woman variance was evaluated using Bartlett's F test for homogeneous variance.

Results

The values for the low and high control specimens fell within 1 SE of the predicted values for all 16 kits that were assayed. This finding indicates the technical success of our laboratory procedures and supports the use of our results in the analyses that follow. Of the 576 paired replicates that were run in the 16 assays, 25 (4.3%) were rejected, according to manufacturer's directions, due to excess variation between the replicates. Results from 10 subjects were excluded from the analyses due to the loss of the specimens' labels during one assay procedure. An additional five women were excluded due to insufficient quantity of specimen to complete the planned assays. All CA 125 values fell between 2.5 and 15 units/ml. Within-woman differences ranged from −1.4 to 1.7 units/ml. The SD of the within-subject differences in subjects with CA 125 values above the median was 0.1953 (in log-transformed units); the SD in subjects with CA 125 values below the median was 0.1687 ($F_{52,52} = 1.34$; $P > 0.10$), which indicates that the magnitude of the differences was unrelated to the CA 125 values and that the assumption of constant variance is justified.

Table 1 Components of variability in CA 125

Source	Value
Within-woman change (σ_δ)	0.030
Between-assay (σ_k)	0.085
Within-assay (σ_ϵ)	0.095
Total ($[\sigma_\delta^2 + \sigma_k^2 + \sigma_\epsilon^2]^{1/2}$)	0.131

Table 2 True CA 125s associated with various observed values

Observed value (x)	95% confidence interval of true value (exp[x ± 1.96 $(\sigma_\delta^2 + \sigma_k^2 + \sigma_\epsilon^2)^{1/2}$])
12	9.3–15.5
14	10.8–18.1
16	12.4–20.7
18	13.9–23.3
20	15.5–25.9
22	17.0–28.4

Table 1 shows the values of the error components expressed as units of the log-transformed CA 125 values and based on analyses of the 104 women with complete results. Because these calculations were performed using the natural logarithm of the CA 125 values, and because the variance estimates are all small (that is, close to zero), the resulting estimates of σ_δ, σ_k, and σ_ϵ can be interpreted as percentages of the original untransformed CA 125 value. Thus, the total error from all sources of 0.131 can be interpreted as a SE of about ±13% of the observed values. The within-assay (or replicate) error of 0.095 reflects the replicate error that remained after rejecting all of those specimens where the replicates had differences in excess of the manufacturer's standards. The within-assay coefficient of variation was 2.2%, and the between-assay coefficient of variation was 1.9%. The major component of the error was within-assay variability; the between-kit variability and the within-woman variability made smaller contributions to the total error.

It is possible to use the estimate of the total error shown in Table 1 to calculate from any observed result a 95% confidence interval of a woman's true CA 125 value. These calculations are shown in Table 2 for a range of low observed values and account for the log transformation. Because the distribution of CA 125 values is not normally distributed, estimations of the confidence interval of the true value will be inappropriately wide for low values unless calculations are performed using log-transformed values. Our analyses reveal that a single, low

observed value for CA 125 is a more precise estimate of a woman's true CA 125 value than has been previously appreciated.

Discussion

The possibility of screening women for ovarian cancer is attracting both scientific and popular interest. There are several large trials of screening now beginning or under way in the United States and Europe. CA 125, with a positive threshold of 30 or 35 units/ml, is one of the tests being evaluated in these trials.

Because CA 125 is relatively insensitive for identifying Stage I ovarian cancer at a threshold of 35 units/ml, some investigators are pessimistic about its potential utility as a screening test (4). In a nested case-control analysis of stored sera from a large cohort of Norwegian women (5), those 105 women who developed ovarian cancer up to 6 years later had a median CA 125 level of 18 units/ml. This finding further supports the notion that a threshold of 35 units/ml is insufficiently sensitive for screening purposes. Among the 323 matched controls in that study, the median CA 125 level was only 11 units/ml. This large and highly significant difference between cases and controls suggests, however, that CA 125 may prove to be useful for screening if a threshold much lower than 35 units/ml is used. In that study the sensitivity increased from 17% to 42% when the threshold was changed from 35 to 20 units/ml.

The reluctance to use a lower threshold of positivity stems from two test concerns. First is the concern that lower CA 125 values have unacceptable precision and validity due to biological variability within women and due to the intrinsic characteristics of the immunoradiometric assay. In the present study we have shown among women with very low CA 125 values that a single observed value is a precise estimate of the woman's true value and in these menopausal women that the biological contribution to the observed variability is small. This means that the use of a lower threshold is not contraindicated due to these assay characteristics. The present analysis, however, somewhat underestimates the assay contribution to the total variability. First we excluded those replicates (4% of the total) with excess variability according to the manufacturer's standards; thus, there is an artificial upper limit on the variability of the specimens included in our analyses. Second, all assays were performed with reagents from the same lot; in general usage there will be additional variability due to differing lots of reagents. However, our conclusion that biological (*i.e.*, within-woman) variability is small is not biased by these factors.

A second reason for the reluctance to use a lower threshold of positivity is that test specificity will be lower. For this relatively rare disease, lower test specificity means a large increase in the proportion and number of women who would be incorrectly labeled as positive. Because the CA 125 distributions of normal and affected women overlap, this problem is only slightly mitigated by our finding of excellent precision in this assay. Studies in mixed-age populations suggested that the test specificity may be unacceptable even at a threshold of 35 units/ml; however, CA 125 levels are much lower in menopausal women (who would be the appropriate target population for screening). Only 1% of menopausal women have a positive test at a threshold of 35 units/ml (2, 6); however, between 2% (2) and 8% (6) may be positive at a threshold of 20 units/ml. A constructive approach to the likely problem of low specificity would be to use a second-level screening test for those women who have a positive CA 125 value. Such a modality may already be available in the use of pelvic sonography or in the use of other serum tumor markers.

The value of screening for ovarian cancer can be assessed only through the evaluation of mortality rates in women who are enrolled in sufficiently large trials. Because such trials are extremely difficult and expensive to carry out, it is necessary to determine that the proposed screening modalities perform well enough to merit such an evaluation. The findings in the present study indicate that the measurement of CA 125 with a low threshold for positivity deserves to be evaluated in any trials of multimodal screening for ovarian cancer.

References

1. Jacobs, I., and Bast, R. C. The CA 125 tumour-associated antigen: a review of the literature. Hum. Reprod. (Oxf.), *4:* 1-12, 1989.

2. Westhoff, C., Gollub, E., Patel, J., Rivera, H., and Bast, R. CA 125 levels in menopausal women. Obstet. Gynecol., *76:* 428-431, 1990.

3. Pittaway, D. Assay characteristics of the immunoradiometric method for CA 125. Fertil. Steril., *51:* 348-350, 1989.

4. Mann, W. J., Patsner, B., Cohen, H., and Loesch, M. Preoperative serum CA 125 levels in patients with surgical Stage I invasive ovarian adenocarcinoma. J. Natl. Cancer Inst., *80:* 208-209, 1988.

5. Zurawski, V. R., Orjaseter, H., Andersen, A., and Jellum, E. Elevated serum CA 125 levels prior to diagnosis of ovarian neoplasia: relevance for early detection of ovarian cancer. Int. J. Cancer, *42:* 677-680, 1988.

6. Zurawski, V., Sjovall, K., Schoenfeld, D., *et al.* Prospective evaluation of serum CA 125 levels in a normal population, Phase I: the specificities of single and serial determination in testing for ovarian cancer. Gynecol. Oncol., *36:* 299-305, 1990.