Division of Immunology and Hematology Devices
Office of In Vitro Diagnostic Device Evaluation and Safety
Center for Devices and Radiological Health

Document Mail Center
9200 Corporate Boulevard
Rockville, MD 20850

FDA Briefing Document

Fujirebio Diagnostics, Incorporated

HE4 EIA Assay

Immunology Devices Panel Meeting
December 3, 2008

FDA Executive Summary

Ovarian cancer is a prime target for improved diagnostics because of its subtle onset, high morbidity and mortality, and narrow opportunities for effective intervention. When a patient presents with low stage ovarian cancer, the expertise and skill of the surgical team affect her prospects for long term survival. The same attributes affect the palliation of disease that is already advanced at presentation.

Through more accurate patient referrals, serum based biomarkers that help to distinguish benign from malignant ovarian disease could help increase the proportion of ovarian cancer cases initially treated in oncology centers and allow more patients with benign disease to be treated efficiently in the community setting. The clinical validation of such a proposed use requires study of the biomarkers in the appropriate intended use population (i.e., community based or oncology referred) to demonstrate acceptable clinical performance characteristics and clinical implications. Such study might assess the biomarker either as a stand-alone test or in the context of other clinical and laboratory findings.

An enzyme immunoassay (Fujirebio Diagnostics Incorporated, sponsor) of biomarker HE4 is proposed for use in conjunction with an assay for another serum biomarker (CA125) and a Risk of Malignancy Algorithm (ROMA), to report a predictive probability (PP) of epithelial ovarian cancer. PP is computed differently for premenopausal and postmenopausal patients. “Subjects categorized as low risk for epithelial ovarian cancer using the ROMA value may have surgical intervention performed by a non-oncology specialist. The results must be interpreted in conjunction with other clinical findings in accordance with standard clinical management guidelines. The assay is not indicated as an aid in a decision to proceed to surgery.”

The test was studied in 530 cases of women with pelvic mass scheduled for surgery at fourteen U.S. centers with surgical oncology expertise. Epithelial ovarian cancer (EpOvCa) was found in 129 patients. Ovarian tumors with low malignant potential (LMP) were found in 22 patients. Other kinds of cancer (OtherCa) were present in 28 patients, and 351 patients had benign surgical pathology findings.

Some patients were originally excluded from final data analyses because their menopausal status, according to originally specified criteria, was indeterminate. For these patients, the sponsor devised two additional rules to assign menopausal status according to a) the patient’s age, and b) prior surgical history or absence of a known date for last menstrual period, and c) ovarian function testing based on measurement of follicle-stimulating hormone in serum.

Using ROMA equations developed from two pilot studies, the sponsor set assay cut-offs for the pivotal study, such that 75% of the studied women who had a benign pelvic mass were classified correctly. The cut-offs were 13.1% PP for premenopausal patients and 27.7% PP for postmenopausal patients. For a success criterion, the sponsor set a requirement that (using the cut-offs as defined) the lower bound of the 95% Confidence Interval for Sensitivity must be >80% for premenopausal and postmenopausal women.

Sponsor’s presentation of clinical performance characteristics concentrates on the 502 patients who had EpOvCa, LMP or benign findings. The observed sensitivity for EpOvCa or LMP combined was 88.7% (95% CI: 82.6%, 93.3%) for all 502 patients, 76.5% (95% CI: 58.8%, 89.3%) for the 234 premenopausal patients, and 92.3% (95% CI: 85.9%, 96.4%) for the 268 postmenopausal patients. The corresponding negative predictive value results were 93.9% (95% CI: 90.4%, 96.4%) for all patients, 94.9% (95% CI: 90.2%, 97.8%) for premenopausal, and 92.6% (95% CI: 86.5%, 96.6%) for postmenopausal.

When all cancers (EpOvCa, LMP and OtherCa) were included in the analysis, the observed sensitivity for these findings combined was 86.0% (95% CI: 80.1%, 90.8%) for all 530 patients, 67.4% (95% CI: 52.0%, 80.5%) for premenopausal patients, and 92.5% (95% CI: 86.6%, 96.3%) for the 284 postmenopausal

patients. The corresponding negative predictive value results were 91.3% (95% CI: 87.4%, 94.3%) for all patients, 90.9% (95% CI: 85.4%, 94.8%) for premenopausal, and 91.9% (95% CI: 85.6%, 96.0%) for postmenopausal.

The pivotal study focused on patients who were evaluated and treated at oncology centers. The study design and statistical analysis did not provide information about performance of the test in a community setting. The design and analysis did not provide information about interaction of PP results with other clinical or pathological data that are available in either a community setting or an oncology center setting. Various analyses concerning subsets of the pivotal study patients were apparently not pre-specified, and raise FDA concerns about their statistical validity.

FDA requests advice from the Immunology Devices Panel, on the basis of the data available for review, concerning safety and effectiveness of the test for the intended use proposed by the sponsor. The questions for the panel focus most on the defensible scope of the intended use population, on acceptability of the test's performance in selecting patients for surgical intervention by a non-oncology specialist, on how the test's results might be knowledgably combined with other clinical and laboratory data, and on differences in the impact of the test according to stage or type of malignant ovarian disease. We complete the request for advice with questions concerning mitigation of misclassifications by the test, and about adequate assessment of menopausal status.

Chapter 1 Introduction

Medical Issues and Need for Interested Party Comment

Women with image-documented adnexal masses and clinical symptoms often undergo intervention by laparotomy or laparoscopy. Many patients are referred to an oncology or specialty gynecologic care center for surgical intervention because of the possibility of gynecologic cancer. Given that surgery is required, a major clinical question (determined by the likelihood of benign vs malignant disease) is whose surgical services will give the best clinical outcome.

It is well documented in the medical literature that optimal survival of ovarian cancer patients depends upon appropriate initial surgery: tumor de-bulking and accurate clinical staging by lymphadenectomy. Less optimal initial surgery is also documented in the literature to give less optimal outcome. However, it is often difficult to discern pre-operatively whether the disease requiring surgery is benign or malignant. Partly because of this, surgery for benign conditions often is nevertheless performed by an oncology specialist. Likewise, it is not uncommon that initial surgery for what turns out to be a malignant condition is undertaken by a non-oncology specialist.

There are two apparent modes for improving selection of the kind of surgical intervention for patients with adnexal mass who do need surgery. One is to increase the proportion of patients with malignant disease who are sent forward for assessment and treatment by an oncology specialist. This requires study of the test's performance in the primary care setting. The other is to increase the proportion of patients with benign disease who receive treatment from a non-oncology specialist. One way to assess test performance in this mode is to study the test in the referred care (oncology) setting, to identify patients for whom surgery by a non-oncology specialist is suitable.

Fujirebio Diagnostics Incorporated has submitted for FDA review a test that has been studied according to the latter mode described above. Since there is no predicate (i.e., previously cleared device) for the test's intended use, a determination of the device's safety and effectiveness is in order. To help FDA with this determination, advice from the Immunology Devices Panel is needed.

A key question hinges on the fact that designating some referred patients who do not need oncology surgery carries a risk that some patients with malignant disease will not be identified; i.e., the test or procedure will falsely classify them as “negative.” As already discussed, non-oncology surgery for such patients with malignant disease might be suboptimal. Measures to minimize such misidentifications and to mitigate their effect must be considered. Assuming that such measures are well understood and put in place, the question bearing on safety and effectiveness is: what is the clinically tolerable percentage, or percentage range, of patients or malignant disease who/that are/is misidentified as benign? This question can be considered both in terms of the proportion of patients identified as negative who in fact do have malignant disease (1 – Negative Predictive Value) and in terms of the proportion of malignant disease that is missed in the population (1 - Sensitivity).

Given that an acceptable false negative rate for the test can be identified and achieved, an additional key question is whether findings of malignant disease among patients identified as “positive” by the test or procedure are sufficiently frequent to justify use of the test (i.e., is the test’s positive predictive sufficiently high?). The answer depends partly on the prevalence of disease in the tested population, which also helps drive the positive predictive value of the test.

It appears that opinions about the answers to these questions vary. Since the FDA is not in a position to establish or recommend guidelines for medical practice and since there is no scientific literature appropriate to assist us in deciding these questions, an advisory panel meeting is convened to discuss and make recommendations regarding this submission.

Chapter 2 Regulatory Background and Device Description

Proposed Intended Use and Regulatory Status of Device

For the proposed device under review and panel comment, the sponsor has proposed the following Intended Use/Indications for Use:

The HE4 EIA is an enzyme immunometric assay for the quantitative determination of HE4 in human serum. The assay used in conjunction with the ARCHITECT CA 125 II assay creates a predictive probability of epithelial ovarian cancer using a mathematical function referred to as the Risk of Ovarian Malignancy Algorithm (ROMA), for use in premenopausal and post-menopausal women presenting with an adnexal mass who have already been referred to an oncologic specialist and are scheduled for surgery. Subjects categorized as low risk for epithelial ovarian cancer using the ROMA value may have surgical intervention performed by a non-oncology specialist. The results must be interpreted in conjunction with other clinical findings in accordance with standard clinical management guidelines. The assay is not indicated as an aid in a decision to proceed to surgery.

It is particularly important to note that the device creates, from a prespecified mathematical model, a predictive probability using two separately measured biomarkers in pre- and post-menopausal women who have an adnexal mass and have already been referred to an oncologic specialist and are scheduled for surgery. The predictive probability is not used as an aid in a decision to proceed to surgery. In the intended use setting, predictive probability is not utilized to make a decision on referring patients to an oncologic specialist.

The proposed Intended Use has not been the subject of a premarket notification clearance or premarket approval in the past. The device is potentially a candidate for initial classification under Section 513(f)(2) of the Federal Food, Drug, and Cosmetic Act. The risks and benefits of this device for its proposed Intended Use are currently being evaluated and a decision will be made at a later date, though the panel is free to offer its judgment on these aspects or other safety and effectiveness aspects of this device.

Regulatory status of component assays in the proposed device

The component assays used in this device currently are cleared for the following Intended Uses:

The HE4 EIA is an enzyme immunometric assay for the quantitative determination of HE4 in human serum. The assay is to be used as an aid in monitoring recurrence or progressive disease in patients with epithelial ovarian cancer. Serial testing for patient HE4 assay values should be used in conjunction with other clinical methods used for monitoring ovarian cancer.

See the FDA decision summary for K072939 at http://www.fda.gov/cdrh/reviews/K072939.pdf for detailed information.

The ARCHITECT CA 125 II assay is a Chemiluminescent Microparticle Immunoassay (CMIA) for the quantitative determination of OC 125 defined antigen in human serum and plasma on the ARCHITECT i System. The ARCHITECT CA 125 II assay is to be used as an aid in monitoring response to therapy for patients with epithelial ovarian cancer. Serial testing for patient CA 125 II assay values should be used in conjunction with other clinical methods used for monitoring ovarian cancer.

See the FDA decision summary for K042731 at http://www.fda.gov/cdrh/reviews/K042731.pdf for detailed information.

Chapter 3 Analytical Characteristics of CA125 and HE4 Assays

Both the HE4 and CA125 assay are dual antibody sandwich immunoassays. In the HE4 assay a biotinylated anti-HE4 antibody captures the antigen and is absorbed to an avidin-coated standard micro-well solid phase. Enzyme-labeled anti-HE4 antibody detects solid-phased bound HE4 and produces visible color signal on addition of enzyme substrate. Color is measured in a standard spectrophotometer and optical density measurements are related to known concentrations of assay calibrator. Results are expressed in picomoles/liter (pM). The HE4 assay is a manual assay and is not currently for use with standard laboratory clinical analyzers. In the CA 125 assay an anti-CA125 antibody captures the antigen directly to a microparticle solid phase. Luminescent-labeled anti-CA125 antibody detects bound CA 125 and produces a signal on addition of a chemiluminescent solution. Luminsecent signal is measured by the Abbott Architect analyzer. Signal is related to known concentrations of assay calibrator. Results are expressed as Units/ml (U/ml).

Both the HE4 and CA125 assays utilize arbitrary calibration since there are no reference standard calibrators for either assay. The measurement range for the HE4 assay is 15 to 900 pM. The measurement range for the CA125 assay is 0 to 1000 U/ml. The Limit of blank (lowest concentration distinguishable from zero) for the HE4 assay is 1.1 – 2.2 pM. The limit of quantitation (lowest concentration at which the sample coefficient of variation was 20%) for the HE4 assay is < 5 pM. The limit of blank for the CA 125 assay is $\leq$ 1 U/ml.

Using percentile ranking analysis, 95% of approximately 200 apparently healthy pre-menopausal women, post-menopausal women and pregnant females had HE4 values below 150 pM. Using approximately the same number of pre-menopausal and post-menopausal subjects, 94% of apparently healthy subjects had CA 125 values below 35 U/ml.

The following tables, taken from the package inserts of the respective assays, indicate the distribution of values in apparently healthy subjects, benign disease subjects, and cancer subjects.

Distribution of HE4 Assay Values (number of subjects per category)

	Number of subjects	0 - 150 pM	150.1 - 300 pM	300.1 - 500 pM	> 500pM
APPARENTLY HEALTHY					
Females (Premenopausal)	76	72	3	0	1
Females (Postmenopausal)	103	97	5	0	1
BENIGN CONDITIONS					
Pregnancy	22	21	1	0	0
Benign Gynecological Disease	347	324	18	1	4
Other Benign Disease	108	82	8	7	11
Hypertension/Cong. Heart Failure	96	75	16	2	3
CANCER					
Ovarian Cancer	127	27	18	21	61
Breast Cancer	46	40	4	2	0
Lung Cancer	50	29	15	6	0
Endometrial Cancer	116	86	15	4	11
Gastrointestinal Cancer	56	47	8	0	1

	Distribution of ARCHITECT CA 125 II Assay Values				
	Number of Subjects	Percent (%) of subjects per category			
		0-35 U/mL	35.1-65 U/mL	65.1-100 U/mL	>100 U/mL
APPARENTLY HEALTHY					
Females (Premenopausal)	99	89.9	6.1	4.0	0.0
Females (Postmenopausal)	97	99.0	1.0	0.0	0.0
MALIGNANT CONDITIONS					
Ovarian Cancer	166	49.9	14.3	4.8	32.8
Breast Cancer	50	80.0	20.0	0.0	0.0
Colorectal Cancer	50	84.0	4.0	10.0	2.0
Endometrial Cancer	25	96.0	4.0	0.0	0.0
Lung Cancer	50	60.0	18.0	10.0	12.0
NONMALIGNANT CONDITIONS					
Ovarian Disease	100	90.0	9.0	1.0	0.0
Urogenital Disease	49	83.7	14.3	2.0	0.0
Hypertension/CHD	100	88.0	11.0	0.0	1.0
Benign Endometrial	25	84.0	8.0	4.0	4.0

Chapter 4 Risk of Malignancy Algorithm Development and Clinical Study Evaluation

The sponsor initially performed a pilot study of 201 women presenting with a pelvic mass who were to have surgery at a single university hospital. The first pilot study was to evaluate multiple serum biomarkers to predict the probability of epithelial ovarian cancer. Blood and urine samples were collected from women under informed consent and a local IRB approved protocol. Tissue removed at surgery was utilized to make a histopathological diagnosis of cancer or non-cancer. Logistic regression and receiver-operator characteristics curve analyses were utilized to evaluate multiple biomarker values for a statistically significant association with ovarian cancer. To improve the accuracy of marker analysis, 236 similar women from a different university hospital were included at a later analysis. Women from the second site were over-sampled for invasive epithelial ovarian cancer. Evaluation of multiple markers indicated that CA125 and HE4 provided the highest sensitivity at set specificities of 80%, 90%, 95%, and 98%. At initial development a single regression equation stratified women from the first group of 201 subjects into 3 groups regardless of menopausal status: low risk, moderate risk, and high risk. In this model, women with either invasive epithelial ovarian cancer only or benign disease were categorized by CA125 and HE4 values. The model was evaluated in the final validation study in what is described as the "Protocol Analysis", using the single equation that defined it, though this model was later superceded by a more complex model as described below.

Subsequent to their realization that menopausal status was a statistically significant factor in the probability of cancer, the sponsor built a new model based on the data from both pilot studies and using two equations: one for pre-menopausal women and one for post-menopausal women, using CA125 and HE4 values in the logistic regression classifier. These model equations were evaluated in the final validation study, via the "Additional Analysis."

The Additional Analysis of the two equations chose predictive probability cutoff values after the Protocol Analysis in the validation study indicated that 75% specificity would yield a sensitivity above 80% at its lower 95% confidence limit. In essence, the Additional Analysis for the validation study utilized cutoffs informed by the Protocol Analysis of the final validation study. However, the parameters for the two equations used for the Additional Analysis in the validation study were estimated based on subjects from the pilot studies.

The validation study was a prospective, multi-center, study enrolling female subjects ≥18 years of age referred to a gynecologist or gynecologic oncologist with a pelvic mass and who were scheduled to undergo surgery. A pelvic mass was demonstrated by ultrasound, CT scan or MRI. Subjects were excluded if:

- A subject received treatment for any malignancy (with the exception of non-melanoma skin cancer) within the last five years;
- Subjects were receiving cytotoxic chemotherapies, such as cyclophosphamide or methotrexate;
- Subjects with previous bilateral oophorectomy;
- Any subject known to be pregnant.

The study was conducted at 14 clinical geographically-disperse clinical sites across the US. A total of 566 patients were enrolled of which 43 patients (7.6%) were non-evaluable for various reasons. Of the 43 non-evaluable subjects, the most common reason for exclusion was inclusion/exclusion criteria failures (21 of 43, 48%). The two next most common reasons for exclusion were the subject did not undergo surgery (15 of 43, 34%) and discordant pathology results (6 of 43, 13%). There were 28 subjects in addition to the 43 subjects that were not evaluable in the Protocol Analysis and the Additional Analysis because they had

malignancies other than invasive epithelial ovarian cancer or low malignant potential (borderline) tumors. Seven of the non-evaluable patients were later returned to the study for the Additional Analysis.

Serum samples were drawn from all subjects and tested in the HE4 EIA at three sites. Sera were also tested in the ARCHITECT CA 125 II assay at one site. The Study Protocol was reviewed and approved by local or central Institutional Review Boards. Written informed consent was obtained from each subject prior to the collection of blood for this study. All patients who were evaluated and had results analyzed in this study underwent laparoscopic surgery or laparotomy, and all tissues were examined by local pathologists and centrally reviewed by 2 or 3 other pathologists. Final review of histopathology results from site and central review for final consensus determination was performed by two gynecologic oncologists blinded to laboratory values. All patient management was done by clinicians blinded to laboratory results.

Menopausal status was determined in the study based on the date of the last menstrual period. Post-menopausal status was defined as one year past the last menstrual period or hysterectomy with bilateral oophorectomy. In the case that the last menstrual period was unknown, pre-menopausal status was defined as 48 yrs and younger with no prior history of a hysterectomy with bilateral oophorectomy and post-menopausal status was defined as 56 years and older. Women between the ages of 49 and 55 years of age who did not have a date for their last menstrual period were considered unknown menopausal status and were not included in the initial submission of the Additional Analysis. In a later re-determination of the menopausal status of 54 women, additional rules were developed to assign menopausal status according to a) the patient's age, and b) prior surgical history or absence of a known date for last menstrual period, and c) ovarian function testing based on measurement of follicle-stimulating hormone in serum (Abbott Architect FSH assay using a cutoff of 22 mIU/mL). The redetermination reclassified, as premenopausal, 39 women who were originally considered postmenopausal. It also determined the menopausal status of 7 women who were previously indeterminate (enabling their inclusion in the Additional Analysis).

All of the major racial groups were represented with approximately 15% representing minority groups. Of the 516 evaluable subjects with known menopausal status, 315 (61%) were post-menopausal and 201 (39%) were pre-menopausal. A total of 176 (34%) women were found to have cancer; 127 (72%) of the cancers were invasive epithelial ovarian cancers and 21 (12%) were low malignant potential (borderline) tumors. Of 123 staged epithelial ovarian cancers, 34 (28%) were Stage I/II, and 89 (72%) were Stage III/IV. Of 127 subjects with epithelial ovarian cancer, 9% were grade I, 24% were grade 2, 65% were grade 3, and less than 2% had unknown grade.

In the protocol analysis, CA 125 and HE4 serum levels from the model equation developed in the pilot study calculated predicted probabilities that separated subjects into low, moderate, and high risk groups for ovarian cancer:

- Low Risk: < 7.7% predictive probability of ovarian cancer
- Moderate Risk: 7.8% to 15.0% predictive probability
- High Risk: > 15.1% predictive probability

In the protocol analysis, the primary endpoint (cancer/non-cancer by histopathology) was compared with risk category using benign disease subjects and cancer patients with invasive epithelial ovarian cancer only (N=474). In additional analysis, the primary endpoint was compared with risk category across menopausal status strata using benign disease patients and cancer subjects with invasive epithelial cancer and epithelial ovarian cancer subjects with low malignant potential tumors (N=495). The sponsor chosen cut-points for the "Additional Analyses" were initially 12.9% Predictive Probability for pre-menopausal patients and 24.7% Predictive Probability for post-menopausal patients. These cutoffs shifted to 13.1% Predictive Probability for the pre-menopausal patients and 27.7% Predictive Probability for the post-menopausal patients after the menopausal status redeterminations. The final cutoffs were used to test the separation of pre-menopausal and post-menopausal patients into low risk and high risk subsets.

The sponsor initially combined data from some sites to form 10 sites from the starting 14 sites on which to perform analysis for pooling all subjects for analysis. Subjects with less than 10 subjects per site were pooled when in the same demographic area. Demographic characteristics for comparison included age of the patient at the time of consent, menopausal status, or specimen type (i.e., benign disease or cancer). For those patients with epithelial ovarian cancer, the additional characteristics of stage of disease, grade of disease and histology were also evaluated. No significant difference in these demographic characteristics was apparent. The sponsor states that based upon the analysis, pooling the data across all site was justified.

For the Protocol Analysis, the sponsor evaluated their primary protocol endpoint using only patients with benign diseases or invasive epithelial ovarian cancer. Patients were categorized as low risk, moderate risk, or high risk for ovarian cancer based upon serum CA125 and HE4 concentrations using cutoff values developed from the pilot study. The Chi-square p-values were < 0.05 signifying that the proportion of cancer and benign disease patients in each of the risk subgroups is statistically significantly different.

Table 20 Primary Endpoint analysis using the single equation algorithm from the Protocol Analysis. Cancer patients - Invasive epithelial ovarian cancer only.

	Low Risk ≤7.7%	Moderate Risk 7.8 - 15.0%	High Risk ≥15.1%	Total
Benign	53	208	86	347
Invasive Epithelial Ovarian Ca (I/II)	0	7	27	34
Invasive Epithelial Ovarian Ca (III/IV)	1	1	87	89
Invasive Epithelial Ovarian Ca (Unstaged)	0	2	2	4
	54	218	202	474

	Low Risk ≤7.7%	Mod Risk 7.8 - 15.0%	High Risk ≥15.1%	
Benign	53 (15%)	208 (60%)	86 (25%)	347 (73%)
Epithelial Ovarian Cancer	1 (1%)	10 (8%)	116 (91%)	127 (27%)
Total	54 (11%)	218 (46%)	202 (43%)	474 (100%)

Pearson's chi^2 = 168.5678	p = 0.0000
Likelihood-ratio chi^2 = 184.2785	p = 0.0000
Fisher's Exact Test ---	p = 0.000

In the Protocol Analysis, the sponsor evaluated the secondary protocol endpoint in which patients with low malignant potential (borderline) tumors were included. The secondary endpoint was intended to be a multivariate predictive algorithm combining multiple serum biomarkers (CA 125 and HE4), radiologic imaging results, and patient risk factors (such as age, menopausal status and ethnicity) for estimation of the risk of ovarian cancer, including low malignant potential (borderline) tumors, at the time of surgery in patients presenting with a pelvic mass.

Results from the Protocol Analysis for a completed secondary protocol endpoint evaluation were not submitted. Inclusion of imaging results is an ongoing study and was not described. Ethnicity was not

included in the analysis due to the small numbers in the study. Age was assessed and provided no definitive conclusion so it was not included in the analysis. However, the sponsor notes evaluation of menopausal status and inclusion of low malignant potential (borderline) tumors. The Chi-square p-values were < 0.05 signifying that the proportion of patients with cancers and benign diseases were significantly different among the three risk groups for the primary and secondary protocol endpoint.

Chapter 5 Performance Evaluation for the "Additional Analysis"

Sponsor's presentation of the results from their Additional Analysis are detailed in the document "Clinical Study Report, Protocol FDI-03A2" included in this panel packet. FDA's independent analysis, from sponsor-provided line listings of the study data, is in close (though not exact) agreement with the sponsor's analysis for the major clinical performance characteristics, i.e., the two-by-two contingency tables assessing discrimination of three patient groups (all patients, premenopausal patients and postmenopausal patients) for Epithelial Ovarian Cancer or LMP tumors versus benign disease. There is similar close agreement for the discrimination of these patient groups with regard to Epithelial Ovarian Cancer or LMP tumors or Other Cancers versus benign disease. These results are presented below with brief discussion of their implications. Other analyses by sponsor are the subject of more in depth critique concerning their significance and validity.

Basic Statistical Points

The sponsor developed an algorithm that combines CA 125 and HE4 concentrations in a logistic model, and provides a probability of finding cancer in a given patient. This algorithm was developed using a Training Set obtained by pooling data across two separate Pilot Studies at Women's and Infants Hospital of Rhode Island (WIHRI) in Providence, RI, and Massachusetts General Hospital (MGH) in Boston, MA.

Premenopausal women

PI (Predictive Index) = -12.0 + 2.38*LN(HE4) + 0.0626*LN(CA125)

Postmenopausal women

PI (Predictive Index) = -8.09 + 1.04*LN(HE4) + 0.732*LN(CA125)

ROMA = exp(PI) / [1 + exp(PI)]

Our comments and concerns are related to the clinical study, an independent data set for the evaluation of the ROMA test (Validation Set).

1) Target Population

The clinical study (validation set) was conducted at 14 clinical sites. Patients evaluated in the clinical study were patients referred to gynecologic oncology centers. Patients who were not referred to the oncology centers by the referring physicians were not included in the clinical study and therefore, performance of the ROMA test was not evaluated for this group of patients. Patients who were referred to the oncology centers had an increased risk for ovarian cancer (this risk for ovarian cancer is evaluated by the referring physicians) and therefore, the clinical study subjects are not a representative samples from the population of community setting, women who have been diagnosed with an adnexal mass and who are candidates for surgical intervention.

According to the sponsor's information, all patients in the study were subjects referred by a physician or self referred to oncology centers. One can speculate that the clinical study includes two groups of patients 1)

subjects who were assessed by a referring physician as high risk subjects; and 2) subjects who were assessed by a referring physician as low risk subjects but were operated in the oncology centers by some different reasons.

Table 1

Pre-Surgical Assessment by Physician		
Malignant (High Risk)	**Non-Malignant (Low Risk)**	
All subjects were referred to oncology centers	Subjects were operated in oncology centers	Subjects were operated in places other than oncology centers
Subjects of the Clinical Study		

So, the subjects of the clinical study with low risk assessed by a referring physician and operated in community setting were not present in the sponsor's study, and therefore are not part of the population cited in the intended use.

The clinical study subjects can be considered as a representative sample for the target population as "women who were referred to a gynecologic oncologist and are scheduled for surgery" and the ROMA assay will be used as an aid to stratify women as low risk subjects in order to refer these women back to their gynecologist.

2) No Information About Pre-Surgical Clinicopathologic Assessment as Risk Estimator Alone or in Conjunction with the New Test

General gynecologists involved in the management of women with pelvic mass use available information (e.g., symptoms consistent with ovarian cancer, family history, pelvic exam, imaging studies concerning for ovarian cancer, an elevated levels of CA125) for assessing the risk of ovarian cancer. It is supposed that the ROMA test will be used in conjunction with other clinical findings in patients with pelvic mass who were referred to an oncology center and scheduled for surgery. However, no ancillary pre-surgical clinicopathologic information was provided for evaluation beside or in combination with test results. Nor was information about referring physicians' independent pre-surgical assessment of the likelihood of ovarian cancer provided by the sponsor. Therefore, performance of the ROMA test can be evaluated only as a stand-alone test. This circumstance is contrary to the statement of intended use. Please note that the evaluation of a medical test as a stand alone test does not provide information about whether the medical test (clinically and statistically) improves patient assessment beyond what is possible with the available pre-surgical information alone.

3) Selection of the Cutoffs In the Validation Set

The sponsor obtained mathematical forms of the ROMA test for the pre-menopausal and post-menopausal women based on the training set:
PI = -12.0 + 2.38*LN(HE4) + 0.0626*LN(CA 125) for pre-menopausal women;
PI = -8.09 + 1.04*LN(HE4) + 0.732*LN(CA 125) for post-menopausal women;
and ROMA = exp(PI) / [1 + exp(PI)] (PI is "Predictive Index").

The cutoffs for defining "Low Risk" and "High Risk" were defined by the sponsor based on the pre-specified level of specificity of 75% separately for the pre-menopausal and post-menopausal subjects using the validation data set. Please note that the same study is used to find the cutoff for a medical test and to evaluate the clinical accuracy (sensitivity and specificity) of the medical test for this cutoff.
Comments on this approach:
a) If the level of, for example, specificity is pre-specified (for example, 75%) and the cutoff is selected as an unbiased estimate of the corresponding percentile (75th percentile) using only the subjects "Non-

Diseased" (Benign by pathology) of the validation set then the estimates of sensitivity and specificity with this cutoff using the validation set are unbiased (see paper by Kondratovich M.V. et al (2005)[1]).

b) If the cutoff is selected before the validation set (from earlier studies), then the conventional binomial statistical methods may be applied to estimate confidence limits for the estimates of sensitivity and specificity (because the cutoff is a constant for a validation data set). If the cutoff is an estimate of the percentile then the variability of the estimates of sensitivity and specificity with this cutoff is larger because the cutoff is a subject to variation (see Linnet K. (1987)[2]). The bootstrap technique can be used for estimation of confidence limits of sensitivity and specificity (see Platt R.W. et al (2000)[3].

The approach built into the sponsor's pivotal study was to select the cutoff as the 75th percentile using the subjects with "benign" pathology results. The estimates of sensitivity and specificity are unbiased, but they have larger variances compare to the binomial distribution calculation. Please note that all calculations of the confidence intervals below do not take into account the increase in variability due to selection of the cutoffs in the validation study. Confidence intervals for the sensitivity and specificity were calculated using binomial distribution (a score method[4]) and the confidence intervals for the predictive values were calculated based on the exact confidence intervals for the positive and negative likelihood ratios.

4) Performance of the ROMA Test as a Stand-Alone Test

A) Pre-menopausal Subjects

Estimation of cutoff for the ROMA test

There were 200 pre-menopausal subjects with pathology results "Benign". For the level of specificity of 75%, the estimate of 75th percentile was **13.4%[5].**

The ROMA test result: "Low Risk" if ROMA <13.4%

"High Risk" if ROMA≥ 13.4%.

Using a line listing dataset provided by the sponsor, the following table shows results of the ROMA test for the subjects with "Benign" disease and Epithelial Ovarian Cancer (EOC) or Low Malignant Potential (LMP).

Table 2

	Benign	LMP + Epithelial Ovarian Cancer				
		LMP	Epithelial Ovarian Cancer			
			EOC Stage I-II	EOC Stage III-IV	EOC Unstaged	
ROMA<13.4%	150	6	1	0	1	158
ROMA≥13.4%	50	10	6	9	1	76
		16	7	9	2	
	200	34				234

Estimates of ROMA performance for Benign vs LMP + EOC

Prevalence = 14.5% (34/234)

[1] Kondratovich, M.V. et al (2005) "Evaluation of Accuracy and Optimal Cutoff of Diagnostic Devices in the Same Study". *Proceedings of the 2005 Joint Statistical Meeting, Biopharmaceutical Section*, p. 2547-2551.

[2] Linnet K. (1987) "Comparison of Quantitative Diagnostic Tests: Type I Error, Power, and sample Size". *Statistics in Medicine*, Vol. 6, 147-158.

[3] Platt R.W. at al (2000) "Bootstrap Confidence Intervals for the Sensitivity of a Quantitative Diagnostic Test". *Statistics in Medicine*, Vol.19, 313-322.

[4] Altman D.A. et al (2000). *Statistics with Confidence.* 2nd ed. British Medical Journal; 2000

[5] Estimate of pth percentile is an order statistic p*(N+1). There are 200 subjects with "Benign" pathology. The estimate of 75th percentile is an order statistic 150.75 (=0.75 * (200+1)). X(150) is 13.31% and X(151) is 13.440%, then X(150.75) is X(150) + 0.75 *(X(151)-X(150))=13.4%.

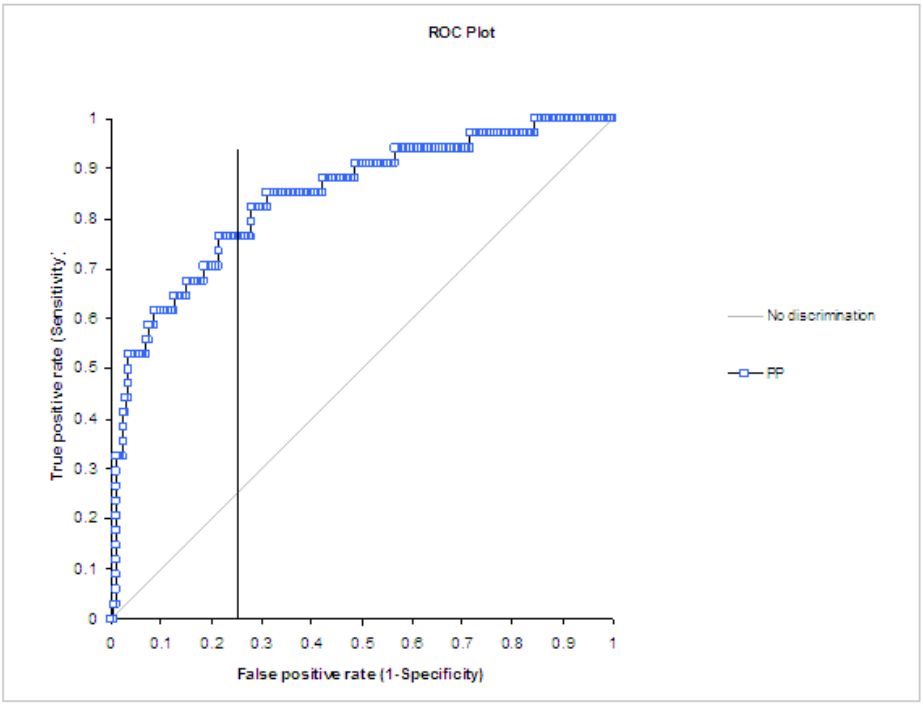
Table 3

	Estimate	95% CI
sensitivity	76.5% (26/34)	60.0% to 87.6%
specificity	75% (150/200)	68.6% to 80.5%
PPV	34.2% (26/76)	27.2% to 41.2%
NPV	94.9% (150/158)	91.6% to 97.3%

- Among the pre-menopausal subjects referred to oncology centers, 67.5% (158/234) subjects were characterized as "Low Risk" by ROMA test (and potentially would be referred back to their community gynecologist if the ROMA test were used as a stand-alone test).
- *For the subjects with "Low Risk" by the ROMA test, there is a risk of 5.1% (1-NPV) that a subject has Epithelial Ovarian Cancer or LMP.* This risk may be high as 8.4% (see the lower limit of 95% CI of NPV above). In other words, among 100 pre-menopausal subjects with "Low Risk" by the ROMA, approximately 5 subjects have EOC or LMP.
- Among the subjects with Epithelial Ovarian Cancer or LMP, 23.5% (1-Sensitivity) subjects were evaluated as "Low Risk" subjects by the ROMA test. Sensitivity was 76.5% (26/34) with a lower limit of 95% CI of 60.0%.
- Among the subjects with LMP, 37.5% (6 out of 16) subjects were evaluated as "Low Risk" by the ROMA. Sensitivity for LMP patients was 62.5% (10/16) with a lower limit of 95% CI of 38.6%.
- Among the subjects with Epithelial Ovarian Cancer Stage I-II, 14.3% (1 out of 7) subjects were evaluated as "Low Risk" subjects by the ROMA test. Sensitivity for Stage I-II EOC was 85.7% (6/7) with a lower limit of 95% CI of 48.7%.

The ROC curve for the pre-menopausal subjects is presented in Figure 1 below.

Figure 1. Pre-menopausal subjects: Benign vs LMP+EOC



AUC for the ROMA test is 0.843 with 95% CI: 0.766 to 0.920.

B) Post-menopausal Subjects

Estimation of cutoff for the ROMA test
There were 151 post-menopausal subjects with pathology results "Benign". For the level of specificity of 75%, the estimate of 75th percentile was **27.7%**[6]**.**
The ROMA test result: "Low Risk" if ROMA <27.7%
"High Risk" if ROMA≥ 27.7%.

Using a line listing dataset provided by the sponsor, the following table shows results of the ROMA test for the subjects with "Benign" disease and Epithelial Ovarian Cancer (EOC) or Low Malignant Potential (LMP).

Table 4

	Benign	LMP + Epithelial Ovarian Cancer				
		LMP	Epithelial Ovarian Cancer			
			EOC Stage I-II	EOC Stage III-IV	EOC Unstaged	
ROMA<27.7%	113	3	4	1	1	122
ROMA≥27.7%	38	4	25	80	1	148
		7	29	81	2	
	151	119				270

Estimates of ROMA performance for Benign vs LMP + EOC
Prevalence = 44.1% (119/270)

Table 5

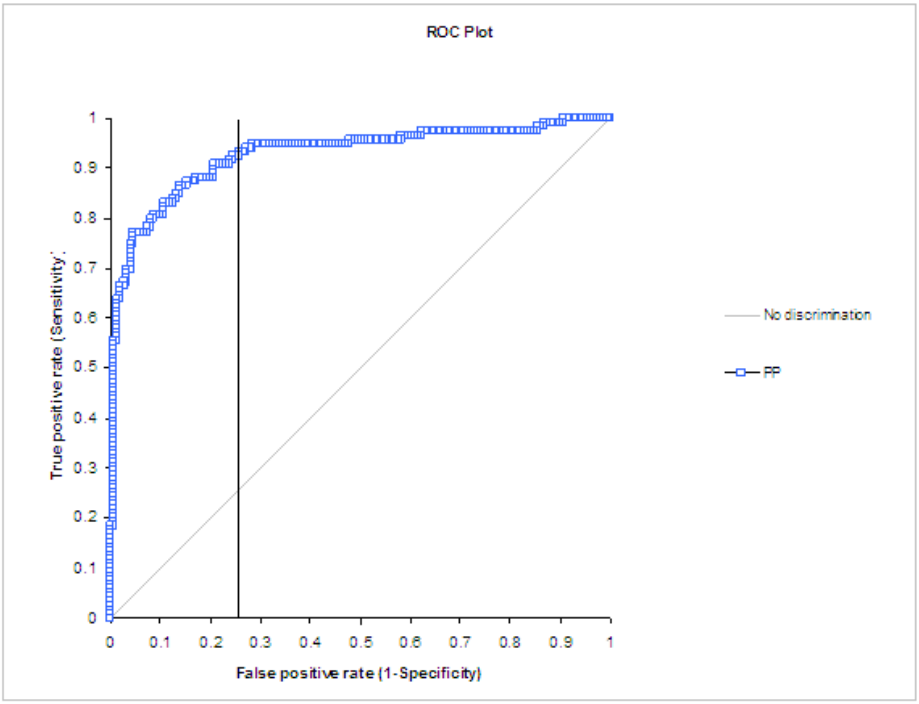
	Estimate	**95% CI**
sensitivity	92.4% (110/119)	86.3% to 96.0%
specificity	74.8% (113/151)	67.4% to 81.1%
PPV	74.3% (110/148)	68.9% to 79.5%
NPV	92.6% (113/122)	87.3% to 96.0%

- Among the post-menopausal subjects referred to oncology centers, 45.2% (122/270) subjects were characterized as "Low Risk" by ROMA test (and potentially would be referred back to their community gynecologist if the ROMA test were used as a stand-alone test).
- *For the subjects with "Low Risk" by the ROMA test, there is a risk of 7.4% (1-NPV) that a subject has Epithelial Ovarian Cancer or LMP.* This risk may be high as 12.7% (see the lower limit of 95% CI of NPV above). In other words, among 100 post-menopausal subjects with "Low Risk" by ROMA, approximately 7 subjects have EOC or LMP.
- Among the subjects with Epithelial Ovarian Cancer or LMP, 7.6% (1-Sensitivity) subjects were evaluated as "Low Risk" subjects by the ROMA test. Sensitivity was 92.4% (110/119) with a lower limit of 95% CI of 86.3%.
- Among the subjects with LMP, 42.9% (3 out of 7) subjects were evaluated as "Low Risk" subjects by the ROMA test. Sensitivity for LMP was 57.1% (4/7) with a lower limit of 95% CI of 25.0%.
- Among the subjects with Epithelial Ovarian Cancer Stage I-II, 13.8% (4 out of 29) subjects were evaluated as "Low Risk" subjects by the ROMA test. Sensitivity for Stage I-II EOC was 86.2% (25/29) with a lower limit of 95% CI of 69.4%.

[6] There are 151 post-menopausal patients with "Benign" pathology. The 75th percentile is 114th order statistic: 0.75*(151+1). X(114) is 27.7%.

The ROC curve for the post-menopausal subjects is presented in Figure 2 below.

Figure 2. Post-menopausal Patients: Benign vs LMP+EOC



AUC for the ROMA test is 0.927 with 95% CI: 0.893 to 0.961.

C) Performance of the ROMA <u>Qualitative</u> Test for Pre-menopausal and Post-menopausal Subjects Combined

The sponsor calculated performance of the ROMA test for the combined data (pre-menopausal +post-menopausal subjects). Though stratified performance characteristics are more informative, a valid analysis of the combined dataset is possible when ROMA is evaluated as a qualitative test. See comments in section 7 about defects in ROC analyses that combine quantitative ROMA results for the pre- and post-menopausal patients.

Table 6

	Benign	LMP + Epithelial Ovarian Cancer	Total
ROMA Low Risk	263	17	280
ROMA High Risk	88	136	224
Total	351	153	504

Estimates of ROMA performance for Benign vs LMP + EOC
Prevalence = 30.4% (153/504)

Table 7

	Estimate	**95% CI**
sensitivity	88.9% (136/153)	82.9% to 93.0%
specificity	74.9% (263/351)	70.1% to 79.2%
PPV	60.7% (136/224)	56.2% to 65.2%
NPV	93.9% (263/280)	90.9% to 96.1%

Clinical interpretation of the performance of the ROMA test for the combined data should be done carefully. For example, sensitivity was
76.5% (26/34) for the pre-menopausal subjects (95% CI: 60.0% to 87.6%) and
92.4% (110/119) for the post-menopausal subjects (95% CI: 86.3% to 96.0%).
Combined sensitivity was 88.9% (136/153) with 95% CI: 82.9% to 93.0%.
Please note that:

a) Sensitivities of the ROMA test for the pre-menopausal and post-menopausal subjects were statistically different: 76.5% vs 92.4%. The difference in sensitivities between post-menopausal and pre-menopausal patients was 16.0% with 95% CI: 2.5% to 33.1%.
b) In the combined data set, there were 153 subjects with "Epithelial Ovarian Cancer & LMP" (34 pre-menopausal subjects + 119 post-menopausal subjects). So, among the subjects with LMP+EOC, the post-menopausal subjects comprise 78%. This means, when one calculates the sensitivity of the combined data, that the weighted combination of sensitivity for pre-menopausal subjects and sensitivity for the post-menopausal subjects is:
$Sensitivity_{Combined} = 0.22 \times Sen_{Premeno} + 0.78 \times Sen_{Postmeno}$.
Therefore, *the combined sensitivity is difficult to interpret from the clinical point of view when the sensitivities for two groups of subjects (pre-menopausal and post-menopausal) are different.*

Similarly, for NPVs:
NPV = 94.9% for 158 pre-menopausal subjects with "Low Risk" and
NPV = 92.6% for the 122 post-menopausal subjects with "Low Risk".
NPV = 93.9% for the 280 pre- and post-menopausal subjects combined =>
$NPV_{Combined} = 0.56 \times NPV_{Premeno} + 0.44 \times NPV_{Postmeno}$

5) Performances of the ROMA Test vs CA125 alone vs HE4 alone

The ROMA test for each patient is a combination of the CA125 and HE4 values of this patient, accounting also for menopausal status. Let us investigate *whether the ROMA test with the specified premenopausal and postmenopausal cutoffs outperforms the single analytes included in the combination.*

A) Pre-Menopausal Subjects

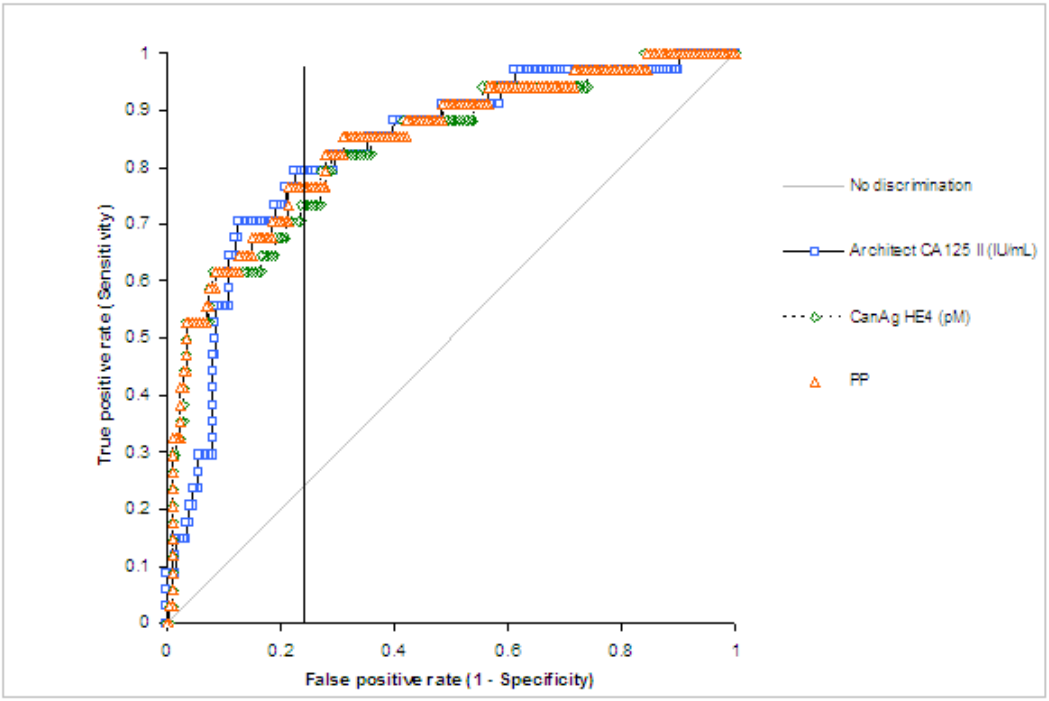
The cutoffs corresponding to the level of specificity of 75% are:

Table 8

Medical Test	Cutoff corresponding to specificity of 75%
ROMA	13.4%
CA125 alone	60.4 IU/mL[7]
HE4 alone	63.6 pM[8]

The ROC curves for the ROMA test, CA125 alone, and HE4 alone for Benign vs LMP+EOC are presented in Figure 3.

Figure 3. Pre-menopausal Patients: Benign vs LMP+EOC



There were 34 pre-menopausal patients with LMP or EOC. With the cutoffs corresponding to specificity of 75%, 26 patients were "High Risk" by the ROMA test, 27 patients were Positive (≥60.4 IU/mL) by the CA125 alone and 25 patients were Positive (≥63.6 pM) by the HE4 alone.

[7] There are 200 pre-menopausal patients with "Benign" pathology. 75th percentile is an order statistic X(150.75). For the CA125 values, X(150) is 60.2 IU/mL and X(151) is 60.5 IU/mL; then X(150.75) is 60.4 IU/mL.

[8] For the HE4 values, X(150) is 63.442 pM and X(151) is 63.700 pM; then X(150.75) is 63.6 pM.

Table 9

Medical test	Sensitivity (level of specificity=75%)
ROMA	76.5% (26/34)
CA125 alone	79.4% (27/34)
HE4 alone	73.5% (25/34)

Performance of the ROMA and CA125 tests for the 34 pre-menopausal patients with LMP+EOC is presented in the table below.

Table 10

	CA125≥60.4 IU/mL	CA125<60.4 IU/mL	Total
ROMA≥13.4%	23	3	26
ROMA<13.4%	4	4	
Total	27		34

$Sensitivity_{ROMA}$ – $Sensitivity_{CA125\ alone}$ was -2.9% (-1/34) with 95% CI: -19.3% to 13.5%.

Conclusion:
The data of the clinical study did not demonstrate that there is a statistically significant contribution of the HE4 test beyond the CA125 in the combination ROMA for the pre-menopausal women. Indeed, for the same level of specificity of 75%, sensitivity of CA125 alone is 79.5% and sensitivity of the combination of CA125 and HE4 (the ROMA test) is 76.5%.

B) Post-Menopausal Subjects

The cutoffs corresponding to the level of specificity of 75% are:

Table 11

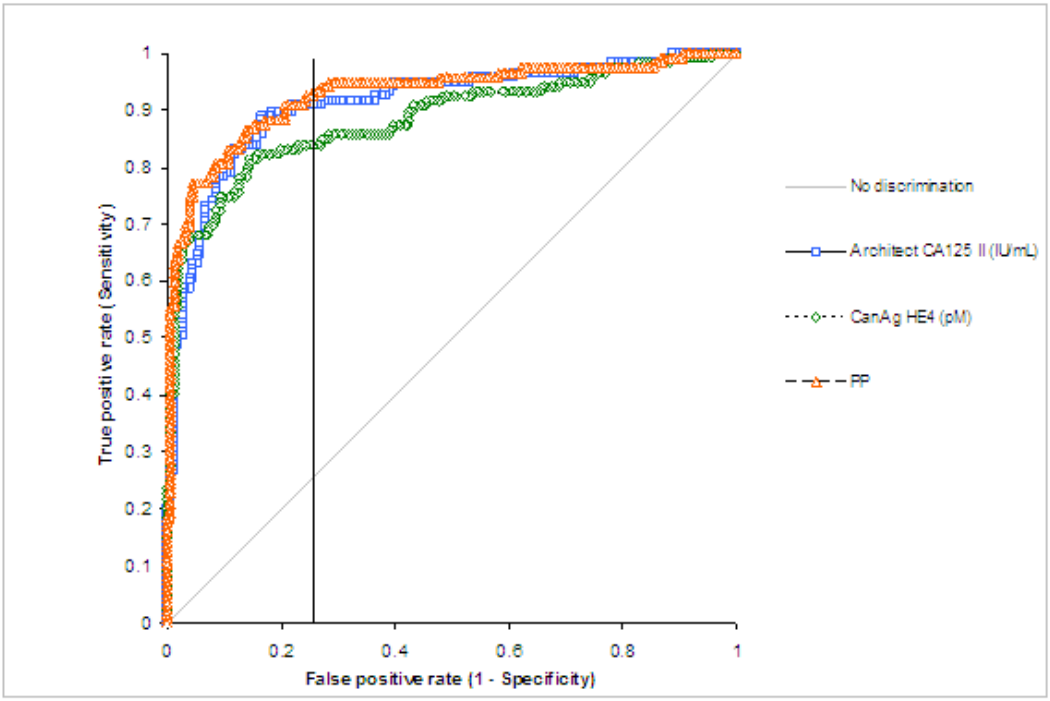
Medical Test	Cutoff corresponding to specificity of 75%
ROMA	27.7%
CA125 alone	30.0 IU/mL[9]
HE4 alone	102.7 pM[10]

The ROC curves for the ROMA test, CA125 alone, and HE4 alone for Benign vs LMP+EOC are presented in Figure 4.

[9] For the CA125 values, the estimate of 75th percentile is X(114)=30.0 IU/mL.

[10] For the HE4 values, the estimate of 75th percentile is X(114)=102.7 pM.

Figure 4. Post-menopausal Patients: Benign vs LMP+EOC



There were 119 post-menopausal patients with LMP or EOC. With the cutoffs corresponding to specificity of 75%, 110 patients were "High Risk" by the ROMA test, 108 patients were Positive (≥30.0 IU/mL) by the CA125 alone and 100 patients were Positive (≥102.7 pM) by the HE4 alone.

Table 12

Medical test	Sensitivity (level of specificity=75%)
ROMA	92.4% (110/119)
CA125 alone	90.8% (108/119)
HE4 alone	84.0% (100/119)

Performance of the ROMA and CA125 tests for the 119 post-menopausal patients with LMP+EOC is presented in the table below.

Table 13

	CA125≥30.0 IU/mL	CA125<30.0 IU/mL	Total
ROMA≥27.7%	107	3	110
ROMA<27.7%	1	8	
Total	108		119

$Sensitivity_{ROMA}$ – $Sensitivity_{CA125\ alone}$ was +1.7% (2/119) with 95% CI: -2.6% to 6.3%.

Conclusion:

The data of the clinical study did not demonstrate that there is a statistically significant contribution of the HE4 test beyond the CA125 in the combination ROMA for the post-menopausal women. For the same level of specificity of 75%, sensitivity of CA125 alone is 90.8% and sensitivity of the combination of CA125 and HE4 (the ROMA test) is 92.4%.

6) Investigation of Performances of the ROMA, CA125 and HE4 tests for the Patients with Stages I-II Epithelial Ovarian Cancer

A) Pre-Menopausal Subjects

Among 34 pre-menopausal women with LMP+EOC, there were 16 patients with LMP, 7 patients with Stage I-II EOC, 9 patients with Stage III-IV EOC and 2 patients with unstaged EOC.

Table 14

	Sensitivity (level of specificity=75%)				
	LMP+EOC **N=34**	**LMP** **N_1=16**	**EOC Stage I-II** **N_2=7**	**EOC Stage III-IV** **N_3=9**	**EOC unstaged** **N_4=2**
ROMA	76.5% (26/34)	62.5% (10/16)	85.7% (6/7)	100% (9/9)	50% (1/2)
CA125	79.4% (27/34)	75.0% (12/16)	57.1% (4/7)	100% (9/9)	100% (2/2)
HE4	73.5% (25/24)	56.3% (9/16)	85.7% (6/7)	100% (9/9)	50% (1/2)

Figure 5. Pre-menopausal Patients: Benign vs LMP (N_1=16)

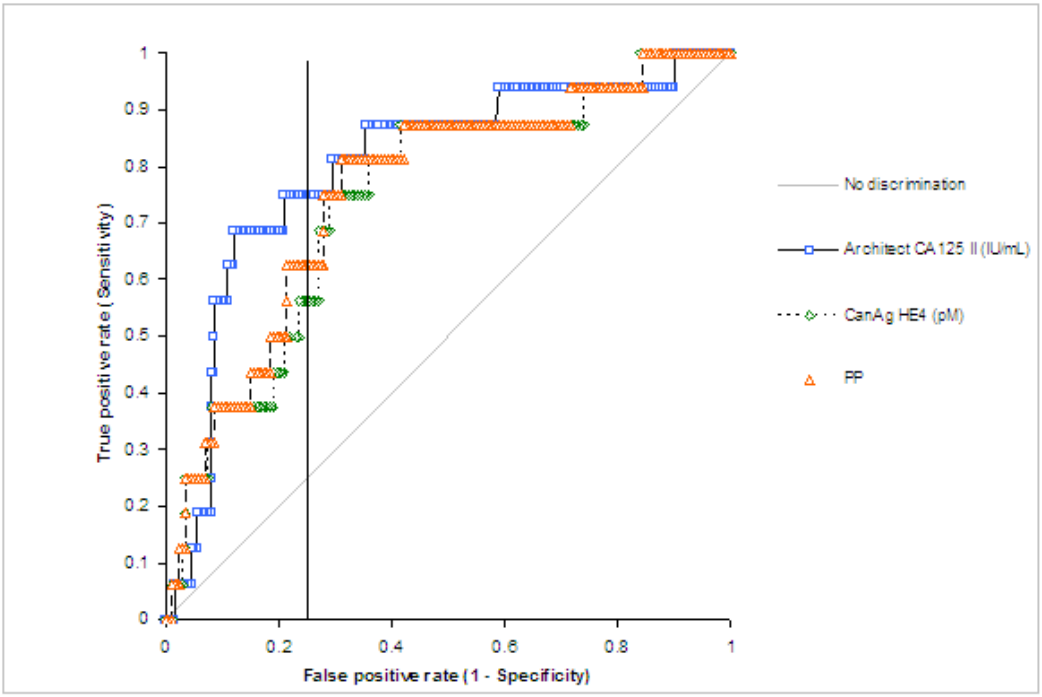


Table 15

Medical test	AUC	95% CI
ROMA	**0.758**	0.634 to 0.882
CA125	**0.801**	0.680 to 0.922
HE4	0.745	0.620 to 0.870

Difference in AUC between the ROMA and CA125 is not statistically significant.

Figure 6. Pre-Menopausal Patients Benign vs Stage I-II EOC (N_2=7)

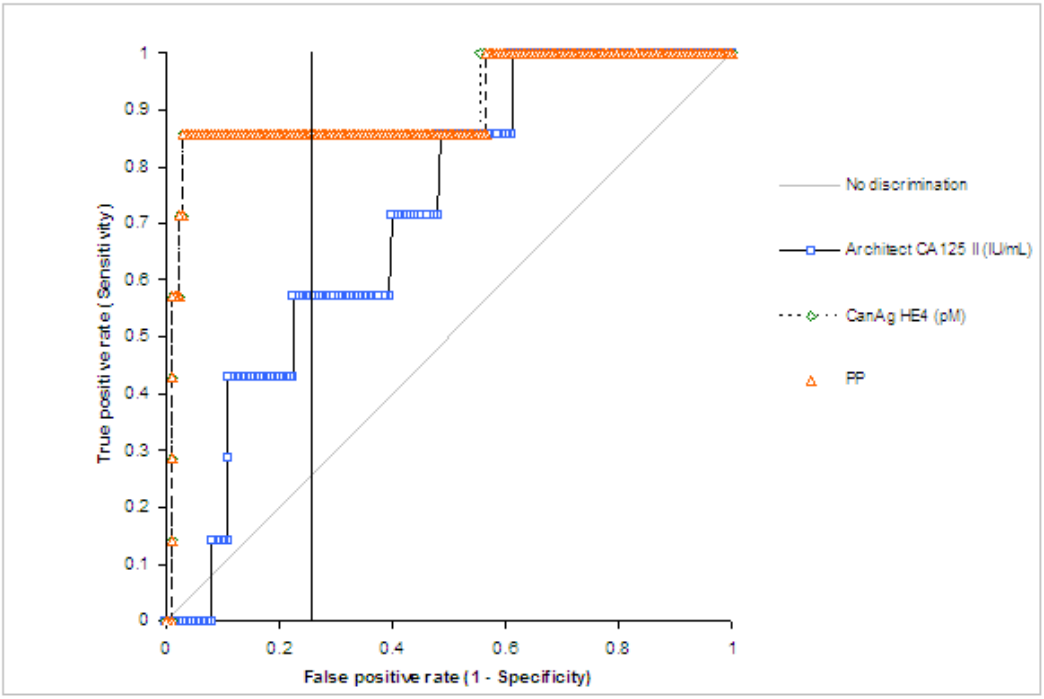


Table 16

Medical test	AUC	95% CI
ROMA	**0.906**	0.751 to 1.00
CA125	**0.712**	0.551 to 0.873
HE4	0.907	0.751 to 1.00

- Difference in AUC between the ROMA and CA125 is 0.195 with 95% CI: 0.024 to 0.366 (p-value=0.025).

- AUC presents a sensitivity averaged over all specificities, while the ROMA test is used with the cutoff corresponding to specificity of 75%. At this cutoff, there is no statistically significant difference between sensitivities of ROMA and CA125:

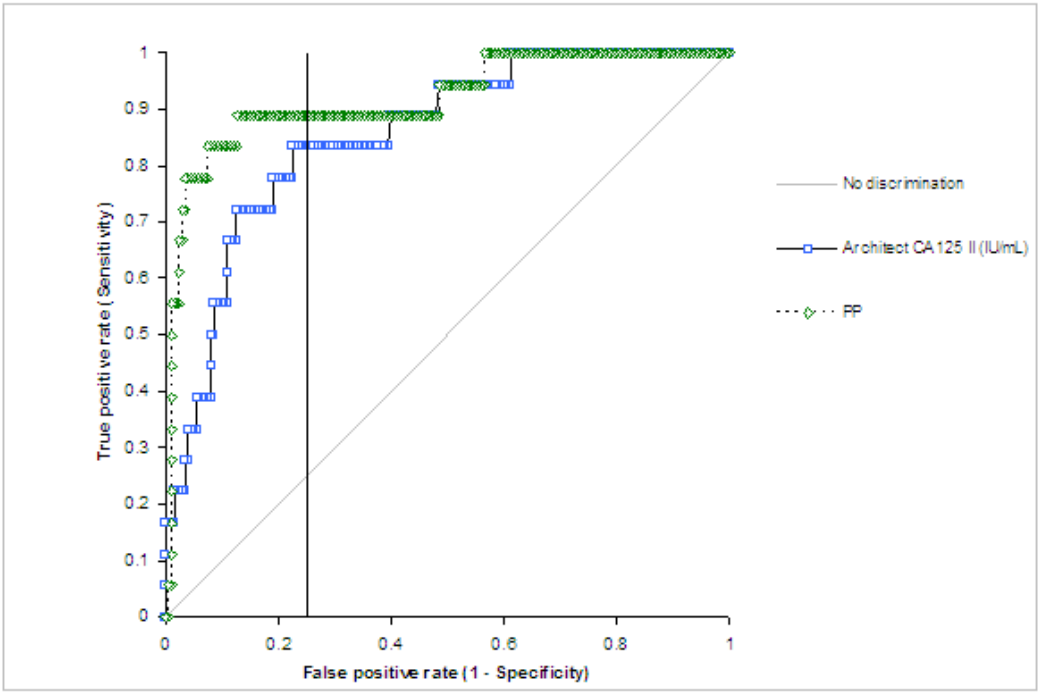
Table 17

	CA125≥60.4 IU/mL	CA125<60.4 IU/mL	Total
ROMA≥13.4%	4	2	6
ROMA<13.4%	0	1	
Total	4		7

- $\text{Sensitivity}_{\text{ROMA}}$ – $\text{Sensitivity}_{\text{CA125 alone}}$ was +28.6% (2/7) with 95% CI: -16.0% to 62.1%.

There were 2 patients in the study with unstaged EOC. Treating these subjects as Stage I-II, the ROC curves for Benign vs (EOC Stage I-II + 2 unstaged) are presented in Figure 7. The difference between AUC was not statistically significant (p-value=0.158).

Figure 7. Pre-menopausal Patients: Benign vs (EOC Stage I-II + 2 unstaged EOC)



Conclusion:

- *The difference in AUC between the ROMA test and CA125 for the pre-menopausal patients with Stage I-II of EOC was statistically significant (p-value is 0.025). But the AUC presents a sensitivity averaged over all specificities. The ROC curves of the ROMA and CA125 tests can be different in the regions that are clinically useless (low levels of sensitivity).*
- *For the pre-menopausal patients with Stage I-II of EOC, there was an observed improvement in sensitivity of +28.6% (2 out of 7) for the ROMA test vs CA125 test with the cutoffs corresponding specificity of 75% (6/7 vs 4/7). This improvement in sensitivity for Stage I-II of EOC was not statistically significant.*

B) Post-Menopausal Subjects

Among 119 post-menopausal women with LMP+EOC, there were 7 patients with LMP, 29 patients with Stage I-II EOC, 81 patients with Stage III-IV EOC and 2 patients with unstaged EOC.

Table 18

	Sensitivity (level of specificity=75%)				
	LMP+EOC **N=119**	**LMP** **N_1=7**	**EOC** **Stage I-II** **N_2=29**	**EOC** **Stage III-IV** **N_3=81**	**EOC** **unstaged** **N_4=2**
ROMA	92.4% (110/119)	57.1% (4/7)	86.2% (25/29)	98.8% (80/81)	50% (1/2)
CA125	90.8% (108/119)	42.9% (3/7)	82.8% (24/29)	98.8% (80/81)	50% (1/2)
HE4	84.0% (100/119)	42.9% (3/7)	75.9% (22/29)	90.1% (73/81)	100% (2/2)

Figure 8. Post-menopausal Women Benign vs LMP (N_1=7)

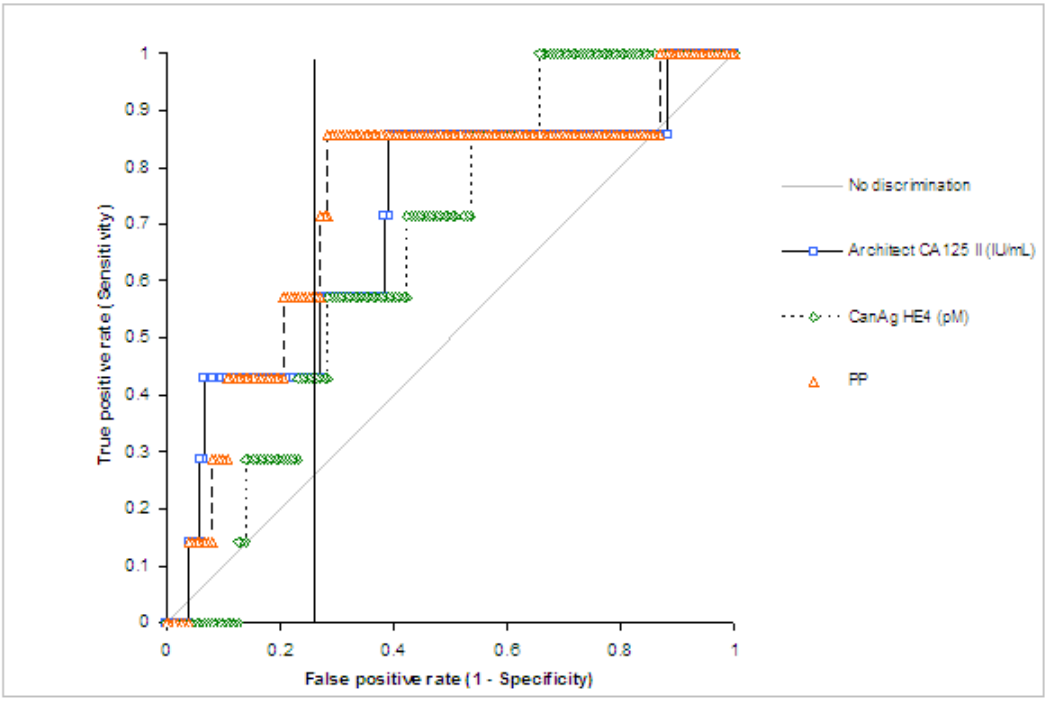


Table 19

Medical test	AUC	95% CI
ROMA	**0.735**	0.522 to 0.948
CA125	**0.701**	0.476 to 0.926
HE4	0.658	0.498 to 0.818

Difference in AUC between the ROMA and CA125 is not statistically significant.

Figure 9. Post-menopausal Patients: Benign vs EOC Stage I-II

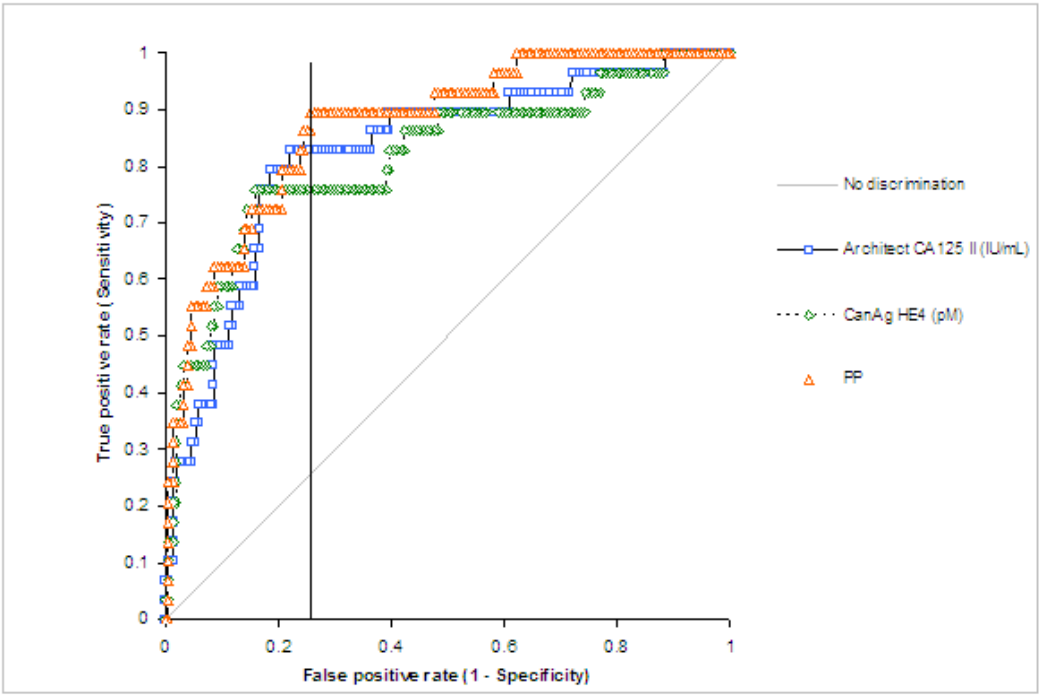


Table 20

Medical test	AUC	95% CI
ROMA	**0.871**	0.802 to 0.939
CA125	**0.826**	0.739 to 0.913
HE4	0.817	0.721 to 0.913

- Difference in AUC between the ROMA and CA125 is 0.045 with 95% CI: -0.007 to 0.097 (p-value=0.087). Difference in AUC is not statistically significant.

The ROMA test is used with the cutoff corresponding to specificity of 75%. At this cutoff, there is no statistically significant difference between sensitivities of ROMA and CA125:

Table 21

	CA125≥30.0 IU/mL	CA125<30.0 IU/mL	Total
ROMA≥27.7%	23	2	25
ROMA<27.7%	1	3	
Total	24		29

- $\text{Sensitivity}_{\text{ROMA}}$ - $\text{Sensitivity}_{\text{CA125 alone}}$ was +3.4% (1/29) with 95% CI: -11.4% to 18.6%. The difference in sensitivities was not statistically significant.

Conclusion:

- *The difference in AUC for the ROMA and CA125 for the Stage I-II of EOC was not statistically significant.*
- *For the post--menopausal patients with Stage I-II of EOC, there was an observed improvement in sensitivity of +3.4% (1 out of 29) for the ROMA test vs CA125 test with the cutoffs corresponding specificity of 75% (25/29 vs 24/29). This improvement in sensitivity for Stage I-II of EOC was not statistically significant.*

c) Pre-Menopausal and Post-Menopausal Subjects Combined

As described already (see section 7, below), ROC curves combining the pre-menopausal and post-menopausal patients in this data set have defects. Still, an informative qualitative analysis comparing sensitivities can be performed, as in Table 22.

Table 22

36 patients with Stage I-II of EOC (pre- and post-menopausal)

	CA125 Positive	CA125 Negative	Total
ROMA High Risk	27	4	31
ROMA Low Risk	1	4	
Total	28		36

$Sensitivity_{ROMA}$ - $Sensitivity_{CA125\ alone}$ was +8.3% (3/36) with 95% CI: -5.3% to 22.4%.

There is no statistically significant improvement in sensitivity of the combination of CA125 and HE4 (ROMA test) vs CA125 for the pre-menopausal and post-menopausal patients of Stage I-II combined.

7) Misuse of ROC Analysis and Area Under ROC Curve

The sponsor used ROC analysis for the evaluation of quantitative values of the ROMA, CA125, and HE4 tests for the pre-menopausal and post-menopausal patients combined (see Figures 11-14 in the sponsor's HE4 clinical report).

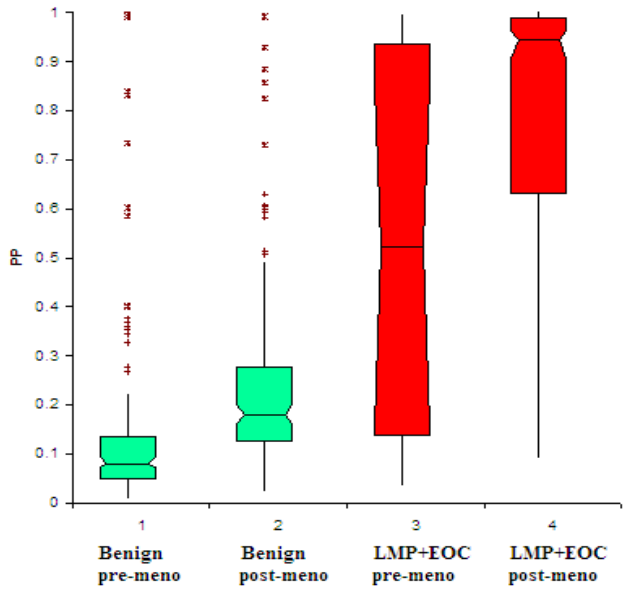
The ROC analysis for the combined data has logical flaws.

As an example, consider the quantitative ROMA tests results for the pre-menopausal and post-menopausal subjects. (The same logical flaws affect combined ROC curves for CA125 and HE4 quantitative tests.)

In this data set, the quantitative results of the ROMA test for the post-menopausal subjects with "Benign" pathology are usually higher than the ROMA test results for the pre-menopausal subjects with "Benign" pathology. Likewise, the quantitative ROMA test results for the post-menopausal subjects with LMP+EOC is usually higher than the ROMA test results for the pre-menopausal subjects with LMP+EOC.

Box-and-whisker plots on Figure 10 display the distribution of the ROMA values for pre- and post-menopausal patients.

Figure 10. Box-and-whisker plots



For example, the median of ROMA results for the patients in the clinical study are

Table 23

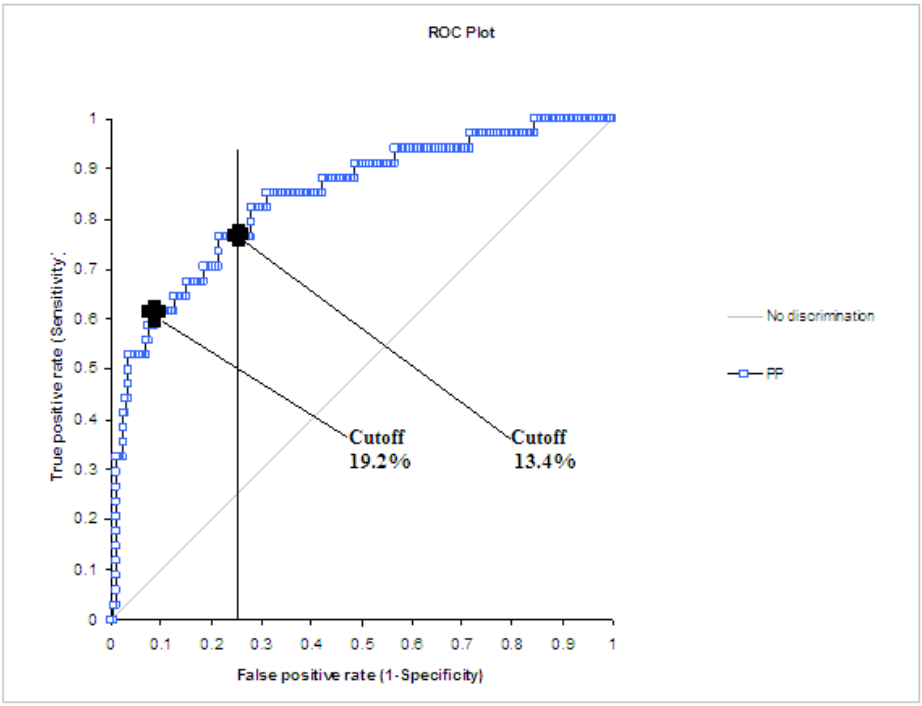
Patients	Median ROMA
Benign pre-menopausal (n_1=200)	8.0%
Benign post-menopausal (n_2=151)	17.9%
LMP+EOC pre-menopausal (m_1=34)	52.3%
LMP+EOC post-menopausal (m_2=119)	94.4%

A) ROC Curve for the Combined Data

The construction of the ROC curve for the combined data implies indirectly that the pre-menopausal and post-menopausal subjects have the same cutoff. Indeed, an ROC curve is a plot of (1-specificity) (calculated based on the ROMA results of the "Benign" pre-menopausal and post-menopausal subjects combined) versus Sensitivity (calculated based on the ROMA results of "LMP+EOC" pre-menopausal and post-menopausal subjects combined) for the different possible cutoffs. As the cut-off moves through the range of possible values, the ROC curve is generated. Consideration of one cutoff for the pre-menopausal and post-menopausal subjects contradicts to the suggested use of the ROMA test with different cutoffs for the pre-menopausal and post-menopausal patients (13.4% for the pre-menopausal and 27.7% for the post-menopausal patients). Therefore, the *ROC curve for the combined data with the same cutoff for the pre- and post-menopausal patients is difficult for interpretation.*

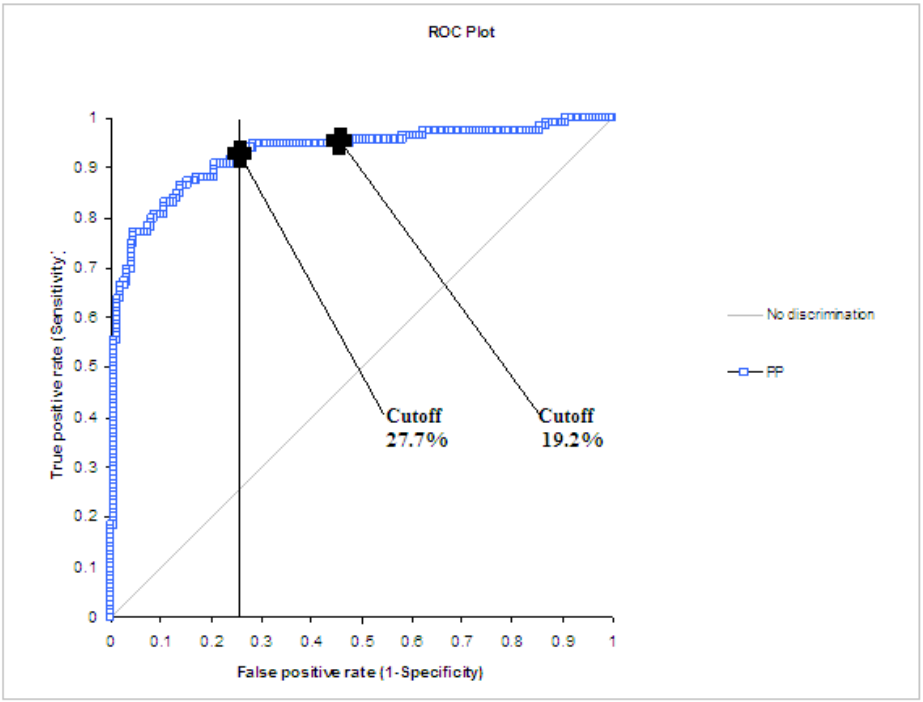
- Using the percentile-based method adopted by sponsor to ensure 75% specificity for the test, the cutoff (based on the 351 "Benign" subjects [200 pre- and 151 post- menopausal subjects]) for the combined data set is 19.2%. This one cutoff is applied to the predictive probabilities that are provided by the two ROMA equations that treat the two menopausal statuses.
- For the pre-menopausal patients, the 19.2% cutoff corresponds to specificity of 89.0% (178/200) in that patient group. This exceeds the 75% design specification, while providing a sensitivity of only 61.8% (21/34).

Figure 11. Pre-menopausal patients



- For the post-menopausal patients, the 19.2% cutoff corresponds to specificity of 56.3% (85/151) in that patient group. This fall short of the 75% design specification, while providing an increased sensitivity of 95.0% (113/119).

Figure 12. Post-menopausal patients



With these divergent results for premenopausal and postmenopausal subgroups, the usual insight provided by an ROC curve (i.e., the trade-off between a test's sensitivity and specificity) becomes garbled. The effect of movement along the ROC curve is opposite for the two subpopulations, in a complex manner. The clinical implications of such behavior for the overall population are difficult to discern.

B) AUC for the pre-menopausal and post-menopausal patients combined

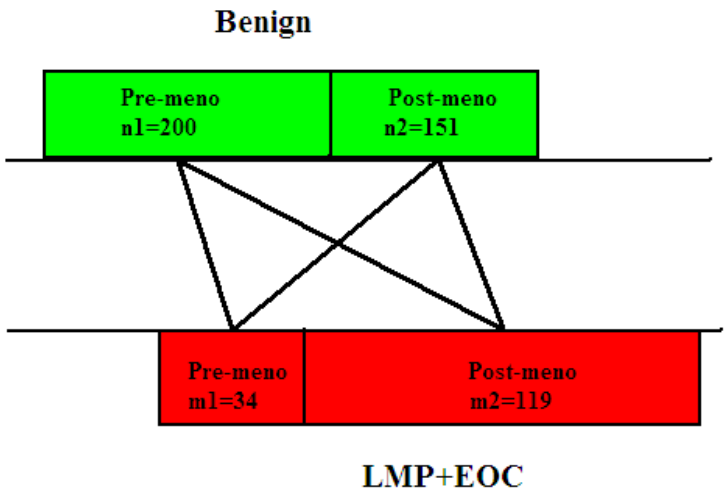
AUC expresses the probability that a randomly selected Benign patient (X) has a lower medical test result than a randomly selected diseased (LMP+EOC) patient (Y): (Prob(X<Y)). AUC is easily estimated as Wilcoxon statistic. For the N benign subjects and M diseased subjects, the Wilcoxon statistic estimates the probability Prob(X<Y). This is based on a count, for each benign subject (Xi), of how many malignant subjects (Y_j) have results higher than the result for that benign subject.

$$W = AUC = \frac{1}{N \cdot M} \sum_i \sum_j I(X_i, Y_j),$$

$$where\ I(X_i, Y_j) = 1 \text{ if } X_i < Y_j;\ \ 1/2 \text{ if } X_i = Y_j;\ 0 \text{ if } X_i > Y_j.$$

Because Xi and Yj are not specific to either population of patients, the area under the ROC curve for the benign pre-menopausal and post-menopausal subjects combined and malignant pre-menopausal and post-menopausal subjects combined can be presented by a sum of four terms (instead of the single term associated with the calculation for one population):

Figure 13



1) ROMA results of the benign pre-menopausal subjects are compared with ROMA results of the malignant pre-menopausal subjects;
2) ROMA results of the benign pre-menopausal subjects are compared with ROMA results of the malignant post-menopausal subjects;
3) ROMA results of the benign post-menopausal subjects are compared with ROMA results of the malignant pre-menopausal subjects; and
4) ROMA results of the benign post-menopausal subjects are compared with ROMA results of the malignant post-menopausal subjects.

There is no apparent biological basis for comparing pre-menopausal benign subjects with post-menopausal malignant subjects and post-menopausal benign subjects with pre-menopausal malignant subjects. The clinical meaning of an AUC built from such comparisons is undefined.

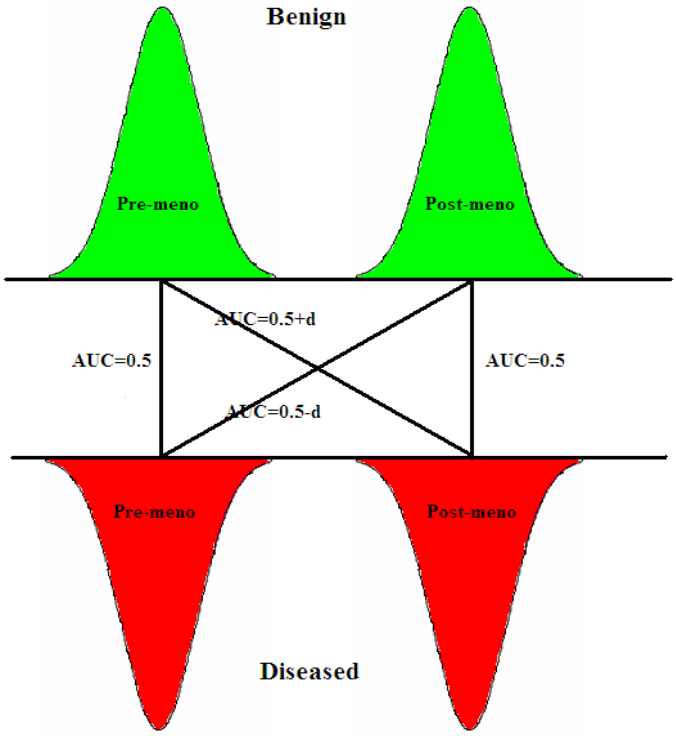
The effect of such ill-posed comparisons in calculating an AUC can be illustrated by an example whose characteristics make it absurd to use AUC as a measure of diagnostic accuracy. Consider a hypothetical medical test T, applied to premenopausal and postmenopausal, benign and malignant populations with sizes as in Figure 13, but which discriminates neither benign from malignant pre-menopausal subjects nor benign from malignant post-menopausal subjects. If the distributions for the premenopausal and postmenopausal populations differ, then the calculation can yield an absurd result (i.e. AUC > 0.5 for this uninformative test).

The four terms in the AUC for the combined data are included with weights corresponding to the proportions of pre-menopausal and postmenopausal subjects among the benign subjects and to the proportions of pre-menopausal and post-menopausal subjects among the malignant subjects, as in the following equation.

$$AUC = \frac{n_1}{n_1+n_2}\frac{m_1}{m_1+m_2}AUC_{\mathrm{Pre},\mathrm{Pre}} + \frac{n_1}{n_1+n_2}\frac{m_2}{m_1+m_2}AUC_{\mathrm{Pre},\mathrm{Post}} + \frac{n_2}{n_1+n_2}\frac{m_1}{m_1+m_2}AUC_{\mathrm{Post},\mathrm{Pre}} + \frac{n_2}{n_1+n_2}\frac{m_2}{m_1+m_2}AUC_{\mathrm{Post},\mathrm{Post}}$$

where n_1, n_2, m_1, *and* m_2 *are quantities depicted in Figure* 13.

Figure 14. Hypothetical example of the medical test T, which discriminates neither pre-menopausal nor post-menopausal patients with respect to benign or malignant cases



- Because the medical test T does not discriminate the pre-menopausal subjects, the $AUC_{Pre,Pre}$ is 0.5.
- Likewise, failure of the medical test T to discriminate the post-menopausal subjects means that $AUC_{Post,Post}$ is 0.5.
- But the $AUC_{Pre,Post}$ is larger than 0.5 because the pre-menopausal benign subjects have test results lower than the post-menopausal malignant subjects; and
- the $AUC_{Post,\ Pre}$ is smaller than 0.5 because the post-menopausal benign subjects have test results higher than the pre-menopausal malignant subjects. (see Figure 14 above).

The weights of these four AUCs differ. The weight of $AUC_{Pre,Post}$ is the largest because the proportion of pre-menopausal benign subjects is the largest among the benign subjects, 0.570=200/351, and the proportion of post-menopausal diseased subjects is the largest among the diseased subjects, 0.778=119/163.
So, AUC for combined data is
0.570 x 0.222 x 0.5 + 0.570 x 0.778 x (0.5+d) +0.430 x 0.222 x (0.5-d) +0.430 x 0.778 x 0.5 =
0.5 + d x (0.570 x 0.778 – 0.430 x 0.222) > 0.5, an absurd result for a non-informative test.

The **AUC for the combined data is usually overstated,** and does not reflect the diagnostic accuracy of the medical test.

Conclusion:

The ROC analysis for the pre-menopausal and post-menopausal patients combined has logical flaws and therefore, it is not an appropriate statistical tool for use in evaluating the test's performance characteristics. In ROC analysis, pre-menopausal and post-menopausal patients should be used separately to estimate test performance characteristics.