# FDA Executive Summary

Lead Reviewers:
Mridu Virmani, Ph.D.
Joyce M. Whang, Ph.D.

*prepared for the*
*May 17, 2005, meeting of the*
*Obstetrics & Gynecology Devices Panel*

*corrected and redacted, April 29, 2005*

P040028
MediSpectra, Inc.
LUMA™ Cervical Imaging System

## Contents

I. Indication for use

II. Device Description

- II. A. Device Components
- II. B. Device Input Data
- II. C. Display Output
- II. D. Algorithm

III. Pre-clinical Studies

- III. A. Pre-clinical Testing
- III. B. Pre-clinical Review Issues
  - III. B. 1. Optical Radiation and Safety Review Issues
  - III. B. 2. Algorithm Review Issues

IV. Overview of Pivotal Studies I and II (PS I and PS II)

V. Pivotal Study I (PS I)

- V. A. PS I Study Design
  - V. A. 1. PS I Primary Effectiveness Endpoints
  - V. A. 2. PS I Secondary Effectiveness Endpoints
  - V. A. 3. PS I Inclusion and Exclusion Criteria
  - V. A. 4. PS I Procedure
- V. B. PS I Results: Overview and Primary Endpoints
  - V. B. 1. PS I Safety Results
  - V. B. 2. PS I Primary Effectiveness Endpoints
- V. C. PS I Results: Additional Panel Review Issues
  - V. C. 1. PS I Analysis by Pap Stratum
  - V. C. 2. PS I Joint Analysis
  - V. C. 3. PS I Biopsy-Level PPV
  - V. C. 4. PS I False Negatives
  - V. C. 5. PS I Relative Increase in TP (or Sensitivity) and FP
  - V. C. 6. PS I Absolute Increase in Sensitivity
  - V. C. 7. PS I Age-Related Device Performance
- V. D. PS I Results: Other Issues
  - V. D. 1. PS I Reader Variability
  - V. D. 2. PS I Site Variability
  - V. D. 3. PS I Study Design: Performance Bias

VI. Pivotal Study II (PS II)

VI. A. PS II Study Design
- VI. A. 1. PS II Primary Effectiveness Endpoints
- VI. A. 2. PS II Secondary Effectiveness Endpoints
- VI. A. 3. PS II Inclusion and Exclusion Criteria
- VI. A. 4. PS II Procedure

VI. B. PS II Results: Overview and Primary Endpoints
- VI. B. 1. PS II Safety Results
- VI. B. 2. PS II Primary Effectiveness Endpoints

VI. C. PS II Results: Additional Panel Review Issues
- VI. C. 1. PS II Analysis by Pap Stratum
- VI. C. 2. PS II Changes to Primary Endpoints
- VI. C. 3. PS II Study Design: Biopsy Requirements
- VI. C. 4. PS II False Negatives
- VI. C. 5. PS II Relative Increase in TP (or Sensitivity) and FP
- VI. C. 6. PS II Absolute Increase in Sensitivity

VI. D. PSII Results: Other Issues
- VI. D. 1. PS II Reader Variability
- VI. D. 2. PS II Site Variability
- VI. D. 3. PS II Study Design: Performance Bias
- VI. D. 4. PS II Study Design: Subjects excluded due to “incomplete pathology”
- VI. D. 5. PS II Early Stopping

## I. Proposed Indication for use

The proposed indication for use for the LUMA™ Cervical Imaging System is for use as an adjunct to colposcopy for the identification of high-grade disease (CIN 2/3+) in patients referred to colposcopy with an ASCUS or LSIL cervical cytology result. LUMA is not intended to replace colposcopy. A thorough colposcopic evaluation with an identification or selection of biopsy sites must be performed independently and prior to the viewing of the LUMA results.

## II. Device Description

The MediSpectra LUMA Cervical Imaging System is a stand-alone, non-contact optical analysis system. Earlier the sponsor called this device Optical Detection System (ODS). "**ODS**" is used in the original PMA and in the Sponsor's Panel Package to refer to the LUMA device and to refer LUMA study results.

### II. A. Device Components

The LUMA system is a self-contained, mobile unit comprising the following components:

- Console and accessories
- Illumination probe
- Disposable probe cover.

*Console and Accessories*

The console contains the computer, control electronics, ultraviolet laser, broadband flashlamp assembly, camera controller, spectrometer, and the visible "targeting" laser used for centering and focusing. Interaction with the system is by way of a keyboard, touchpad and foot-actuated dual-pedals.

Output data are provided to the user by way of an LCD monitor and a printer. The LCD monitor is mounted on the system mast, and is used to display the software control screens and cervical images captured during the scan. A color printer on top of the console, produces a report containing the cervical image captured during the scanning process, both with and without the results display.

A calibration port is located in the front center of the console, and is used to calibrate the system prior to each patient exam; it can also be used to store the probe when not in use. An internal video camera reads the bar code on the disposable probe cover upon insertion into the calibration port; it ensures single-patient use and use of a valid disposable.

*Illumination Probe*

The illumination probe is attached to the console by way of an articulating arm. The articulating arm enables the user to manipulate and position the illumination probe during the exam. The probe can be moved in several different axes for optimum positioning, and contains an assembly that allows for fine focusing. The articulated arm joints and the probe joints can be tightened for user preference.

*Disposable Probe Cover*

The disposable probe cover is a single-patient use accessory that minimizes the risk of contamination to the system. It consists of ABS plastic and acrylic with an anti-fog coated optical window. A protective plastic strip over the anti-fog window is removed after the disposable cover is placed on the probe, but prior to inserting into the calibration port. Each disposable cover has a unique bar code label that is used to verify single-patient use. The probe with the valid disposable probe cover is then automatically calibrated prior to use on a patient.

II. B. Device Input Data

Diagnostic scans of the cervix are performed by interrogating the ectocervix using laser-induced fluorescence spectroscopy, white light diffuse reflectance spectroscopy, and video imaging. The cervix is scanned by the illumination probe, which is positioned near the proximal opening of the speculum as part of a colposcopy exam, following the application of acetic acid. The scan is a hands-free operation that takes approximately 12 seconds.

The fluorescence measurements are made by illuminating the cervix with ultraviolet (UV) light pulses at 337 nm from a nitrogen laser and recording the spectrally resolved intensities (spectra) of longer wavelength UV and visible light (360-720 nm) emitted from the cervix. Reflectance measurements are made by illuminating the cervix with pulses of broadband UV and visible light (360-720nm) from xenon flash lamps and recording the spectra of reflected light at those same wavelengths. Fluorescence and reflectance spectra are obtained from 499 distinct, closely-packed sites on the cervix via a scanning mechanism.

Prior to and during the measurement sequence, the cervix is also illuminated with flashlamps for capturing video images of the cervix. The video images prior to the measurements provide a method to align the device, a method which requires the cervix to be further illuminated with four bright spots of narrowband continuous wave visible light at 532 nm from a frequency doubled Nd:YAG laser. The video images are used for alignment, to control for movement during scan, and to create masks (more on the masks below). An interrogation point is rescanned if there is movement greater than 0.55 mm. An entire rescan is performed if accumulated movement is greater than 2.5 mm.

For each interrogation point (IP), the three measured spectra, one Fluorescence (FL) and two Broadband (BB), undergo some preprocessing: background subtraction, factory/reference/energy calibration, replication at top and bottom of spectrum, a 5 pixel median filter, a [redacted]polynomial regression[redacted] filter, and interpolation to 1 nm samples. The two BB spectra are then averaged constrained to spectral arbitration. Spectral arbitration can throw out one or both BB spectra based on glare, shadows, or low signal. If both BB spectra are thrown out, the IP is marked as indeterminate. The resulting FL and averaged BB spectra are then used to produce masks[redacted].

II. C. Display Output

Following the approximately 12-second optical scan of the cervix, the device analyzes the calibrated spectra and images of the cervix via a mathematical algorithm encoded in

its software. Output results are calculated and displayed within approximately 30 seconds. Results are displayed as a false-color overlay superimposed on a digital image of the cervix obtained during the scan. This overlay is color-coded to indicate the probability of identifying high-grade disease (CIN 2/3+) at different locations on the cervix follows:

- blue: highest probability of being CIN 2/3+
- yellow: lower probability of being CIN 2/3+
- green square cross-hatching: necrotic tissue(rare)
- gray diamond-shaped cross-hatching: indeterminate, i.e. an optical report is not being made.

II. D. Algorithm

The algorithm is made up of masks and two classifiers. The first classifier yields a score of no evidence of disease (NED). The second classifier yields a CIN 2/3 score (note: The sponsor uses the terms “probabilities” and “likelihoods” instead of score. In this document the term “scores” is used, because the relationship to probabilities is not consistent with probability theory). The masks are themselves classifiers that result in a positive or negative (zero or one) indication for a particular result: necrotic tissue (NEC), region of interest, vaginal wall, speculum blades, smoke tubes, fluid or foam, cervical opening (os), blood, glare, and mucus.

The output of the algorithm is depicted by a false-color image overlaid on a color video image of the cervix. This false-color overlay depicts points that are either indeterminate (IND, gray crosshatch pattern), necrotic (NEC, green trellis pattern), have no evidence of disease (NED, no marks), or have some “probability” score for the presence of CIN 2/3. The CIN 2/3 scores are colored yellow to blue: scores less than 0.2 are invisible (and thus equivalent to a NED result), from 0.2 to 1.0 the color map goes from yellow to blue and the transparency goes from 0% to 100%.

*Masks*

There are three kinds of masks: hard, soft, and NEC.

Hard masks lead to indeterminate (gray crosshatch) masks and are based on either spectral data or video image data. The hard masks determined by the spectral data identify mucus/fluids and the cervical edge/vaginal wall. The hard masks determined by the video data identify the region of interest, vaginal wall, speculum blades, smoke tubes, and fluid/foam.

Soft masks also lead to indeterminate marks, but are based on the video data only. Soft masks are intended to identify biologically benign areas that the spectral data cannot evaluate, such as the cervical opening (os), blood, glare, mucus. [redacted]

The NEC mask leads to the NEC (green trellis) marks [redacted].

The specific structure of each mask is complicated. In general, features are compared to thresholds and the results are combined with logical ANDs and ORs. Features from spectral images are absolute and relative intensities at specific wavelengths, as well as mean and maximum intensities across wavelengths. Features from video images are derived from pixel histograms (peaks, valleys, percentiles) and much image processing: color representation, spatial filtering, segmentation, segmentation smoothing (dilation, erosion), histogram processing, and component labeling.

*Classifier*

The BB spectrum is used to evaluate a score representing the likelihood, or probability of one of five classes: normal columnar, normal squamous, CIN 1, CIN 2/3, and metaplasia. This is done in two ways [redacted]. The [redacted] likelihoods are then combined; specifically, the likelihood for each class is multiplied together and normalized to produce a “probability” for each class. [redacted]

The [redacted] methods are similar in that each selects a small number of features that are linear combinations of the [redacted] spectrum. Each organizes the features into feature vectors. Then for each feature vector the mean vector, covariance matrix, and prior probability for each tissue class are estimated from training data. These parameters are then used to design (Gaussian) likelihood functions for unclassified BB spectra. So for an unknown input [redacted] spectrum, the outputs of each method is a set of five likelihoods: the likelihood that the input [redacted] spectrum is normal columnar, normal squamous, CIN 1, CIN 2/3, and metaplasia.

[redacted]

[redacted]

## III. Pre-clinical Studies

### III. A. Pre-clinical Testing

For the PMA application, a complete battery of pre-clinical testing results is provided for the LUMA system. This includes engineering testing, optical exposure, software testing, failure mode and effect analysis, hazard analysis, electrical, thermal and mechanical testing, electromagnetic compatibility testing, biocompatibility testing, disinfection and shelf life for the disposable cover.

The proposed Summary of Safety and Effectiveness Data (Sponsor's Panel Package, Volume 1, Tab 2) includes the summaries of the pre-clinical test data provided in the PMA.

**Bench/Engineering Testing**

The LUMA Cervical Imaging System and subassembly verification testing was conducted to demonstrate that the system fulfills the requirements identified in the product specification, risk management guidelines and associated failure modes and effects analyses, and the design requirements document.

**Optical Exposure**

A complete optical radiation assessment was performed. The worst-case maximum laser single pulse exposure is 35 µJ (mean+6SD) corresponding to a maximum radiant dose of 0.36 $mJ/cm^2$ when using a 3.5 mm limiting aperture as recommended by the American Conference of Governmental Industrial Hygienists (ACGIH). This value is more than a factor of 10 below the ACGIH Threshold Limit Value (TLV) for exposure to a single pulse. The worst-case radiant power from the xenon broadband lamps is 4.6 $mW/cm^2$, which is more than a factor of 40 below the ACGIH Threshold Limit Value. An evaluation of the total effective dose between 300 and 400 nm from all sources was also evaluated, including a worst-case 10-minute exposure from the xenon source optics, a 10-minute exposure to the colposcope, and a worst-case 10-scan exposure from the UV laser. The maximum optical emission was found to be more than a factor of 10 below the ACGIH Threshold Limit Value.

**Electrical, Thermal and Mechanical Testing**

The LUMA console was evaluated for electric shock, fire and mechanical safety including, for example leakage current tests, rigidity and rough handling tests, and a normal temperature test. The LUMA system was found to be compliant with the following standards: UL 60601-1, IEC 60601-1, and CAN/CSA-C22.2 No. 601.1-M90.

**EMC Testing**

The LUMA system was evaluated for electromagnetic compatibility including emissions and immunity tests performed by an independent laboratory. The LUMA system was found to be compliant with IEC 60601-1-2.

**Environmental Stress and Lifetime Testing**

Packaging and shipping containers for the console and the disposable covers were evaluated to ensure that they can withstand the expected distribution environment. Ship testing was successfully conducted on the LUMA single-use disposable containers, as well as on the console-shipping container. Testing was performed to the standard ASTM D4169-01.

**Biocompatibility**

The LUMA disposable probe cover is the only component of the LUMA system that has potential to contact the patient in normal use. It is made from ABS plastic and acrylic. The disposable probe cover is used to cover the front of the probe and is positioned near the proximal opening of the speculum during cervical scanning. Testing per the ISO 10993 matrix for an externally communicating, skin contact, short duration device was performed using good laboratory practices. Testing was done as per ISO 10993-1 guidance as required for this type of contact.

**Disinfection**

Risk analysis determined that a user could accidentally insert the probe with a used disposable cover into the calibration port of the console. To mitigate this risk, (1) instructions are provided both in the User's Guide and on the system display to discard the disposable cover after use and (2) the LUMA system was designed with an electronic eye that checks the disposable bar code; if the bar code has been used before, the software, will prompt the user to disinfect the calibration port door. The calibration port has been designed such that, if a used disposable cover is inserted into the port, only the calibration port door could become contaminated. Instructions for disinfection are provided in the User's Guide.

**Software Testing**

System level software quality assurance (SQA) testing, unit level testing and integration testing were performed on the LUMA and manufacturing software. User test cases were developed for SQA testing to verify all LUMA software. Defects affecting the algorithm were corrected between Pivotal Studies I and II and are discussed in Section III.B.2 (FDA Summary).

**Failure Mode and Effects Analysis (FMEA) and Hazard Analysis**

Risk Management was initiated in the Definition and Requirement phases of product development and conducted throughout the design and development activities for the LUMA system. The purpose of the Risk Management activities was to identify and control potential hardware and software hazards associated with the LUMA System. Risk Management involved identifying potential hazards, estimating and evaluating the associated risks, and reducing these risks to acceptable levels. The criteria for risk acceptability were based on ISO 14971.

III. B. Pre-clinical Review Issues

III. B. 1. Optical Radiation and Safety Review Issues

*Possible interactions between UV light and HPV*

Human papilloma virus (HPV) is known to be a precursor to cervical dysplasia. To our best knowledge, no data are available on activation of HPV by UV light. Other viruses (HSV, HIV) require much higher UV dosages to be activated than the dosage from the LUMA device [Norval et al, Photochem Photobiol, 64(2):242-5, 1996; Zmudzka et al, Photochem Photobiol, 64(2):246-53, 1996; Beer et al, Photochem Photobiol, 59(6):643-9, 1994]. Pending further studies in this area, we believe that there is minimal risk of this device to activate HPV.

There are a number of published studies which suggest that HPV-infected cells are more sensitive to UV exposure than non-infected cells [El-Mahdy et al, Mutat Res, 459(2):135-45, 2000; McKay et al, Neoplasia, 1(3):276-84, 1999; Giampieri et al, Br J Cancer, 90(11):2203-9, 2004]. Most of the research in this area reports preferential damage and/or changes to HPV-infected cells due to higher energy UV light than what the LUMA device uses. This topic is still under review by FDA and the sponsor.

*Potential interference from medication*

*S*ome topical, oral or intravenous medications used by the patient at the time of a LUMA procedure could introduce safety concerns and/or interfere with LUMA's optical measurements. The sponsor addressed this issue, during the pivotal studies by excluding individuals who had used such medications, e.g. sulfa, amplicillin and tetracycline. The sponsor also indicated this exclusion in the User's Guide (Sponsor's Panel Package, Volume 1, Tab 3). Although it has been documented (Pfefer et al, Photochem Photobiol, 73(6), 2001) that certain drugs used for photodynamic therapy (e.g. porfimer sodium) can remain in the human body for months and significantly influence some types of fluorescence measurements, there is no evidence to suggest that these particular medications would affect the LUMA results, nor compromise patient safety.

*Patients with cervicitis*

Published studies [Ramanujam et al, Photochem Photobiol, 64(4):720-735, 1996; Heintzelman et al, Photochem Photobiol, 71(3):327-332, 2000] have suggested that inflammation can affect the performance of optical methods to detect dysplasia which are similar to that used by the LUMA device. In light of these studies, FDA requested the sponsor to review histology and analyze data from the pilot and pivotal clinical studies to determine the potential impact of cervicitis on the effectiveness of the LUMA device. Although the data were limited, there was no evidence that the effectiveness of the LUMA device was influenced by cervicitis any differently than the effectiveness of colposcopy.

*Fluorescence calibration*

The calibration methods incorporated into the LUMA device are rigorous and necessary to ensure reliable data are input into the algorithm. However, the fluorescence

calibration lacks a pre-patient calibration method to correct the shape of fluorescence spectra for variations in the transmittance of the optical collection path. These variations could be introduced by degradation of fixed optical components in the probe and console, and by differences between the windows of the disposable probe covers. In light of this concern, the sponsor has demonstrated that the hardware and algorithm of the device are robust enough to make the impact of the expected variations negligible.

III. B. 2. Algorithm Review Issues

The sponsor has attempted to present a clear physical and physiological basis for the diagnostic ability of their device. However, the bulk of the algorithm is a "black box." Only the spectral mask which identifies necrotic tissue (Sponsor's Panel Package, Volume 1, Tab 4, Section 4.4.3.4) appears to purposefully include features which are pertinent to the physiology (FAD/NADH fluorescence ratio, protoporphyrin-IX fluorescence). In this case, FDA must rely almost exclusively on statistical analyses of diagnostic efficacy. We cannot be certain of how much of the algorithm's ability to discriminate CIN 2/3+ from other tissue types is actually due to neoplastic processes instead of unrelated, coincidental processes.

There are three major concerns that were noted during review of the algorithm development:

1. The algorithm was changed between Pivotal Studies I and II (PS I and PS II).

2. The results from PS I and PS II do not account for reader variability.

3. Sensitivity is estimated and presented throughout PMA without any indication regarding the uncertainty of the estimate. The pivotal studies were not designed to determine sensitivity.

The algorithm was changed between PS I and PS II. Below we summarize the changes, how they were evaluated, and our conclusions. The issues of reader variability (FDA Summary, Sections V.D.1 and VI.D.1) and sensitivity (FDA Summary, Section V.C.5, V.C.6, VI.C.5, VI.C.6) will be addressed in elsewhere in this executive summary.

Algorithm defects were discovered as part of ongoing research into next generation products and from investigations into potential causes of reported events from the PS I clinical trials. Clinical study raw data and images were used to identify these errors and to validate the fixes, but clinical outcomes were not. The defects were coding errors. The fixes affected only the NED (no evidence of disease) mask, the os mask, and the cervical edge mask, but did not affect the CIN 2/3 disease scoring module. Consequently, when the fixes impacted the LUMA display, they did so in essentially one of two ways: either a CIN 2/3 interrogation point was changed to NED or IND (indeterminate), or a NED/IND interrogation point was changed to CIN 2/3. There were no instances of a CIN 2/3 value changing to some other measurable CIN 2/3 value (in a very few instances, NEC interrogation points changed to something else).

The algorithm changes were evaluated quantitatively and clinically. Two basic quantitative evaluations included accounting of the number of interrogation points and patients from PS I that had different algorithm outputs and LUMA display. Other quantitative evaluations compared the ROC curves of the PS I algorithm scores and the

PS II algorithm scores given the Pilot Study biopsy data; no clinician was involved. Clinical evaluation included the review of images determined to have a higher impact on clinical decisions in addition to an Algorithm Equivalence Study.

Our review of the algorithm change and the testing concludes the following:

- The changes corrected coding errors rather than being related to algorithm design and/or performance results.

- On a per interrogation basis, the changes affected a small number of total interrogation points.

- On a per patient basis, while an appreciable number of patients had interrogation points that changed, the potential impact seems to be low for most of the patients, and when the potential impact is high, the algorithm change favors more biopsies (impact for each patient equals the number of changed points divided by the number of diseased points).

- Based on a review of 56 images from PS I on which the change to the current algorithm was scored as having the highest impact, the current algorithm would not be expected to lower the TP rate in the LUMA+colpo arm of PS I. This assessment was made by seeing which of the subjects in the LUMA + colpo arm might have had fewer biopsies using the current algorithm, and seeing that none of these subjects was a TP. However, the current algorithm might have increased the TP rate in the LUMA+ colpo arm of PS I. Therefore, if the current algorithm had been used in PS I, the increase in the LUMA + Colpo arm TP rate over that of colpo only arm might have been larger (better) by a small amount.

- On a per patient basis, the algorithm equivalence study shows intra-reader agreement for images with no changes is very similar to intra-reader agreement for images with changes (agreement equals the proportion of biopsies that have matching biopsies from other algorithm, where a matching biopsy is within 3mm).

In summary, it appears that the algorithm change has little impact on the study results for PS I. Both PS I and PS II can be considered in this submission because of these conclusions and the following: results from the two studies are not combined (data was collected under different protocols, and the current indication for use explicitly states that a thorough colposcopy exam and biopsy selection is made independently of and prior to viewing the LUMA display.

## IV. Overview of Pivotal Studies I and II (PS I and PS II)

MediSpectra has conducted two pivotal clinical studies to evaluate the safety and efficacy of the LUMA System, Pivotal Study I (PS I) and Pivotal Study II (PS II). These studies did not require Investigational Device Exemption (IDE) regulatory consent from FDA, because for the studies as designed, the LUMA System was considered a “non-significant risk” (NSR) device.

Both PS I and PS II were prospective, multi-center clinical trials enrolling women referred for colposcopy following a recent history of the following abnormal Pap smear results:

first-time ASC-US/-H, repeat ASC-US/-H, LSIL, HSIL or squamous cell cervical cancer. The two studies differ in how LUMA is used as an adjunct to colposcopy, and the study designs have different limitations. In Tab 2 of the Sponsor's Panel Package (Volume 1, page 2-014), a diagram of the designs of the two studies is presented.

**PS I** was a 1:1 randomized comparison of the rate of diagnosis of CIN 2/3+ in the following two study arms (FDA Summary, Section V; Sponsor's Panel Pack, Volume 1, Tab 5):

- **colpo-only** arm: Women underwent traditional colposcopy followed by cervical biopsy.
- **LUMA+colpo** arm: Women underwent *simultaneous* colposcopy and LUMA cervical imaging. The decision to biopsy was based on the simultaneous review of the LUMA and colposcopy images.

In PS I, a total of 2299 subjects were enrolled (Intent to Diagnose population) and 2186 completed the study without being excluded (the "Per Protocol" population). Thirteen sites and 51 colposcopists participated in the study.

**PS II** was the subject of an agreement letter with FDA. It was a single arm, internally-controlled study (FDA Summary, Section VI; Sponsor's Panel Pack, Volume 1, Tab 6).

- **all patients**: Biopsy sites were first identified via colposcopy, and then additional sites were identified using LUMA.

In PS II, the protocol called for the enrollment of 788 subjects. The sponsor terminated the study after enrolling 227 subjects at 7 institutions.

Because of the difference in the designs of PS I and PS II, the results of these studies cannot be combined. A summary of these differences is provided in the table below.

**Comparison of PS I and PS II**

| | **PS I** | **PS II** |
|---|---|---|
| **Study Design** | randomized controlled trial | one-armed, paired design |
| **Sample Size** | 2299 | 227 (788 planned, stopped early) |
| **Algorithm Version** | 2.0.05-09 | 2.0.10-11 |
| **Treatment Group** | simultaneous LUMA + colpo evaluation | colpo followed by LUMA increment |
| **Primary Endpoints** | TP rate increase > 0%<br>FP rate increase < 8% | TP rate increment > 2%<br>FP rate increment < 15%<br>biopsy-level PPV (no hypothesis) |

Elsewhere in the panel package, the PS I is sometimes called an ***intended-use study*** and PS II a ***device-performance study***. However, the design of PS II more closely matches the proposed indication for the device.

As discussed in Section III.B.2 (FDA Summary), there were changes made to the LUMA's algorithm after the completion of PS I and the beginning of PS II.

The primary effectiveness endpoints in both PS I and PS II pertain to the incremental changes due to LUMA in the subject "true positive" (TP) and "false positive" (FP) rates for colposcopy. Biopsies are either taken or not taken according to the study protocols.

Individual patient outcome is described as follows:

- **TP:** patient had at least one CIN2/3 positive biopsy
- **FP:** patient had one or more biopsies and they were all negative
- **negative:** patient did not have biopsy taken.

The TP and FP rates are defined as follows:

- **TP rate**: the number of TP patients divided by the total number of patients[1]
- **FP rate**: the number of FP patients divided by the total number of patients[2]

The TP and FP rates defined above should not be confused with the traditional definitions. Traditionally, the TP rate divides the number of TP patients by the number of *diseased patients* and the FP rate divides the number of FP patients by the number of *non-diseased patients*. Traditionally, the TP rate is called sensitivity and the FP rate equals one minus specificity. The design of the two clinical studies does not permit direct calculation of the sensitivity and specificity. This is because neither study established the true underlying disease status for each patient. Ideally, this would be achieved by sampling the entire transformation zone or, less optimally, by extended (2-3 year) follow up of subjects. Truth regarding disease status can be known only where biopsies are taken and are confirmed to be CIN 2/3+ by histology. If no biopsy specimens in a subject contain CIN 2/3+, the subject should not be presumed to be disease-free, because some nonbiopsied sites could contain disease.

Both studies had two primary endpoints: one for TP rate and one for FP rate. In addition, although sensitivity and specificity cannot be estimated from either study, relative sensitivity can be estimated in both studies (FDA Summary, Sections V.C.5 and VI.C.5), and absolute sensitivity can be imputed by assuming a value for the prevalence of CIN 2/3+ (FDA Summary, Sections V.C.6 and VI.C.6)

Note that true negative (TN) and false negative (FN) were poorly defined in this study, because subjects without positive biopsies cannot be assumed to be without CIN 2/3+ disease. The biopsies may have been in the wrong locations.

---

[1] This rate is more appropriately called the *detection rate (Pepe, 2004, Statistical Evaluation of Medical Tests for Classification and Prediction, Oxford).*

[2] This rate is more appropriately called the *false referral rate (Pepe, 2004, Statistical Evaluation of Medical Tests for Classification and Prediction, Oxford).*

## V. Pivotal Study I (PS I) (Sponsor's Panel Package, Volume 1, Tab 5)

### V. A. PS I Study Design

PS I was a prospective, randomized, controlled, multi-center clinical trial enrolling women referred for colposcopy who had an abnormal Pap smear in the last 6 months. The definition of abnormal was first-time ASC-US/-H, repeat ASC-US/-H, LSIL, HSIL or squamous cell cervical cancer. The study was a 1:1 randomized comparison of the rate of diagnosis of CIN 2/3+ between women who undergo traditional colposcopy followed by cervical biopsy against women who undergo *simultaneous* colposcopy and LUMA cervical imaging followed by cervical biopsy.

Randomization was stratified by Pap smear diagnosis and colposcopist. The stratified randomization was useful for ensuring that each colposcopist had a comparable number of colposcopic exams within each Pap smear diagnosis.

### V. A. 1. PS I Primary Effectiveness Endpoints

The two primary endpoints were subject-level true positive (TP) and false positive (FP) rates. The primary hypotheses were that, for the per protocol population,

- TP rate was greater in LUMA+colpo arm than in the colpo-only arm (one-sided p-value <0.025)
- FP rate was no more than 8% greater in LUMA+colpo arm than in the colpo-only arm (one-sided p-value <0.025)

### V. A. 2. PS I Secondary Effectiveness Endpoints

Secondary effectiveness endpoints were as follows:

- TP and FP rates of the two study arms within strata (i.e. Pap smear diagnosis category) at the subject level
- Number of TP and FP biopsies associated with subjects as well as biopsy-specific sites
- TN and FN rates associated with subject-level outcomes in negative biopsy subjects
- TP and FP rates associated with biopsy-level outcomes
- LUMA TN and FN rates associated with biopsy-level outcomes
- Subject-level agreement rate between the blinded LUMA scan and colposcopy findings in the colposcopy-only arm; subject-level agreement rate between the LUMA scan and colposcopy findings in the combination group

V. A. 3. PS I Inclusion and Exclusion Criteria

The inclusion criteria were as follows:

- Signed informed consent
- Age 18 and older
- Documented abnormal Pap smear defined as first-time or repeat ASC-US/-H, LSIL, HSIL or cancer or equivalent classification using Bethesda II
- Subjects referred for LEEP must be willing to undergo biopsy prior to LEEP

The exclusion criteria were as follows:

- Pregnant through six weeks post-partum
- Menstruating (heavy flow) at the time of the LUMA
- History of hysterectomy or congenital anatomical cervical variants, e.g. double cervix
- Cervical stenosis (unable to insert Q-tip or cytobrush into endocervical canal)
- History of DES exposure in utero
- Punch biopsy or ablative/excisional therapeutic procedure since Index Pap smear (e.g. cervical conization, LEEP, laser, cryosurgery, radiation treatment, etc)
- Use of vaginal creams or vaginal medications (including all douches except saline) within the 48-hours immediately prior to LUMA examination
- Current use of photosensitizing agents (within 72-hours) such as sulfa, ampicillin and tetracycline
- History of photosensitivity or other diseases affected by UV radiation (e.g. porphyria, Systemic Lupus Erythematosis, etc)
- Undergoing phototherapy
- Previously enrolled in current study or in the Pilot I study, Pap smear Study or Pivotal Verification Studies
- Participated in an investigational drug or device study within the past 30 days
- Subjects with observable and untreated gynecological infections, such as acute cervicitis, acute inflammation and acute vaginitis
- Subjects who required a Pap smear at the time of the LUMA examination
- Index Pap smear that is greater than 6-months old at time of study visit, or within prior 7 days

V. A. 4. PS I Procedure

Randomization occurred as soon as the subject ID number and Pap smear result were entered into the LUMA system. A speculum was inserted into the vagina and the LUMA probe was placed at the opening to the speculum in such a way as to center the cervix in the field of view. The cervix was then gently washed with saline. Acetic acid was liberally applied to the cervix and pooled fluid was removed from the vagina.

The LUMA system was focused and a scan was made over a 15-second interval. (Up to 5 additional scans could be taken if the previous scan was deemed inadequate, e.g. due to motion by the subject.) Next, the LUMA probe was moved aside and the colposcope positioned.

The procedure then varied depending on study arm.

- Colpo-only arm: The investigator was blinded to LUMA results. The investigator took colposcopy-indicated biopsies and records biopsy sites.

- LUMA + colposcopy arm: The LUMA system displayed a color image of the cervix with the LUMA color overlay superimposed over the image of the cervix. The LUMA overlay identified areas on the cervix that potentially had CIN 2/3. The investigator decided whether cervical biopsies were indicated, and took any LUMA-directed biopsies first. All biopsy sites were recorded as well as the indication for the biopsy (e.g. LUMA only, colposcopy only or both LUMA and colposcopy).

Details of the biopsy selection procedure are provided in the Sponsor's Panel Package (Volume 1, Tab 5, Sub-tab "Interpretation of ODS…"; Volume 2, Question 30).

All biopsy specimens were sent to an on-site pathologist who prepared two diagnostically equivalent slides from each specimen. The "study slide" was sent to the contract research organization (CRO) and from there to the Central Pathology Review for independent reading by two pathologists. The two reviewing pathologists were blinded to the source of the slides and provided with no specific patient information except patient age (Sponsor's Panel Package, Volume 1, Tab 5, page 5-044). If the two pathologists disagreed on the diagnosis, the slide was reviewed by a third pathologist.

V. B. PS I Results: Overview and Primary Endpoints

A total of 2299 subjects were enrolled (Intent to Diagnose population). 113 were discontinued, and 2186 completed the study without being excluded (the "Per Protocol" population, 1090 in LUMA+colpo; 1096 in colpo-only). Thirteen sites and 51 colposcopists participated in the study. The first enrollment took place on August 22, 2002. The last subject was evaluated on September 4, 2003.

Of the 113 (4.9%) discontinued subjects, 27 (1.2%) were discontinued by the investigator before randomization, and 86 (3.7%) were excluded after randomization. Of these 86, 49 were removed from the LUMA+colpo arm, and 37 were removed from the colpo-only arm. Among randomized subjects, the primary reasons for exclusion were labeled as "no majority Pathology diagnosis," i.e. every biopsy for a subject was either unreadable or had a disagreement among the reference pathologists, (41/2272 or 1.8%), and device malfunction identified after the LUMA scan (26/2272 or 1.1%). Within randomized subjects of the ASCUS/LSIL sub-population, 1.5% (27/1808) were excluded because of this "no majority Pathology diagnosis" designation.

The mean age of study participants was 30.4 (SD 11.1) years.

PS I racial demographics were as follows:

| | |
|---|---|
| Caucasian | 44.8% |
| African American | 20.8% |
| Hispanic | 29.3% |
| Asian | 2.3% |
| Native American | 1.1% |

The referral Pap smear study profile for PS I was:

| | |
|---|---|
| ASCUS | 36.3% |
| LSIL | 43.6% |
| HSIL/Cancer | 20.1% |

There were no statistically significant differences in any of these groups across the two arms of the study.

Some subjects did not undergo any biopsy, as shown in the following table (Sponsor's Panel Package, Volume 1, Tab 5, page 5-142):

**PS I Percent of Subjects with Any Biopsies by Arm and Strata**

| Stratum | Colpo-Only Arm | LUMA+Colpo Arm |
|---|---|---|
| **Overall** | 77.3% (847/1096) | 82.3% (897/1090) |
| **ASCUS/LSIL** | 72.6% (632/871) | 79.6% (697/876) |
| **HSIL/Cancer** | 95.6% (215/225) | 93.5% (200/214) |

The average number of biopsies obtained increased in the LUMA plus colposcopy arm as shown in the table below.

**PS I Mean Number of Biopsies by Arm and Strata within Per Protocol Population (95% confidence interval)**

| Strata | Colpo-Only Arm | LUMA+Colpo Arm |
|---|---|---|
| **Overall** | 1.03 (0.97, 1.09)* | 1.30 (1.23, 1.37) |
| **ASCUS/LSIL** | 0.93 (0.87, 1.00) | 1.24 (1.16, 1.31) |
| **HSIL** | 1.40 (1.25, 1.55) | 1.57 (1.40, 1.74) |

**95% Confidence Intervals calculated by FDA based on the Poisson distribution*

V. B. 1. PS I Safety Results

There were 7 reports of adverse events in PS I reported in the PMA. Four of these events occurred in the same patient. These events consisted of cramping, vomiting, weakness, vaginal bleeding and fainting. None of the events was serious, and none was believed to be related to the LUMA device.

V. B. 2. PS I Primary Effectiveness Endpoints

The results for the two primary effectiveness endpoints are provided in the table below.

- The TP primary endpoint was not met, i.e. for the TP rate difference, the lower bound of the 95% confidence interval was less than 0.
- The FP primary endpoint was met, i.e. the upper bound on the 95% confidence interval for the FP rate difference was less than 8%.

**PS I: Primary Effectiveness Endpoints (All Subjects)**

| | Colpo-only % (n/N) | LUMA+Colpo % (n/N) | Difference (95% CI) | Met Hypothesis? |
|---|---|---|---|---|
| **TP** | 19.9% (218/1096) | 21.8% (238/1090) | 1.9% (-1.5%, 5.3%) | **No**: 95% CI includes 0. |
| **FP** | 57.4% (629/1096) | 60.5% (659/1090) | **3.1% (-1.0%, 7.2%)** | **Yes**: 95% CI <8% |

V. C. PS I, Additional Review Issues

V. C. 1. PS I Analysis by Pap Stratum

The sponsor is presently seeking an indication specifically for the ASCUS/LSIL sub-population. The TP and FP results are provided by entry Pap result in the tables below. Although analysis by individual Pap strata was pre-specified as a secondary endpoint, analysis of ASCUS/LSIL as a group was not.

**PS I Subject-level True Positive Rates by Entry Pap Result**
*Original Hypothesis: 95% confidence (two-sided) that the TP rate in the LUMA+colpo arm is more than in the colpo-only arm*

| Pap stratum (% total) | Colpo-only % (n/N) | LUMA+Colpo % (n/N) | Difference (95% CI) |
|---|---|---|---|
| **Overall (100.0%)** | 19.9% (218/1096) | 21.8% (238/1090) | 1.9% (-1.5%, 5.3%) |
| **ASCUS/LSIL (79.9%)** | 11.4% (99/871) | 14.4% (126/876) | +3.0% (-0.1%, 6.1%) |
| **ASCUS (36.3%)** | 10.9% (42/387) | 13.8% (56/407) | +2.9% (-1.7%, 7.5%) |
| **LSIL (43.6%)** | 11.8% (57/484) | 14.9% (70/469) | +3.1% (-1.2%, 7.4%) |
| **HSIL (20.1%)** | 52.9% (119/225) | 52.3% (112/214) | -0.6% (-9.9%, 8.7%) |

This original hypothesis for the TP primary endpoint was not met for the overall study population or for any of the sub-populations, i.e. for the TP rate difference, the lower bound of the 95% confidence interval was less than 0 for the overall population and for each of the Pap strata. For the ASCUS/LSIL sub-population, the lower bound on the 95% confidence interval was -0.1%. The p-value for the difference was 0.035, just above the significance level of 0.025 (one-sided).

**PS I Subject-level False Positive Rates by Entry Pap Result**
*Original Hypothesis: 95% confidence (two-sided) that the FP rate in the LUMA+colpo arm is **not** more than 8% greater than in the colpo-only arm*

| Pap Strata (% Total) | Colpo-Only % (n/N) | LUMA+Colpo % (n/N) | Difference (95% CI) |
|---|---|---|---|
| **Overall (100.0 %)** | 57.4% (629/1096) | 60.5% (659/1090) | **3.1% (-1.0%, 7.2%)** |
| **ASCUS/LSIL (79.9 %)** | 61.2% (533/871) | 65.2% (571/876) | +4.0% (-0.5%, 8.5%) |
| **ASCUS (36.3 %)** | 54.3% (210/387) | 62.4% (254/407) | +8.1% (1.3%, 14.9%) |
| **LSIL (43.6 %)** | 66.7% (323/484) | 67.6% (317/469) | **+0.9% (-5.1%, 6.9%)** |
| **HSIL (20.1 %)** | 42.7% (96/225) | 41.1% (88/214) | **-1.6% (-10.8%, 7.6%)** |

The original hypothesis for the FP primary endpoint was met for the overall study population, and for the following sub-populations: LSIL and HSIL. That is, the upper bound on the 95% confidence interval for the difference between the LUMA and Colposcopy arm and the colposcopy-only arm was less than 8% in these groups. For both the ASCUS/LSIL and in the ASCUS populations, the upper bound on the 95% confidence limit was greater than 8%. This means that the original hypothesis for FP rate was not met for these groups.

In general, when a study fails its primary endpoint it is not statistically valid to analyze subgroups. If this study had been designed for multiple sub-group analysis, then the sub-group analysis would have to be corrected to control the overall chance of finding a falsely significant result in any of the subgroups. When considering the three PAP strata LSIL, ASCUS, and HSIL, the number of possible subgroups is 7. When considering the four PAP strata of LSIL, first time ASCUS, repeat ASCUS, and HSIL, which were used for stratified randomization of the study, the number of possible subgroups is 15. In principle, the findings in the ASCUS/LSIL subgroup should be discounted and/or adjusted to account for multiple possible subgroups. For example, if for a test of a single group a significant result is obtained with a p-value of <0.05, then with 7 possible subgroups, one statistically valid approach is to require a p-value of <0.0071 (0.05/7).

The TP rate was not met for ASCUS/LSIL at significance level 0.05, so the significance level 0.0071 would not be met either.

V. C. 2. PS I Joint Analysis

The purpose of this analysis was to help understand if an overall difference existed between the arms in the set of three subject-level rates: the TP rate, the FP rate, and the negative rate, i.e., the rate of non-biopsied subjects. Per protocol, only the TP and FP rates were primary endpoints and these were analyzed separately. In the joint analysis, the TP and FP rates were analyzed jointly together with the negative rate (Sponsor's Panel Package, Volume 2, Question 4). The joint analysis takes advantage of correlation among the three rates that exists because they must sum to one.

One joint analysis was based on Fisher's exact test. This test was significant in the overall population (p value 0.0132), indicating that the TP, FP and negative rates were not independent of arm. Fisher's exact test on the overall population was then used in a gatekeeper procedure for subgroup analysis. Because the overall test with this analysis (the gatekeeper) was significant, a test of the ASCUS/LSIL sub-population was permitted at significance level 0.05/7 = 0.0071 to control the overall false positive error rate in testing a possible seven subgroups that can be formed from the three PAP strata (LSIL, ASCUS, HSIL). For the ASCUS/LSIL subgroup, the Fisher's exact test p value was 0.0014, which is less than 0.0071 and therefore indicates a significant difference between arms in the TP, FP, and negative rates.

The joint analysis yielded a significant finding both overall and for the ASCUS/LSIL sub-population, but these results have limitations. First, the analysis was not pre-planned, complicating its interpretation. Second, the significant results obtained are not directly related to the primary endpoints (increase in TP rate, and no more than 8% increase in FP rate). Fisher's exact test only indicates that the TP, FP, and negative rates are collectively not the same in the two arms, not which rates are individually different. The significant finding overall does not appear to be due to any one of the three rates (Table below). The significant finding in ASCUS/LSIL subjects seems to be especially influenced by a lower rate of subjects not biopsied in the LUMA+colpo arm relative to the colpo only arm.

**Percent of TP, FP and negative subjects by arm**

| | Overall | | ASCUS/LSIL | |
|---|---|---|---|---|
| | **colpo** | **LUMA+colpo** | **colpo** | **LUMA+colpo** |
| TP | 19.9% | 21.8% | 11.4% | 14.4% |
| FP | 57.4% | 60.5% | 61.2% | 65.2% |
| negative | 22.7% | 17.7% | 27.4% | 20.4% |

The sponsor has also provided a Grizzle-Starmer-Koch (GSK) joint analysis. Aspects of this analysis remain under review between FDA and the sponsor.

V. C. 3. PS I Biopsy-Level PPV

A main purpose of this analysis is to help understand if LUMA indicated biopsies are better than random biopsies. For each biopsy in the LUMA+colpo arm, the colposcopist designated whether the biopsy was indicated by colpo only, LUMA only, or both colpo and LUMA. Because there is no record of colposcopic biopsy sites for PS I, one can only infer how the colposcopist made these designations.

**Biopsy-Level PPV (%) Results**

| | LUMA+Colpo Arm | | | | Colpo |
|---|---|---|---|---|---|
| | By Indication | | | | only Arm |
| Population | colpo only | LUMA only | both | Over-all | |
| All subjects | **17.5** | 11.3 | **28.8** | 21.9 | 24.0 |
| ASCUS/LSIL | **10.8** | 8.6 | **18.1** | 13.7 | 14.1 |
| HSIL/Cancer | 41.5 | 32.3 | 54.4 | 48.6 | 49.8 |

According to analysis of the biopsies taken in the LUMA+colpo arm, LUMA appears to be selecting biopsy sites that are predictive of CIN 2/3+. For each of these three biopsy reasons (colpo only, LUMA only, or both colpo and LUMA), the positive predictive value (PPV) was computed, i.e., the proportion of biopsies confirmed to be CIN 2/3+ by pathology (Table above). For each Pap stratum, PPV was less for biopsies indicated by colpo-only than biopsies indicated by both colpo and LUMA. The overall PPV was 17.5% for colpo-only-indicated biopsies and 28.8% for both LUMA and colpo indicated biopsies, a statistically significant difference (p value 0.0005). In the ASCUS/LSIL subgroup, the PPVs were 10.8% and 18.1%, again statistically significant (p-value 0.0018). (These p-values are given in Volume 2, Question 20 of the Sponsor's Panel Package and were computed accounting for correlation among biopsies from the same subject.) If LUMA biopsies were no better than random, these two PPVs would be expected to be the same. The interpretation of these significant findings is that among the pool of colpo-indicated biopsies (the "colpo only" and "both" columns of the table above), the LUMA selections for biopsy are better than random selections for biopsy.

The PPV for LUMA-only indicated biopsies (11.3%) was smaller than the PPV for colpo-only indicated biopsies (17.5%). However, for ASCUS/LSIL subjects, LUMA-only indicated biopsies appeared to be better than the random biopsies taken in the Shanxi Province Cervical Cancer Screening Study II (Pretorius et al, Am J Obstet Gynecol, 2004, 191:430-434). For LSIL and ASCUS subjects, respectively, PPVs were 1.7% and 3.6% for random biopsies in that study compared with 9.6% and 7.8% among LUMA-only indicated biopsies in PS I. These results suggest that among LUMA-only locations indicated for biopsy, the ones that were selected for biopsy by the colposcopist (in addition to his/her colpo-indicated biopsies) are much better than random selections for biopsy.

V. C. 4. PS I False Negatives

Although the primary endpoints were based on *subject-level* results, we can further explore the performance of colposcopy and LUMA via a *biopsy-level* analysis of the 126 TP subjects in the ASCUS/LSIL subpopulation of the LUMA+colpo arm. The study Case Report Forms asked the investigator to indicate, for each biopsy, whether the biopsy was directed by colposcopy, LUMA, or both. They found the following sources for the CIN 2/3+ defining biopsies:

- 85/126 (67.5%): both the LUMA and the colposcopic impressions
- 26/126 (20.6%): colposcopic appearance alone
- 15/126 (11.9%): LUMA overlay.

Among the 126 TPs in the ASCUS/LSIL sub-population of the LUMA+colpo arm of PS I,

- 20.6% (26/126) would not have been diagnosed if LUMA were the only tool being used to identify biopsy sites
- 11.9% (15/126) of the 126 TP patients would not have been diagnosed if colposcopy were the only tool being used.

The sponsor reviewed all of the cases in both PS I and PS II where there was discordance between colposcopy and LUMA such that the LUMA did not target a site for biopsy. Because there is no record of colposcopic biopsy sites for PS I, one can only infer how the colposcopist designated colpo only, LUMA only or both. The sponsor's clinical expert concluded that the FNs were due to obscuring features on the cervix,

incomplete visualization of the squamo-columnar junction, and no obvious explanation all in approximately equal thirds (Sponsor's Panel Package, Volume 2, Question 24).

The false negative rate is traditionally defined as the proportion of diseased subjects that test negative. Under this definition and when colposcopy is used as the gold standard to determine true disease status, the false negative rate in ASCUS/LSIL subjects was 23.4% (26/111), the percent of CIN 2,3+ subjects detected by colposcopy not detected by LUMA. Conversely, LUMA sensitivity can be estimated to be 76.6% when colposcopy is used as the gold standard. However, because colposcopy is far from an infallible truth standard, these figures should be interpreted with caution.

In the protocol, the false negative rate was instead defined as the proportion of subjects detected by colposcopy as positive for CIN 2/3+ but not indicated for biopsy by LUMA. The denominator is all subjects. The numerator is the number of CIN 2/3+ subjects not indicated for biopsy by LUMA. Under the per protocol definition the false negative rate in ASCUS/LSIL subjects was 2.1% (18/876). When including in the numerator CIN 2/3+ subjects with all negative LUMA indicated biopsies, the false negative rate is 3.0% (26/876) (Sponsor's Panel Package, Volume 2, Question 24).

The following table summarizes three methods for estimating FN from PS I:

**Calculations of FN results**

| **FN result** | **23.4%** (26/111) | **2.1%** (18/876) | **3.0%** (26/876) |
|---|---|---|---|
| **numerator** | sum of<br>▪ number of CIN 2/3+ subjects not indicated for biopsy by LUMA<br>▪ number of CIN 2/3+ subjects for which all LUMA-indicated biopsies are negative | number of CIN 2/3+ subjects not indicated for biopsy by LUMA | sum of<br>▪ number of CIN 2/3+ subjects not indicated for biopsy by LUMA<br>▪ number of CIN 2/3+ subjects for which all LUMA-indicated biopsies are negative |
| **denominator** | number of CIN 2/3+ subjects identified by colposcopy | all subjects | all subjects |

V. C. 5. PS I Relative Increase in TP (or Sensitivity) and FP

As discussed in Section IV (FDA Summary), the designs of PS I and PS II do not permit direct calculation of the sensitivity and specificity, because neither study established the true underlying disease status for each patient. However, relative changes of TP and FP can be calculated. Relative change of TP is the same as relative change of sensitivity. For the ASCUS/LSIL subgroup, the relative estimates and their 95% confidence intervals (CI) are indicated below.

**PS I Relative Increases of TP (or Sensitivity) and FP, ASCUS/LSIL**

| | Colpo-only | LUMA+colpo | Relative Increase | 95% CI |
|---|---|---|---|---|
| **TP** | 11.4% (99/871) | 14.4% (126/876) | 26.5% | (-1.0%, 61.8%)* |
| **FP** | 61.2% (533/871) | 65.2% (571/876) | 6.5% | (-0.9%, 14.5%) |

**95% Confidence Intervals calculated by FDA based on the Clopper-Pearson method. The sponsor's confidence intervals for the TP were very similar. The sponsor did not provide a confidence interval for the FP.*

Based on the estimated relative increase in sensitivity of 26.5% for the ASCUS/LSIL population, the sponsor claimed the following in the PMA submission:

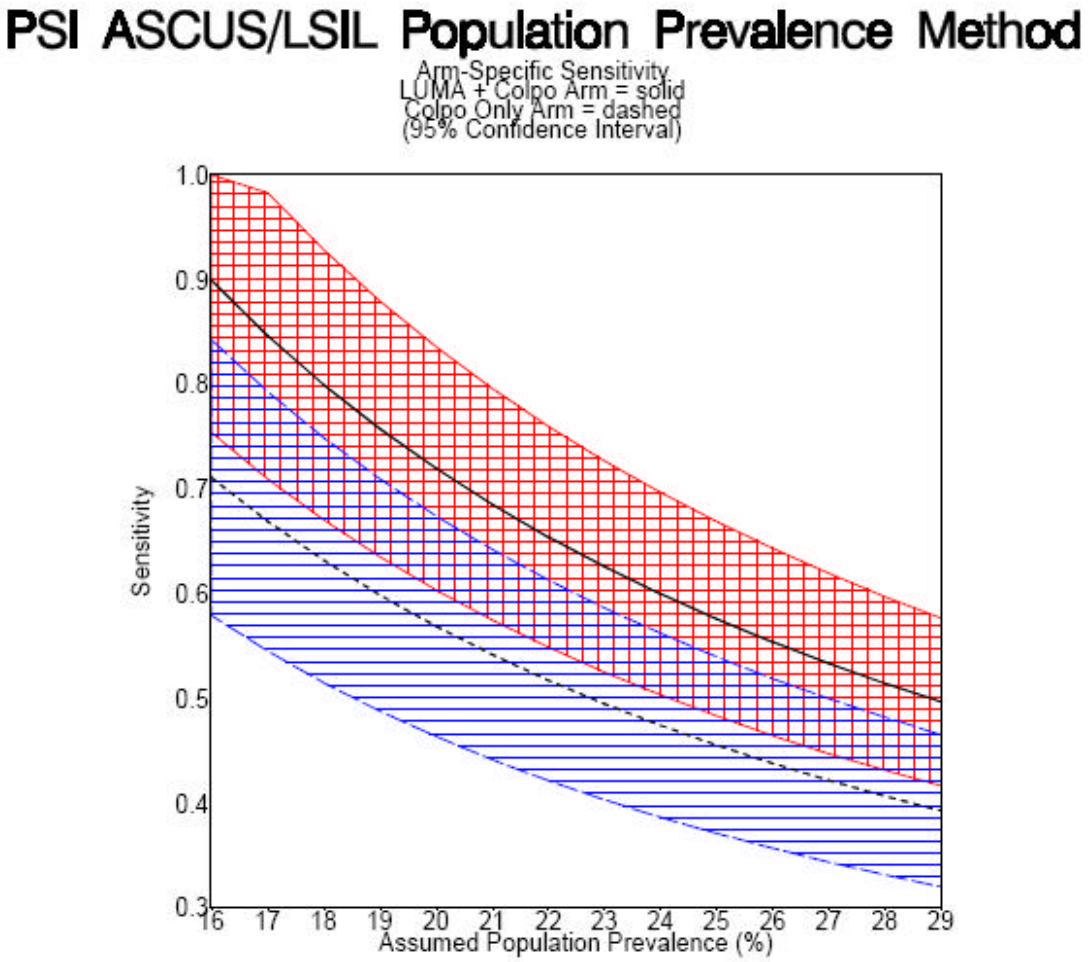
> *Findings from both studies showed that the use of ODS [LUMA] resulted in at least a 25% increase in detection of CIN 2,3+ among women referred to colposcopy for the evaluation of ASCUS or LSIL Pap test result. When extrapolated to the United States as a whole, the improved performance obtained through the use of [LUMA] translates into approximately 100,000 more women per year with C1N 2,3+ being identified and provided with definitive treatment.*

This claim, however, depends on assuming a baseline value sensitivity for colposcopy, or equivalently, a corresponding value for the prevalence of CIN 2/3+. These values cannot be estimated from PS I. Moreover, the claim does not account for the uncertainty associated in the estimate of 26.5%. From the 95% confidence interval lower bound of -1.0%, a decrease in sensitivity cannot even be ruled out.

V. C. 6. PS I Absolute Increase in Sensitivity

Although the design of PS I does not permit direct calculation of the sensitivity, sensitivity can nevertheless be imputed by assuming a value for the prevalence of CIN 2,3+. Because any assumption of the particular value for the prevalence is likely to be false, it is better to compute sensitivities for a range of prevalences. The sponsor provided plots of sensitivity and difference in sensitivity as a function of prevalence (Sponsor's Panel Package, Volume 2, Question 32 and Tab "A010…"). However, the confidence intervals on these quantities were not correct. They were subsequently corrected, but these intervals did not make the Sponsor's Panel Package. They are provided below.

From the first plot, if the prevalence is 16%, sensitivity is estimated to be about 90% in the LUMA+colpo arm and about 70% in the colpo only arm. Note that both sensitivities decrease with increasing prevalence. From the second plot, the difference in sensitivity is not significant for any prevalence and decreases with increasing prevalence. For any prevalence, the sensitivity difference cannot be significant because, as stated above, the TP rate difference is insignificant.

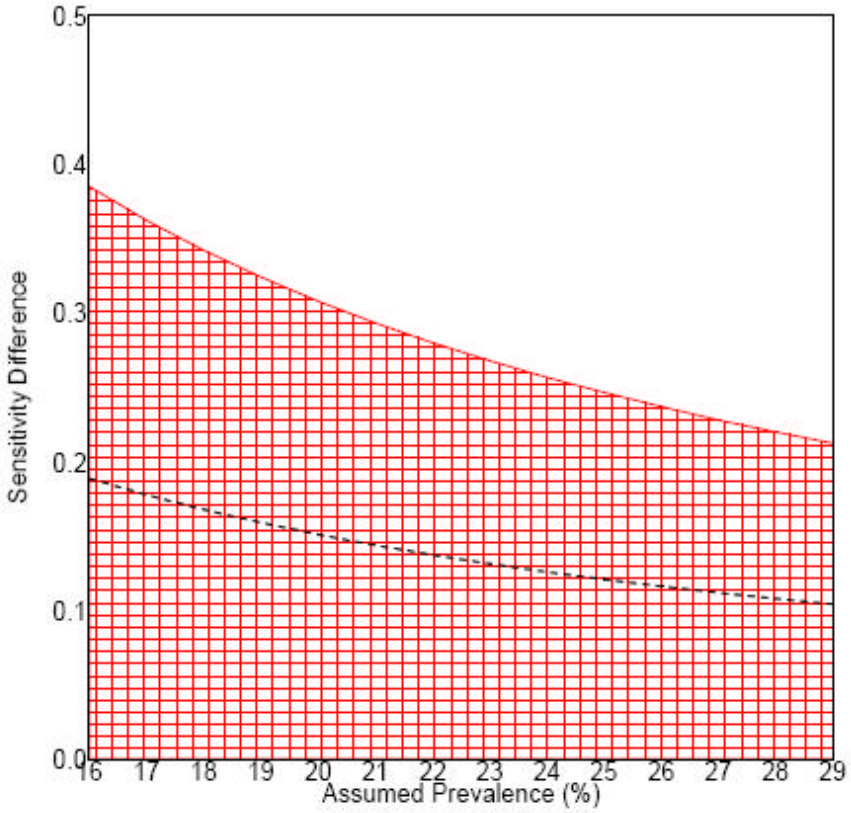



*Solid black line is sensitivity based on LUMA+colpo arm. Red cross-hatching indicates 95% confidence interval.*

*Dashed black line is sensitivity based on colpo-only arm. Blue stripes indicate 95% confidence interval.*

# PSI ASCUS/LSIL Population Prevalence Method

Between-Arm Sensitivity Difference
(95% Confidence Interval)



V. C. 7. PS I Age-Related Device Performance

The purpose of this analysis was to examine if the effect of LUMA depended on age group. As indicated in the table below, the improved TP rate seen with LUMA+colpo in PS I was primarily due to the group of patients under 21 years of age. This trend is consistent across all Pap strata. Within this age group, the FP rate did not increase, suggesting that LUMA increased sensitivity without decreasing specificity for this age group. While this analysis was unplanned, the age groups were pre-specified by the sponsor. The result was not replicated in PS II, perhaps because of the small sample size in that study.

**PS I TP and FP Results by Age Group and Arm**

| | True Positive | | | | | | | | | False Positive | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Age Group | | | | | | | | | Age Group | | | | | | | | | |
| | <21 | | | 21-29 | | | >29 | | | <21 | | | 21-29 | | | >29 | | | |
| | % | # | N | % | # | N | % | # | N | % | # | N | % | # | N | % | # | N | Total N |
| Overall | | | | | | | | | | | | | | | | | | | |
| LUMA+Colpo | **20** | 39 | 191 | 23 | 103 | 447 | 21 | 96 | 452 | **69** | 132 | 191 | 64 | 288 | 447 | 53 | 239 | 452 | 1090 |
| Colpo | **10** | 18 | 179 | 23 | 108 | 464 | 20 | 92 | 453 | **71** | 127 | 179 | 60 | 276 | 464 | 50 | 226 | 453 | 1096 |
| ASCUS | | | | | | | | | | | | | | | | | | | |
| LUMA+Colpo | **17** | 10 | 58 | 18 | 28 | 153 | 9 | 18 | 196 | **72** | 42 | 58 | 66 | 101 | 153 | 57 | 111 | 196 | |
| Colpo | **7** | 4 | 59 | 18 | 26 | 146 | 7 | 12 | 182 | **66** | 39 | 59 | 56 | 81 | 146 | 50 | 90 | 182 | |
| LSIL | | | | | | | | | | | | | | | | | | | |
| LUMA+Colpo | **17** | 19 | 111 | 18 | 38 | 217 | 9 | 13 | 141 | **71** | 79 | 111 | 69 | 149 | 217 | 63 | 89 | 141 | |
| Colpo | **7** | 7 | 97 | 14 | 30 | 218 | 12 | 20 | 169 | **74** | 72 | 97 | 69 | 150 | 218 | 60 | 101 | 169 | |
| HSIL | | | | | | | | | | | | | | | | | | | |
| LUMA+Colpo | **46** | 10 | 22 | 48 | 37 | 77 | 57 | 65 | 115 | **50** | 11 | 22 | 49 | 38 | 77 | 34 | 39 | 115 | |
| Colpo | **30** | 7 | 23 | 52 | 52 | 100 | 59 | 60 | 102 | **70** | 16 | 23 | 45 | 45 | 100 | 34 | 35 | 102 | |
| | | | 370 | | | 911 | | | 905 | | | | | | | | | | |

FDA analyzed these data with a Bayesian multinomial logistic model that included effects for arm, age, PAP stratum, and effects for arm by age, arm by stratum, and age by stratum interactions. The latter two interactions were insignificant, but the age by arm interaction was very significant, indicating that the effect of arm on TP and FP rates depended on age group.

The Bayesian model allowed calculation of the probability that subject TP, FP, and negative (no biopsy) rates were greater in the LUMA+colpo arm than in the colpo-only arm. These probabilities were computed after dropping from the model the two insignificant interaction terms and retaining only the arm by age interaction term. These probabilities indicate significantly improved TP rates within PAP strata for the < 21 age group, but not for the other age groups.

**Bayesian model probabilities that subject true positive (TP), false positive (FP), and test negative (non-biopsied) rates are greater in the LUMA+colpo arm than in the colpo only arm, stratified by PAP stratum and age group**

| | TP | | | FP | | | Negative | | |
|---|---|---|---|---|---|---|---|---|---|
| | Age (years) | | | Age (years) | | | Age (years) | | |
| | <21 | 21-29 | > 29 | <21 | 21-29 | > 29 | <21 | 21-29 | > 29 |
| **LSIL** | **1.00** | **0.79** | **0.54** | **0.62** | **0.96** | **0.90** | **0.01** | **0.01** | **0.10** |
| **ASCUS** | **1.00** | **0.74** | **0.49** | **0.42** | **0.93** | **0.90** | **0.01** | **0.01** | **0.10** |
| **HSIL** | **1.00** | **0.64** | **0.41** | **0.01** | **0.49** | **0.70** | **0.00** | **0.01** | **0.16** |

The sponsor is presently reviewing FDA's Bayesian analysis. The sponsor acknowledged (Sponsor's Panel Package, Volume 2, Question 5) the observed age effect in PS I. They suggested a differential device effect within the =21 patient population, but concluded that the observation was most likely an artifact. The sponsor's conclusion is based on mathematical analysis methods and the sponsor's opinion that "there is no a priori clinical or biological reason to expect device performance to differ in a population of women under vs. over 21 years of age." The sponsor's mathematical analysis suggested that TP rates improved in the LUMA + colpo arm by a greater amount for first time ASCUS subjects than repeat ASCUS subjects. This may explain partially the age effect in PS I.

A theoretical reason why the LUMA might be more sensitive in women =21 is that there may be a lesser degree of cervical surface keratinization among younger women. A difference in keratinization may make the surface of cervical epithelium more reflective and hence give better LUMA reflectance data in younger vs. older women.

V. D. PS I Results: Other Issues

V. D. 1. PS I Reader Variability

The confidence intervals for the TP and FP rates given in the panel packet represent "case", or patient, variability. From Sect. 11.2.3.2 (Sponsor's Panel Package, Volume 1, Tab 5, page 5-031), it is clear that the colposcopist interpretation plays a significant role and contributes an appreciable amount of "reader", or colposcopist, variability to the TP and FP rates as well. This reader variability has not been accounted for in the confidence intervals on TP and FP differences between arms, and is expected to increase their width. Accounting for reader variability would generalize the study from the colposcopists in the study to the population of all colposcopists, assuming that the colposcopists in the study are a representative sample from the population.

V. D. 2. PS I Site Variability

Variability by site in TP and FP rate differences is depicted graphically by plots (Sponsor's Panel Package, Volume 2, Question 23). Formal testing for significant site variability is commonly used to justify pooling of sites. To test for site variability, the sponsor performed logistic regressions to test if the difference in TP and FP rates between arms (colpo or LUMA+colpo) varied significantly by site (Sponsor's Panel Package, Volume 1, Section 12.9.3.13, page 5-156). Sites 7, 10, and 12 were pooled for this analysis because of small sample sizes at these sites. For ASCUS/LSIL subjects, site variability was insignificant for the TP rate (p value 0.4570) but significant for the FP rate (p value 0.0376). In principle, these results indicate that sites can be pooled for the TP rate but not for the FP rate, an undesirable result. For HSIL subjects, site variability was insignificant for both TP and FP rates (p values 0.2190 and 0.2315). However, the power to detect heterogeneity over sites was low in HSIL subjects due to small sample sizes.

V. D. 3. PS I Study Design: Performance Bias

The colposcopists in PS I were not blinded to whether they were in the colpo-only arm or the LUMA+colpo arm; this knowledge could have affected their colposcopy. To assess

performance bias, the rates of colpo-indicated biopsies and biopsy outcomes are compared. For the LUMA+colpo arm, the colpo-indicated biopsies were either indicated by colpo only or indicated by both colpo and LUMA; these were called "any colpo" in the PMA. For the colpo arm, all biopsies were colpo-indicated. As indicated in the table below,

- Among all patients, the number of colpo-indicated biopsies in the LUMA+colpo arm was similar to the number in the colpo-only arm.
- Among patients having at least one biopsy, there were more colpo-indicated biopsies in the LUMA+colpo arm than in the colpo-only arm.
- The biopsy-level PPV among the colpo-indicated biopsies was slightly higher in the LUMA+colpo arm than in the colpo-only arm. (biopsy-level PPV is the proportion of biopsies that are positive for CIN 2/3+).
- Among colpo-indicated biopsies, the percentage of patients that were TP was similar in both arms, but the LUMA+colpo arm had more patients for whom no biopsies were taken and had fewer FP patients.

**Comparison of Colposcopist Performance in Two Arms of PS I**

| | **colpo-only** | **LUMA+colpo** |
|---|---|---|
| | | |
| **colpo-indicated biopsies** | | |
| among all patients | 1.03 (1128/1096) | 1.05 (1140/1090) |
| among patients having at least 1 biopsy | 1.33 (1128/847) | 1.47 (1140/777) |
| | | |
| | | |
| **biopsy-level PPV** *among colpo-indicated biopsies* | | |
| overall | 24.0% | 24.5% |
| ASCUS/LSIL | 14.1% | 15.2% |
| | | |
| **subject-level rates** *among colpo-indicated biopsies* | | |
| TP | 19.9% | 20.0% |
| FP | 57.4% | 50.2% |
| no biopsies taken | 22.7% | 29.8% |

VI. Pivotal Study II (PS II) (Sponsor's Panel Package, Volume 1, Tab 5)

VI. A. PS II Study Design

Pivotal Study II was a prospective, single arm, internally-controlled, multi-center clinical study that was the subject of an agreement letter with FDA. The most important change in design of PS II compared to PS I is that the LUMA overlay was viewed only after a thorough colposcopy was performed and commitment was made to colposcopy biopsy sites. This study design enables conclusions to be drawn regarding the true incremental value LUMA adds to colposcopy. The objective of this study was to evaluate the incremental gain attributed to LUMA once a colposcopic examination has been completed.

The sponsor originally planned to enroll 576 subjects and treat approximately 490 subjects. Based on the prevalence estimated from PS I, the sample size for PS II was increased to 788 patients enrolled to attain 670 subjects. The study was terminated on October 16, 2003 after enrolling 227 subjects at 7 institutions. The first patient was enrolled on July 2, 2003; the last patient was evaluated on October 2, 2003.

In this non-randomized paired study, the colposcopist first committed to biopsy sites based on initial colposcopy and then after being shown the LUMA image committed to additional biopsy sites. With this design, comparison of LUMA+colpo with colpo alone can be made on the same subject and on the same colposcopy. Thus TP and FP rate comparisons are more precise than in PS I, because PS I comparisons included variation in subjects and colposcopy between the LUMA+colpo and colpo-only arms. Furthermore, unlike PS I, TP and FP rate comparisons are not confounded by any systematic differential behavior in colposcopy between the arms. That is, the contribution by LUMA to the TP and FP rates can be isolated. However, the comparisons could be biased if the colposcopists intentionally or unintentionally undercalled (or overcalled) initial colposcopy.

The Per Protocol population was defined as a subset of the Intent to Diagnose population (all enrolled subjects) who completed the protocol, met all eligibility criteria, had a valid LUMA scan, all committed colposcopically directed biopsies taken (if any), all committed LUMA biopsies taken (if any), no un-annotated biopsies obtained, and had majority pathology diagnoses for all biopsies taken (Sponsor's Panel Package, Volume 1, Tab 6, page 6-012).

Per protocol, subjects were excluded if any biopsy had incomplete pathology (unreadable or disagreement between the three pathologists). In contrast, for PS I, subjects were excluded per protocol only if all biopsies for a subject had incomplete pathology.

Per the Study Protocol (Sponsor's Panel Package, Volume 1, Tab 6, page 6-016), an Interim Analysis was planned after 100 to 150 subjects were enrolled in the per-protocol population (fully usable LUMA scans, complete site pathology) representing a minimum of 3 investigational sites. However, this interim analysis never took place because the data were not locked before the decision to terminate the study early was made (Sponsor's Panel Package, Volume 2, Question 27). The interim analysis would have assessed six study assumptions. The study would have been stopped for (a) design

concerns if any of the first three assumptions do not hold, or (b) futility if the conditional power for meeting the primary effectiveness endpoints was <25%.

VI. A. 1. PS II Primary Effectiveness Endpoints

The primary effectiveness endpoints were the TP percent and the FP percent for the incremental LUMA evaluation at the subject level and the Positive Predictive Value (PPV) percent at the biopsy level. The primary hypotheses to be shown were that for the per protocol population,

- TP rate increment was greater than 2% (one-sided p-value <0.025)
- FP rate increment was less than 15% (one-sided p-value <0.025)

PPVs for incremental LUMA evaluation and the initial colposcopic evaluation were to be compared, but no hypothesis was pre-specified for study success.

VI. A. 2. PS II Secondary Effectiveness Endpoints

The secondary effectiveness endpoints were:

- Mean number of biopsies per colposcopy for initial colposcopic examination
- Biopsy-level TP and FP percents, and PPV for the initial colposcopy and LUMA increment
- Subject-level TP and FP percents and biopsy-level PPV for the initial colposcopic examination and for the LUMA increment within strata at the subject level (ASC-US/-H, LSIL and HSIL)
- Subject-level TP and FP percents and number of biopsies for the same colposcopists in Pilot Study II/Pivotal Study I, 3-01803, to be used as an external control
- TP and FP percents for the LUMA increment at the subject level for the subgroup comprised of the per-protocol population and any eligible subjects (with complete pathology) who were excluded from the per-protocol population because of an absent or invalid LUMA result

VI. A. 3. PS II Inclusion and Exclusion Criteria

The inclusion criteria are provided below. *Changes from PS I are in italics*.

- Signed informed consent
- Age 18 (or age of consent per local regulations) and older
- Documented abnormal Pap smear defined as first-time or repeat ASC-US/-H, LSIL, HSIL or cancer or equivalent classification using Bethesda II. *Subjects with a referral Pap diagnosis of Atypical Glandular Cells of Unknown Significance should not be enrolled.*

The exclusion criteria were as follows:

- Pregnant through six weeks post-partum
- Menstruating (heavy flow) at the time of the LUMA
- Medical history of total hysterectomy (without a cervix) or congenital anatomical cervical variants, e.g. double cervix

- Cervical stenosis (unable to insert Q-tip or cytobrush into endocervical canal, *and thus unable to obtain a clean, directed biopsy)*
- History of DES exposure in utero
- Punch biopsy or ablative/excisional therapeutic procedure since Index Pap smear (e.g. cervical conization, LEEP, laser, cryosurgery, radiation treatment, etc)
- Use of vaginal creams or vaginal medications (including all douches except saline) within the 48-hours immediately prior to LUMA examination
- Current use of photosensitizing agents (within 72-hours) such as *sulfasalazine* and tetracycline
- History of photosensitivity or other diseases affected by UV radiation (e.g. porphyria, Systemic Lupus Erythematosis, etc)
- Undergoing phototherapy
- Previously enrolled in prior MediSpectra studies (excluding the MediSpectra Menstrual Study, 3-01350, and Verification Studies 3-01017, 3-01575, and 3-01710)
- Participated in an investigational drug or device study within the past 30 days
- Subjects with untreated gynecological infections, such as acute cervicitis, acute inflammation and acute vaginitis
- Subjects who required a Pap smear at the time of the LUMA examination
- Index Pap smear that is greater than 6-months old at time of study visit, or within prior 7 days
- *Subjects who have been referred for LEEP procedure*

VI. A. 4. PS II Procedure

The subject's identification number and colposcopist's name were entered into the LUMA system. A speculum was inserted into the vagina and the LUMA probe was placed at the opening to the speculum in such a way as to center the cervix in the field of view. The cervix as then gently washed with saline. Acetic acid was liberally applied to the cervix and pooled fluid removed from the vagina.

The LUMA system was focused and a scan made over a 15-second interval. (Up to 5 additional scans could be taken if the previous scan is deemed inadequate, e.g. due to motion by the subject or inadequate focus.) The LUMA results were not displayed but stored for later retrieval. The LUMA probe was moved aside and the colposcope positioned.

The colposcopist performed a routine colposcopic examination, committed to and annotated all intended biopsy sites on an electronic image.

Following the colposcopic examination and annotation of biopsy sites, the LUMA results were unblinded and the LUMA provided a cervical image with an overlay of any potential areas of CIN 2/3+.

The colposcopist then identified and committed to additional biopsy sites based on the LUMA scan. These were annotated on the electronic image. The LUMA printed a cervical image with the committed/electronically annotated biopsy sites indicated. The colposcopist took all annotated colposcopy and LUMA-directed biopsies.

Details of the biopsy selection procedure are provided in the Sponsor's Panel Package, (Volume 1, Tab 6, Sub-tab "Interpretation of ODS…"; Volume 2, Question 30).

All biopsy specimens are sent to an on-site pathologist who prepares two diagnostically equivalent slides from each specimen. The "study slide" is then sent to the contract research organization (CRO) and from there to the Central Pathology Review for independent reading by at least two pathologists. If the two pathologists disagreed on the diagnosis, the slide is reviewed by a third pathologist.

VI. B. PS II Results: Overview and Primary Endpoints

Pivotal Study II was initiated in July 2003. The original protocol had a sample size of 576 subjects. This was based on the assumption that the underlying CIN 2/3 prevalence was 33%. After the study began, the results of PS I became available. As a result, the sponsor reduced the assumed prevalence of CIN 2/3+ from 33% to 27%. To adjust for the lower disease prevalence in a comparable patient population, the sample size for PS II was increased to 788 to allow for a 15% loss of subjects resulting in 670 subjects.

On October 16, 2003, the Board of Directors of MediSpectra decided that PS II could not continue; they indicate this decision was made for business/economic reasons. The PMA states that "this decision was made approximately six weeks before the primary endpoint analysis of Pivotal Study II was conducted." (Sponsor's Panel Package, Volume 1, Tab 6, page 6-095) Enrollment was terminated at 227 subjects at seven institutions.

Of the 227 subjects enrolled, 34 (15.0%) were discontinued. Seven (3.0%) were discontinued by the investigator before the examination was completed, and 27 (11.9%) were excluded after randomization. The primary reason for exclusion post-randomization was labeled as "no majority Pathology diagnosis" (16 subjects, 7.0%), i.e. at least one biopsy for the subject was either unreadable or had a disagreement among the reference pathologists. The device malfunction rate was smaller (0.2%) in PS II compared to PS I (1.1%). Within the ASCUS/LSIL sub-population, 6.2% (12/193) were excluded because of "no majority pathology diagnosis" The "no majority Pathology diagnosis" exclusion was more rigorous in PS II compared to PS I: in PS II, subjects were excluded if *any* specimen lacked a majority pathology diagnosis; in PS I, subjects were excluded only if *all* specimens lacked a majority pathology diagnosis. The possible impact of this change is considered in Section VI.D.4 (FDA Summary).

The mean age of study participants was 28.5 (SD 10.6) years.

Racial demographics were as follows:

| | |
|---|---|
| Caucasian | 43.0% |
| African American | 21.2% |
| Hispanic | 31.1% |
| Asian | 1.6% |
| Native American | 1.6% |

The referral Pap smear study profile was:

| | |
|---|---|
| ASCUS | 39.0% |
| LSIL | 47.0% |
| HSIL/Cancer | 14.0% |

Analogous results for PS I are provided in Section V.B (FDA Summary).

The average number of biopsies obtained increased by about one when the LUMA overlay was used for identifying biopsy sites, as shown in the table below.

**PS II Mean Number of Biopsies per Stratum within Per Protocol Population (95% confidence interval)**

| Strata | Colposcopy | LUMA Increment |
|---|---|---|
| **Overall** | 0.89 (0.80, 0.98)* | 1.02 (0.92, 1.12) |
| **ASCUS/LSIL** | 0.81 (0.72, 0.90) | 0.99 (0.88, 1.10) |
| **HSIL** | 1.42 (1.18, 1.67) | 1.23 (0.94, 1.52) |

**95% Confidence Intervals provided by sponsor*

Numbers of biopsies taken in PS I are provided in Section V.B (FDA Summary).

Not all subjects were biopsied. The no-biopsy rate at colposcopy in the Per Protocol population was 26.9% (52/193). The no-biopsy rate at incremental LUMA for the Per Protocol population was 22.3% (43/193). For the combination of both evaluations, 7.8% (15/193) of the subjects did not undergo any biopsy (Sponsor's Panel Package, Volume 1, Tab 6, page 6-120).

VI. B. 1. PS II Safety Results

With respect to device safety, there were 3 reports of adverse events in PS II reported in the PMA. Two of these events occurred in the same patient. These events consisted of fainting in one patient and abdominal pain and dysuria both of which occurred in another patient. None of the events was serious, and none was believed to be related to the LUMA device.

VI. B. 2. PS II Primary Effectiveness Endpoints

According to the protocol, two primary effectiveness endpoints of PS II pertain to TP and FP LUMA increments. These increments are defined in the protocol as the number of TP and FP subjects additional to initial colposcopy divided by the total number of subjects. The study hypotheses were developed using this definition. These results are provided in the table below.

- The TP primary endpoint was met, i.e. for the TP rate difference, the lower bound of the 95% confidence interval was greater than 2%. This incremental improvement of 4.7% corresponded to 9 additional TPs being detected by LUMA. The colposcopy and LUMA images for these 9 TP subjects are provided in (Sponsor's Panel Package, Volume 1, Tab 6, page 6-137).

- The FP primary endpoint was not met, i.e. the 95% confidence interval for the FP rate difference included 15%.

**PS II: Primary Effectiveness TP and FP Endpoints (All Subjects, per protocol definition of LUMA increment)**

| | Initial Colposcopy Rate (n/N) | 95% CI | LUMA Increment Rate (n/N) | 95% CI | Met Hypothesis? |
|---|---|---|---|---|---|
| **TP** | 21.2% (41/193) | 15.7%, 27.7% | **4.7% (9/193)** | **2.2%, 8.7%** | **Yes**: 95% CI >2% |
| **FP** | 51.8% (100/193) | 44.5%, 59.0% | 18.1% (35/193) | 13.0%, 24.3% | **No**: 95% CI includes 15% |

Note that the sponsor has also created a different denominator for calculating TP and FP increments. Results with this denominator are provided in Section VI.C.2 (FDA Summary). Elsewhere in the panel package, the sponsor refers to the results with the new denominator as "*device performance*," and results with the protocol denominator as "*alternate definition*."

Biopsy level PPV was also a primary endpoint, but there was no statistical hypothesis. From the results in the table below, one cannot tell if LUMA biopsies are better than colpo-indicated biopsies or better than random biopsies. However, LUMA biopsies that were selected by the colposcopist appear to be better than the random biopsies in the Shanxi Province Cervical Cancer Screening Study II (Pretorius et al, Am J Obstet Gynecol, 2004, 191:430-434; cf. Sponsor's Panel Package, Volume 2, Question 20). For LSIL and ASCUS subjects, respectively, PPVs were 1.7% and 3.6% for random biopsies in that study, compared with 10.9% overall for ASCUS/LSIL LUMA-only indicated biopsies in PS II. These results suggest that among LUMA-only locations indicated for biopsy, the ones that were selected for biopsy by the colposcopist (in addition to his/her colpo-indicated biopsies) are much better than random selections for biopsy.

**PS II Biopsy-level PPV% by Cervical Assessment Technique and Referral Strata**

| **Pap Referral Strata** | **PPV % 95% CI** | | |
|---|---|---|---|
| | **Colposcopy** | **LUMA Increment** | **Colposcopy plus LUMA** |
| **Overall N=193** | **26.2%** (45/172) 19.8%, 33.4% | **18.3%** (36/197) 13.1%, 24.4% | **22.0%** (81/369) 17.7%, 26.5% |
| **ASCUS/LSIL N=167** | **19.3%** (26/135) 13.0%, 26.9% | **10.9%** (18/165) 6.6%, 16.7% | **14.7%** (44/300) 10.9%, 19.2% |
| **HSIL N=26** | **51.4%** (19/37) 34.4%, 68.1% | **56.3%** (18/32) 37.7%, 73.6% | **53.6%** (37/69) 41.2%, 65.7% |

VI. C. PS II Results: Panel Review Issues

VI. C. 1. PS II Analysis by Pap Stratum

The sponsor is presently seeking an indication specifically for the ASCUS/LSIL sub-population. The TP and FP results are provided by entry pap result in the tables below. Although analysis of TP and FP rates within Pap strata was pre-specified as a secondary endpoint, analysis of the ASCUS/LSIL subgroup was not pre-specified as either a primary or secondary endpoint. As discussed in Section VI.C.1 (FDA Summary), if this study had been designed with multiple sub-group analysis, then the sub-group analysis would have to be corrected for multiplicity to control the overall chance of finding a falsely significant result.

**PS II Subject-level True Positive Rates by Entry Pap Result (per protocol definition of LUMA increment)**
*Original Hypothesis: 95% confidence (two-sided) that the increment from LUMA will be at least 2%*

| **Strata** | **Initial Colposcopy** | | **LUMA Increment** | |
|---|---|---|---|---|
| | **Rate (n/N)** | **95% CI** | **Rate (n/N)** | **95% CI** |
| **Overall** | 21.2% (41/193) | 15.7%, 27.7% | **4.7% (9/193)** | **2.2%, 8.7%** |
| **ASCUS/LSIL** | 14.4% (24/167) | 9.4%, 20.6% | 3.6% (6/167) | 1.3%, 7.7% |
| **HSIL** | 65.4% (17/26) | 44.3%, 82.8% | **11.5% (3/26)** | **2.4%, 30.2%** |

Although the original hypothesis for the TP primary endpoint was met for the overall study population and the HSIL sub-population, it was not met for the ASCUS/LSIL sub-populations, i.e. for the TP rate LUMA increment, the 95% confidence interval includes 2%.

**PS II Subject-level False Positive Rates by Entry Pap Result (per protocol definition of LUMA increment)**
*Original Hypothesis: 95% confidence (2-sided) that the increment from LUMA will be less than 15%*

| **Strata** | **Initial Colposcopy** | | **LUMA Increment** | |
|---|---|---|---|---|
| | **Rate (n/N)** | **95% CI** | **Rate (n/N)** | **95% CI** |
| **Overall** | 51.8% (100/193) | 44.5%, 59.0% | 18.1% (35/193) | 13.0%, 24.3% |
| **ASCUS/LSIL** | 55.7% (93/167) | 47.8%, 63.4% | 20.4% (34/167) | 14.5%, 27.3% |
| **HSIL** | 26.9% (7/26) | 11.6%, 47.8% | 3.8% (1/26) | 0.1%, 19.6% |

The original hypothesis for the FP primary endpoint was not met for the overall study population or for any of the sub-populations, i.e. the 95% confidence interval (CI) on the increment for all populations include 15%.

Another reasonable way of determining FP increment is as follows: there were 35 additional FPs in the LUMA increment, and there were 7 patients who were FP in the initial colposcopy and became TP with the LUMA increment. Hence, the net increase of FP with the LUMA increment could also be calculated as 14.5% ((35-7)/193) (FDA 95% CI 9.8%, 20.3%)

VI. C. 2. PS II Changes to Primary Endpoints

The sponsor has proposed 2 changes to the analysis of the results of PS II compared to the original protocol:

1. **Success criteria**: The original protocol called for a 2% improvement of TP resulting from LUMA-directed biopsies. The sponsor suggests that this 2% can be changed post-hoc to 1.5% and 4% For the ASCUS/LSIL and HSIL groups to reflect the different prevalences measured in PS I. The sponsor suggests FDA should accept the new threshold values for the ASCUS/LSIL population (Sponsor's Panel Package, Volume 2, Question 7), because FDA agreed to change in number of subjects enrolled based on disease prevalence (FDA Summary, Section VI.A).

2. **Denominator**: Study hypotheses were developed using **per protocol** definitions of the TP and FP rate increments in which the number of true and false positive subjects additional to initial colposcopy is divided by all subjects. These increments are called the *alternative definition* increments elsewhere in the panel package. In a **new definition** of TP and FP rate increments, the sponsor subtracted from the per protocol population the number of subjects who were found to have CIN 2/3+ based on the colposcopic biopsy. This reduced the denominator used in the calculation of the TP rate and increased the rates. The results with the new denominator are called the *device performance* increment in the PMA. Under the latter definition, the TP rate can be regarded as a measure of the attributable risk in not using the device. It is a lower bound on increment is the percent reduction in false negatives, which is an important quantity but not per protocol.

With the new definitions, the TP rates for PS II change as indicated in the table below. Using the per protocol analysis, the TP endpoint was met for the "overall" and HSIL populations. However, using the reduced denominator and new success criteria, the success criteria are met in all three populations.

**PS II TP Results When (1) Different Null Hypothesis Gets Assigned to Each Pap Substratum and (2) Denominator is Calculated Differently from Original Agreement with FDA**

| Stratum | New TP Hypothesis | 95% CI for LUMA Increment | |
|---|---|---|---|
| | | Per protocol, Full Denominator | New Definition, Reduced Denominator |
| **Overall** | >2% | **2.2%, 8.7% (p=0.0164)** | **2.7%, 10.9% (p=0.0037)** |
| **ASCUS/LSIL** | >1.5% | 1.3%, 7.7% (p=0.0411) | **1.6%, 8.9% (p=0.0214)** |
| **HSIL** | >4% | **2.4%, 30.2% (p=0.0840)** | **7.5%, 70.1% (p=0.0045)** |

For analogous changes to the FP rate analysis, the study does not meet the revised FP requirements for either the overall population or the ASCUS/LSIL population.

VI. C. 3. PS II Study Design: Biopsy Requirements

As indicated in the table below, the mean number of colpo-indicated biopsies was slightly higher in the colpo-only arm of PS I (1.03, overall population) than for initial colposcopy in PS II (0.89, overall population). However, when the LUMA increment is added, then the mean number of biopsies in PS II (0.89+1.02 = 1.91, overall population) is considerably more than either the colpo-only arm of PS I (1.03, overall population) or the LUMA+colpo arm of PS I (1.30, overall population).

The larger number of biopsies in PS II could contribute to the increases of both TP and FP in the LUMA increment of this study, and could be a result of the study design and the instructions for taking biopsies.

**PS I and PS II Mean Number of Biopsies by Strata within Per Protocol Population**

| Strata | PS I Colpo-Only | PS I LUMA+Colpo | PS II Colposcopy | PS II LUMA Increment |
|---|---|---|---|---|
| **Overall** | 1.03 | 1.30 | 0.89 | 1.02 |
| **ASCUS/LSIL** | 0.93 | 1.24 | 0.81 | 0.99 |
| **HSIL** | 1.40 | 1.57 | 1.42 | 1.23 |

VI. C. 4. PS II False Negatives

Among the 45 colpo-indicated biopsies that were CIN 2/3+, 10 were LUMA-negative. That is, **22.2%** (10/45) of these 45 TP biopsies would not have been diagnosed if LUMA were the only tool being used to identify biopsy sites. Put another way, the biopsy-level sensitivity for ASCUS/LSIL subjects of LUMA can be estimated as 77.8% when colposcopy is used as the gold standard in determining true disease status. However,

because colposcopy is far from an infallible truth standard, this result should be interpreted with caution.

The sponsor estimated the False Negative rate of LUMA, by identifying cases in which CIN 2/3+ sites were identified by colposcopy but not by LUMA (Sponsor's Panel Package, Volume 2, Question 24). For PS II, there are digital images that indicate where colposcopic biopsies were made as well as the image of the LUMA overlay. There were six FNs within the ASCUS/LSIL strata. For 4/6 of these, the sponsor's expert could not find any condition visible on the image that might be expected to have hindered the performance of LUMA. In 1/6 case, there might have been insufficient acetic acid application to yield optimal LUMA results. In the remaining case, the colposcopically positive biopsy was obtained from an area of white epithelium that had two LUMA blue areas and the biopsy site overlaid a LUMA yellow area. Therefore, this probably should not have been counted as a FN.

In the HSIL referral stratum for PS II, there were 5 biopsy sites that were positive based on colposcopy and which were LUMA negative. No explanation could be found for 4/5. In one, there were some obscuring features that might have interfered with LUMA.

VI. C. 5. PS II Relative Increase in TP (or Sensitivity) and FP

As discussed in Section IV (FDA Summary), the designs of PS I and PS II do not permit direct calculation of the sensitivity and specificity, because neither study established the true underlying disease status for each patient. However, relative changes in TP and FP can be calculated. Relative change of TP is the same as relative change in sensitivity. For the ASCUS/LSIL subgroup, the relative estimates and their 95% confidence intervals (CI) are indicated below

**PS II Relative Increases of TP (or Sensitivity) and FP, ASCUS/LSIL**

| | **Colpo-only** | **LUMA increment** | **Relative Increase** | **(95% CI) – calculation model(s)** |
|---|---|---|---|---|
| **TP** | 14.4% (24/167) | 3.6% (6/167) | 25% | (10%, 47%)* |
| **FP** | 55.7% (93/167) | 20.4% (34/167) | 37% | (27%, 47%) |

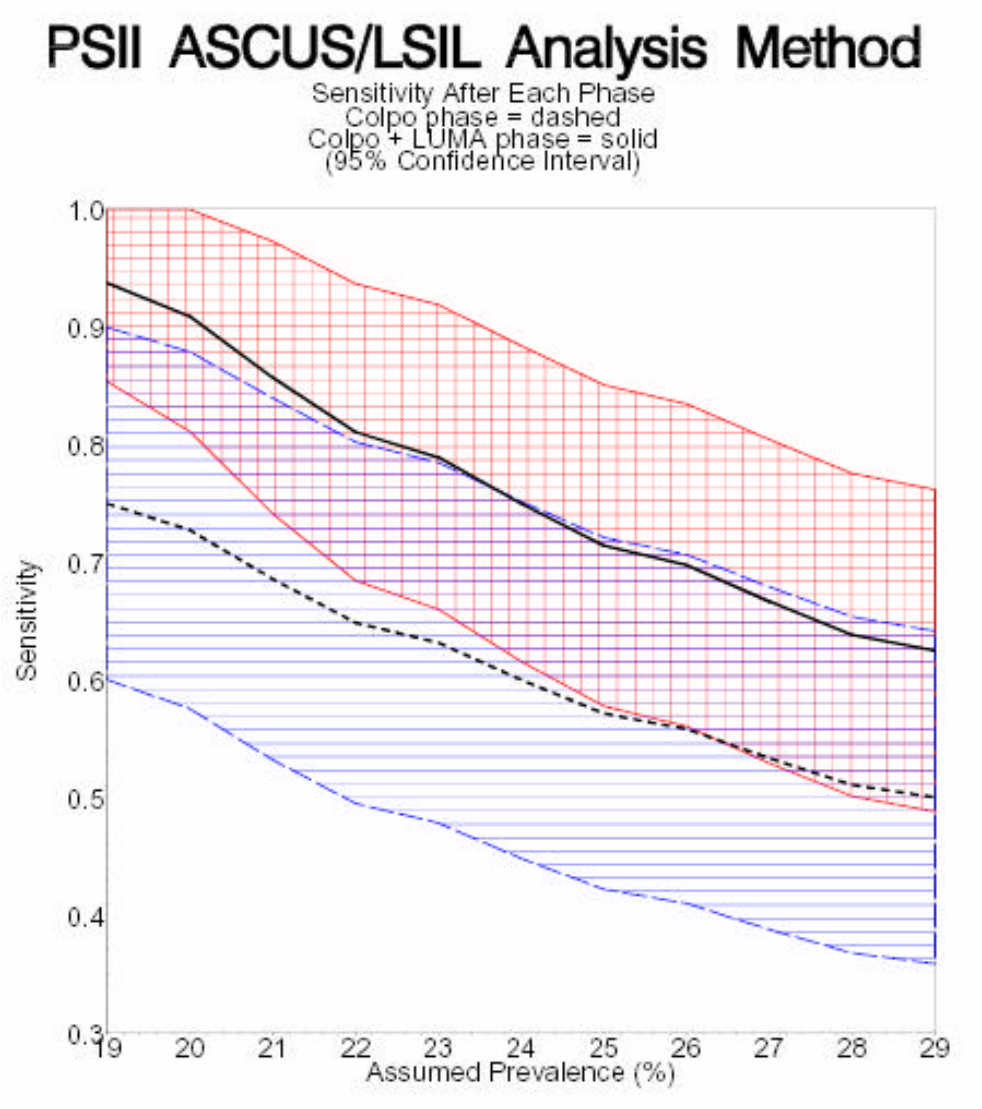
**95% Confidence Intervals calculated by FDA based on the Clopper-Pearson method. The sponsor's confidence interval for the TP was similar. The sponsor did not provide a confidence interval for the FP.*

As discussed in Section V.C.5 (FDA Summary), the sponsor indicates that the use of LUMA resulted in at least a 25% increase in detection of CIN 2,3+ among women referred to colposcopy for the evaluation of ASCUS or LSIL Pap test result, and that this improvement can be extrapolated to approximately 100,000 more women per year with CIN 2,3+ being identified and provided with definitive treatment the United States. This claim, however, depends on assuming a baseline value sensitivity for colposcopy. This value cannot be estimated from PS II. More over, the claim does not account for the uncertainty associated in the estimate of 25%. From the 95% confidence interval, the relative increase in sensitivity can be as low a 10.0%.

VI. C. 6. PS II Absolute Increase in Sensitivity

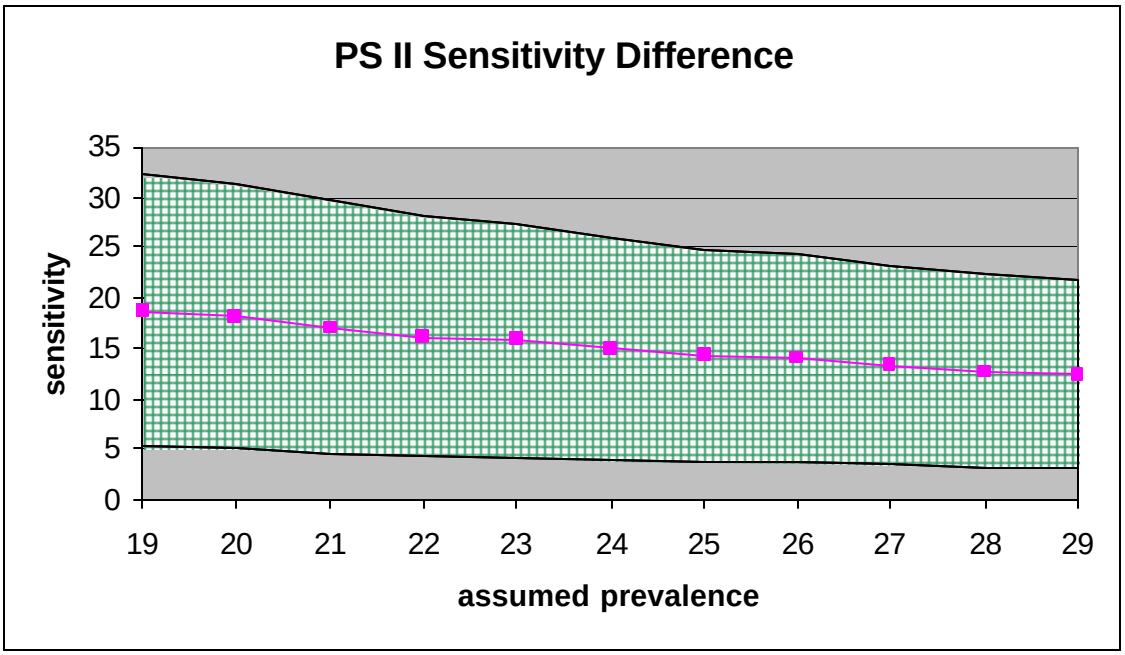
While the design of PS II does not permit direct calculation of the sensitivity, sensitivity can nevertheless be imputed by assuming a value for the prevalence of CIN 2,3+. The sponsor provided plots of sensitivity and difference in sensitivity as a function of prevalence (Sponsor's Panel Package, Volume 2, Tab "A010…"). However, the confidence intervals on these quantities were incorrect. They were subsequently corrected, but these intervals did not make the panel pack. The correct intervals are provided below.

From the first plot, for a prevalence of 19%, sensitivity is estimated to be about 94% in the LUMA+colpo combined phases and about 75% in the initial colpo phase. Note that both sensitivities decrease with increasing prevalence. From the second plot, the difference in sensitivity is significant, with the lower bound on the 95% confidence interval being of about 5% that decreases only slightly with increasing prevalence. By design, the difference is necessarily significant because after initial colposcopy, extra biopsies are taken in the LUMA phase, which necessarily result in some increase in sensitivity.




*Solid black line is sensitivity based on colposcopy with LUMA. Red cross-hatching indicates 95% confidence interval.*

*Dashed black line is sensitivity based on colposcopy only. Blue stripes indicate 95% confidence interval.*




VI. D. PS II Results: Other Issues

VI. D. 1. PS II Reader Variability

As with PS I, the confidence intervals for the LUMA incremental TP and FP rates in PS II represent "case", or patient, variability (FDA Summary, V.D.1), but colposcopist interpretation plays a significant role and contributes an appreciable amount of "reader", or colposcopist, variability to the TP and FP rates as well (Sponsor's Panel Package, Volume 1, Tab 5, Section 11.2.3.2, page 5-031). This reader variability has not been accounted for in the confidence intervals on TP and FP differences between arms, and is expected to increase their width. Accounting for reader variability would generalize the study from the colposcopists in the study to the population of all colposcopists, assuming that the colposcopists in the study are a representative sample from the population.

VI. D. 2. PS II Site Variability

Over the seven sites in the study, the LUMA incremental TP rate ranged from 0.0% (0/18) to 12.2% (5/41) (Sponsor's Panel Package, Volume 1, Tab 6, page 6-115, Table 20). The incremental FP rate ranged from 0.0% (0/17) to 33.9% (20/59). Formal testing for significant site variability is commonly used to justify pooling of sites. The sponsor tested for site variability by using Cochran-Mantel-Haenszel tests that controlled for PAP stratum (Sponsor's Panel Package, Volume 1, Tab 6, page 6-115 to 6-117). For the per protocol ("alternative") definition, site variability was insignificant for the LUMA incremental TP rate (p value 0.1890), but significant for the LUMA incremental FP rate (p value 0.0025). In principle, these results indicate that sites can be pooled for the TP rate but not for the FP rate, an undesirable result.

VI. D. 3. PS II Study Design: Performance Bias

Because of the design of PS II, the study results could be biased in favor of LUMA effectiveness if initial colposcopy were undercalled. The TP and FP rates and biopsy-level PPVs for the initial colpo-biopsies in PS II are similar to those in among the colpo-indicated biopsies in the LUMA+colpo and colpo arms of PS I. This suggests that initial colposcopy was not undercalled in PS II.

**Comparison of colposcopist performance, colpo-indicated biopsies**

| | **PS I colpo-only** | **PS I LUMA+colpo** | **PS II initial colpo** |
|---|---|---|---|
| **subject-level TP and FP** | | | |
| **TP** | 19.9% | 20.0% | 21.2% |
| **FP** | 57.4% | 50.2% | 51.8% |
| | | | |
| **biopsy-level PPV** | | | |
| **overall** | 24.0% | 24.5% | 26.2% |
| **ASCUS/LSIL** | 14.1% | 15.2% | 19.3% |
| **HSIL** | 49.8% | 50.3% | 51.4% |

However, comparisons between PS I and PS II are not adjusted for the fact that not all colposcopists were in both studies. The comparison for the overall and ASCUS/LSIL populations are also not adjusted for any differential between PS I and PS II in the distribution of cases by PAP stratum.

VI. D. 4. PS II Study Design: Subjects Excluded due to “incomplete pathology”

FDA examined the 16 subjects excluded from the PS II per protocol population due to incomplete pathology. Among these subjects, 10 have pathology results for some of the biopsies that provide information on whether a LUMA incremental TP or FP is impossible:

- In six subjects, an incremental TP is impossible because either no biopsies were indicated by LUMA or pathology was available on some biopsies and indicated a TP for initial colposcopy. When these five subjects are included in the TP analysis, the incremental TP rate for the overall population is 4.55% (9/199) with 95% confidence interval (2.09%, 8.41%), so the TP primary endpoint requiring increment greater than 2% is still demonstrated.

- In nine subjects, an incremental FP is impossible because either no biopsies were indicated by LUMA or pathology was available on some biopsies and indicated either a TP or FP for initial colposcopy. In one subject an incremental FP is guaranteed because available pathology indicated a TP for initial colposcopy and a FP for the one LUMA indicated biopsy. When these nine subjects are included in the FP analysis, the incremental FP rate for the overall population is 17.24% (35/203) with 95% confidence interval (12.31%, 23.15%),

so the FP primary endpoint requiring increment no more than 15% is still not demonstrated.

The sponsor also provided a worst-case-scenario analysis on the intent-to-diagnosis population of 227 subjects (Sponsor's Panel Package, Volume 2, Question 28, Table 28.2.B.i). In the worst-case scenario, missing colpo biopsies were assigned as TPs and missing LUMA biopsies were assigned as FPs. FDA is still assessing this analysis.

VI. D. 5. PS II Early Stopping

As discussed above, this study was intended to enroll 788 patients; however, it was terminated after enrolling 227 patients. Apparently, the sponsor did not monitor the trial. An argument can be made that because the study was terminated early with no opportunity to continue, the one-time analysis of the study at its termination does not need an adjustment for interim analysis because Type I error rate is controlled at 5%.