Accuracy of optical spectroscopy for the detection of cervical intraepithelial neoplasia: testing a device as an adjunct to colposcopy

Scott B. Cantor, Ph.D.[1], **Jose-Miguel Yamal, Ph.D.**[2], **Martial Guillaud, Ph.D.**[3], **Dennis D. Cox, Ph.D.**[4], **E. Neely Atkinson, Ph.D.**[1], **J. L. Benedet, M.D.**[3], **Dianne Miller, M.D.**[5], **Thomas Ehlen, M.D.**[5], **Jasenka Matisic, M.D.**[6], **Dirk van Niekerk, M.D.**[6], **Monique Bertrand, M.D.**[5], **Andrea Milbourne, M.D.**[7], **Helen Rhodes, M.D.**[7], **Anais Malpica, M.D.**[7,8], **Gregg Staerkel, M.D.**[8], **Shahla Nader-Eftekhari, M.D.**[9], **Karen Adler-Storthz, Ph.D.**[10], **Michael E. Scheurer, Ph.D.**[11], **Karen Basen-Engquist, Ph.D., M.P.H.**[12], **Eileen Shinn, Ph.D.**[12], **Loyd A. West, M.D.**[7], **Anne-Therese Vlastos, M.D.**[7], **Xia Tao, M.D., Ph.D.**[7], **J. Robert Beck, M.D.**[13], **Calum MacAulay, Ph.D.**[3], and **Michele Follen, M.D., Ph.D.**[7,9,14]

Scott B. Cantor: sbcantor@mdanderson.org; Jose-Miguel Yamal: jose-miguel.yamal@uth.tmc.edu; Martial Guillaud: mguillau@bccrc.ca; Dennis D. Cox: dcox@rice.edu; E. Neely Atkinson: eatkinso@mdanderson.org; J. L. Benedet: lbenedet@interchange.ubc.ca; Dianne Miller: dianne.miller@vch.ca; Thomas Ehlen: Tom.Ehlen@vch.ca; Jasenka Matisic: jmatisic@bccancer.bc.ca; Dirk van Niekerk: dvanniek@bccancer.bc.ca; Andrea Milbourne: amilbour@mdanderson.org; Helen Rhodes: hrhodes@mdanderson.org; Anais Malpica: amalpica@mdanderson.org; Gregg Staerkel: gstaerke@mdanderson.org; Shahla Nader-Eftekhari: Shahla.Nader-Eftekhari@uth.tmc.edu; Karen Adler-Storthz: Karen.A.Storthz@uth.tmc.edu; Michael E. Scheurer: scheurer@bcm.edu; Karen Basen-Engquist: kbasenen@mdanderson.org; Eileen Shinn: eshinn@mdanderson.org; Loyd A. West: Loyd.A.West@dartmouth.edu; Calum MacAulay: cmacaula@bccrc.ca; Michele Follen: michelefollen@hotmail.com

[1] Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas

[2] Division of Biostatistics, The University of Texas School of Public Health, Houston, Texas

[3] Department of Cancer Imaging, British Columbia Cancer Research Centre, Vancouver, British Columbia, Canada

[4] Department of Statistics, Rice University, Houston, Texas

[5] Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, University of British Columbia, Vancouver, British Columbia, Canada

[6] Department of Pathology, University of British Columbia, Vancouver, British Columbia, Canada

[7] Department of Gynecologic Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas

[8] Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, Texas

[9] Department of Obstetrics, Gynecology, and Reproductive Sciences, The University of Texas Health Science Center at Houston, Houston, Texas

[10] The University of Texas Health Science Center at Houston Dental Branch, Houston, Texas

Corresponding Author: Scott B. Cantor, Ph.D., Department of Biostatistics - Unit 1411, The University of Texas M. D. Anderson Cancer Center, P. O. Box 301402, Houston, TX 77230-1402, phone: 713-745-5984, fax: 713-563-4243, sbcantor@mdanderson.org.

[11] Department of Pediatrics and Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas

[12] Department of Behavioral Science, The University of Texas MD Anderson Cancer Center, Houston, Texas

[13] Fox Chase Cancer Center, Philadelphia, Pennsylvania

[14] Department of Obstetrics, Gynecology, and Reproductive Sciences, the Lyndon Baines Johnson Hospital, Houston, Texas

Abstract

Testing emerging technologies involves the evaluation of biologic plausibility, technical efficacy, clinical effectiveness, patient satisfaction, and cost-effectiveness. The objective of this study was to select an effective classification algorithm for optical spectroscopy as an adjunct to colposcopy and obtain preliminary estimates of its accuracy for the detection of CIN 2 or worse. We recruited 1000 patients from screening and prevention clinics and 850 patients from colposcopy clinics at two comprehensive cancer centers and a community hospital. Optical spectroscopy was performed and 4864 biopsies were obtained from the sites measured, including abnormal and normal colposcopic areas. The gold standard was the histologic report of biopsies, read 2–3 times by histopathologists blinded to the cytologic, histopathologic, and spectroscopic results. We calculated sensitivities, specificities, receiver operating characteristic (ROC) curves, and areas under the ROC curves. We identified a cutpoint for an algorithm based on optical spectroscopy that yielded an estimated sensitivity of 1.00 [95% confidence interval (CI) = 0.92 – 1.00] and an estimated specificity of 0.71 [95% CI = 0.62 – 0.79] in a combined screening and diagnostic population. The positive and negative predictive values were 0.58 and 1.00, respectively. The area under the ROC curve was 0.85 (95% CI 0.81 – 0.89). The per-patient and per-site performance were similar in the diagnostic and poorer in the screening settings. Like colposcopy, the device performs best in a diagnostic population. Alternative statistical approaches demonstrate that the analysis is robust and that spectroscopy works as well as or slightly better than colposcopy for the detection of CIN 2 to cancer.

Keywords

sensitivity and specificity; diagnosis; early detection of cancer; uterine cervical neoplasms; cervical intraepithelial neoplasia

INTRODUCTION

Cervical cancer remains a major cause of morbidity and mortality in the developing world, where 85% of cancers arise.[1, 2] Identification of cancerous and precancerous lesions at earlier stages, when interventions are more likely to be effective, is critical for effective cancer control.[3−6] Recent advances in fiber-optic and semiconductor technologies have enabled the development of a new generation of inexpensive, miniature optical sensors that can probe the interaction of light with potentially cancerous tissue in real-time.[7, 8] Work to integrate this new technology with existing detection modalities and clinical screening efforts is ongoing.

Two other recent developments in cervical cancer diagnosis and prevention are also impacting disease detection. These are human papillomavirus (HPV) testing for screening and the HPV vaccine. The HPV test for screening is intended for the developing world.[9] There is preliminary evidence it may be more cost-effective than the Papanicolaou smear.[10] However, this test requires electricity, laboratory access, and effective follow-up after a

positive result. HPV vaccines face similar challenges. While we expect that such vaccines will reduce cervical neoplasia incidence by approximately 70% in the developed world,[11, 12] these results will likely take more than 20 years to achieve, and several hurdles must be overcome before this option is available in developing countries. Where the vaccination program is successful, fewer cases are likely to be discovered in screening and diagnostic settings. Consequently, cost-effective methods will be needed to address this reduced incidence.

Optical spectroscopy is a candidate technology to fill this role, particularly in areas lacking health care resources. The assessment of "emerging technologies" is a complicated process.[13] Figures 1A and 1B depict the current clinical paradigms for cervical screening and diagnosis and show what optical technologies could do to transform the process to one that occurs in real-time, is more automated, and is less energy dependent, i.e., battery powered.[14]

Following the Littenberg paradigm for the assessment of emerging technologies, we examined biologic plausibility, technical efficacy, clinical effectiveness, patient and provider satisfaction, and cost-effectiveness of optical spectroscopy.[13] Biologic plausibility studies demonstrated that 1) tissue fluorescence varied with age and menopausal status,[15] 2) changes in NADH, FAD, and collagen occurring in lesions in studies of fresh tissue were likely responsible for these changes in variation,[16⁻20] and 3) convolution and deconvolution of data from fluorophores measured in the laboratory and spectra from patients could be used to model fluorophores to resemble tissue and tissue to resemble fluorophore measurements.[17, 19, 21⁻29] Technical efficacy studies showed that these research grade devices were safe[28] but performed differently from each other; however, each device did not perform differently over the course of a day or several days.[30⁻35] We developed a quality assurance program and a database to hold the terabytes of data generated from patient measurements. We examined several aspects of clinical effectiveness to reassure ourselves that we were controlling for potential biases or confounders. We compared the screening and diagnostic populations we recruited to be certain that the screening patients were typical of low-risk patients and that the diagnostic patients had a higher number of risk factors for cervical cancer.[36, 37] The patients could be measured any time during the cycle except during menses.[38⁻40] The pathologists had good agreement.[41] Our evaluation of patient satisfaction showed that colposcopy-directed biopsies were more painful and made patients more anxious than spectroscopy or the Papanicolaou smear, which were equally uncomfortable.[42] We also studied provider satisfaction, and these findings are in a forthcoming manuscript.[43] To address the final concept of the Littenberg paradigm, we modeled the cost effectiveness of spectroscopy in a see-and-treat scenario and showed that, if we were able to achieve sensitivities of at least 84% and specificities of at least 76%, there was indeed a huge savings of health care dollars in "biopsies avoided" and in being able to "see-and-treat" more accurately with a loop electrical excision procedure.[44] We reviewed the literature carefully to learn from other investigators who have studied optical instruments for the cervix, noting the trade-off that may exist regarding device accuracy and sample size.[45, 46]

In this report, we present the use of fluorescence spectroscopy as an adjunct to colposcopy, one step in the multi-step process of automating and replacing existing clinical tests in the screening and diagnosis of cervical cancer and precancer.

MATERIAL AND METHODS

Specific Objectives

Our specific objective was to select an effective algorithm using optical spectroscopy and obtain preliminary estimates of its accuracy for the detection of cervical intraepithelial neoplasia (CIN) 2 or worse in both the screening and diagnostic settings. Because we intended to sample normal and abnormal sites, a secondary objective included the evaluation of colposcopy against the gold standard of doubly- or triply-read histopathologic biopsy.

Overview of Study Procedures

Women who were at least 18 years of age and not pregnant were eligible to participate in the study, which took place between October 1998 and November 2005. Patients were recruited either from a screening group of women who had no history of abnormal Papanicolaou smears, or a diagnostic group of women from colposcopy clinics who were referred with an abnormal Papanicolaou smear or previous treatment for CIN. Volunteers for the screening group were solicited from the community through advertising, media coverage, and participant word of mouth.[47] A research nurse described the study to eligible patients and obtained informed consent from those agreeing to participate. All patients were asked to complete an interview conducted by the research nurse that covered demographic variables and aspects of sexual behavior. These data were entered into a database but were not made available to providers.

Each patient provided a complete history and received a physical exam, a Papanicolaou smear, and colposcopic examination of the vulva, vagina, and cervix. A blood sample for the measurement of follicle stimulating hormone, estrogen, and progesterone was taken, and prevention recommendations (regarding tobacco use, sunscreen use, obtaining mammography, increasing calcium intake, and following American Cancer Society screening guidelines) were discussed with the patient. Two cervical smears were obtained using an Ayre's spatula and Cytobrush; the first sample was placed directly on a glass slide and fixed with fixative, and the second specimen was placed in Cytyc liquid-based medium for quantitative cytology and for HPV testing using the Hybrid Capture® II (HCII) test (Digene Corporation, Gaithersburg, MD). Two additional cytobrush specimens were obtained from the endocervical canal for detection of HPV DNA and mRNA by PCR.[48−50] (Table 1 depicts all technologies we evaluated in this clinical trial.) Results from these technology assessments have been published.[51−53] After examining the cervix with white light, acetic acid at 6% strength was placed on cotton balls against the cervix for two minutes. Following colposcopic examination using white and green light at 3.5X, 7X, and 15X magnification, but prior to biopsy, a fiber optic probe 5.1 mm in diameter was advanced through the speculum and placed in gentle contact with the cervix. Spectroscopic measurements were obtained from one or two normal cervical sites covered with squamous epithelium and, when visible, one colposcopically normal cervical site with columnar epithelium. If abnormalities were present and visible, measurements were taken from two colposcopically-abnormal sites. Thus, all patients had sampling of both abnormal and normal areas if colposcopic abnormalities were present.

The center of the probe, which is about the size of a pencil, has both the light emitting and light detecting systems. The probe interrogated an area 2 mm in diameter on the cervix and left a circular impression. Following the spectroscopic measurements, the biopsy was taken from the center of the circular impression so that, as much as possible, the biopsy site was the same as the spectroscopic site. The biopsies were obtained with forceps, yielding specimens that were 2 mm long by 1 mm wide by 1 mm deep. The biopsies were fixed in formalin and submitted for permanent section. Sections four microns thick were stained with

hematoxylin and eosin for routine reading in patient care and Feulgen staining was performed for research on the quantitative measurements of the epithelium.

Colposcopy was performed by five gynecologic oncologists, two generalist obstetrician gynecologists, one family practitioner, and six nurse practitioners specialized in colposcopy. Each colposcopy was recorded with a drawing detailing the cervical areas of squamous and columnar epithelium, the squamo-columnar junction, the transformation zone, the presence of white or erythematous areas, and the presence of vascular atypia (fine or coarse punctation, loose or tile-like mosaicism, and atypical vessels). A study of inter-provider and intra-provider variability in colposcopic technique showed excellent to outstanding agreement (by kappa statistic) on the findings of aceto-whitening, erythema, fine and coarse punctation, mosaiform and mosaic vasculature, and atypical vessels. [data not shown]

Study Population

The study protocol was reviewed and approved by the Institutional Review Boards at The University of Texas M. D. Anderson Cancer Center (MDACC), The University of Texas Health Science Center at Houston, the Harris County Hospital District for the Lyndon Baines Johnson (LBJ) General Hospital, The University of Texas at Austin, Rice University, and the British Columbia Cancer Agency (BCCA). The study was carried out at three clinical sites including The University of Texas M. D. Anderson Cancer Center (Houston, TX), the Lyndon Baines Johnson General Hospital (Houston, TX), and the British Columbia Cancer Agency (Vancouver, BC). At MDACC, patients were self-referred or referred by private physicians. At the LBJ General Hospital, patients were referred by county health department clinics. At the BCCA, patients were referred from the network of physicians in British Columbia.

Laboratory Tests

All routinely stained cytology and pathology specimens were submitted for diagnosis by an experienced pathologist with specialization in gynecology on call for the day and blinded to the results of the spectroscopy. Cytologic smears and histopathologic sections were initially reviewed clinically by the cytologist or pathologist at the respective cancer center institutions. The pool of physicians reading samples included study cytologists and histopathologists. Both the Bethesda System and the World Health Organization (WHO) classification were used for cytology samples; the WHO classification was used for biopsies. All specimens were independently reviewed a second time by a study cytopathologist (JM, GS) or a study histopathologist (DVN, AM), who was blinded to the results of the first review, to colposcopy, and to all other clinical tests including the spectroscopy When diagnoses were discrepant, the specimens were reviewed a third time by the study cytopathologist or histopathologist to resolve the discrepancy. In order to determine whether any bias was introduced by the histopathology being read at the BCCA or MDACC, a study was conducted to be certain that the readings of the biopsies at both institutions were similar. The kappa statistics reported in Malpica et al. for inter-and intra-rater reliability for the expert group of pathologists were in the substantial and almost perfect ranges for the histopathology review of study samples with high-grade lesions. The histopathologic consensus diagnosis was used as the gold standard for the trial.41

HPV typing was performed using the Food and Drug Administration-approved HCII test. The test was performed by Laboratory Corporation of America®, an external clinical laboratory. The HPV HCII tests were performed according to the manufacturer's recommended protocol.

Each patient's menstrual history, hormone use, follicle stimulating hormone level, estrogen level, and progesterone level were reviewed with a reproductive endocrinologist (SNE). Menopausal status was classified into three categories: pre-, peri-, and post-menopausal based on a review of each patient's current and past menstrual history and consistent with the laboratory results.

Spectroscopic Measurements

We developed four research-grade, fiber-optic spectrometers to measure fluorescence and reflectance spectra from cervical tissue *in vivo* at three clinical locations over the five years of the study. In fluorescence spectroscopy, the tissue is illuminated by light of a particular wavelength called the excitation wavelength, and the light is re-emitted at a longer wavelength called the emission wavelength. The devices measured fluorescence emission spectra at 16 different excitation wavelengths ranging from 330 nm to 480 nm and collected at a range of emission wavelengths from 360 nm to 800 nm. These data are referred to as an excitation-emission matrix.

Each measurement took approximately one minute or less and the total light exposure was less than the American National Standards Institute standard for tissue exposure to light.[28] Details of the devices used during the trial can be found in Freeberg et al.[34] Of note, there were two generations of the device used during the seven years of the trial. The second-generation device was cheaper to construct and took measurements more quickly than its predecessor. The details of the processing of the data from the devices can be found in Marin et al.[35]

All spectra were reviewed three times, beginning with four independent investigators blinded to the histology. Spectra were independently reviewed a second and third time by two medical physicists who were blinded to the histology data. Spectra were excluded for several indications: it was inferred that the probe slipped during the measurement, blood obscured the site, a chipped probe lead to altered reflectance measurements, autofluorescence from the probe was present, fluorescence or reflectance spectra were saturated, the signal was weak, or the device failed during a measurement. Spectra were discarded if both reviewers agreed that at least one of the aforementioned abnormalities was present. For the 4864 fluorescence and reflectance spectra, the two medical physicists disagreed in only one situation, thus demonstrating outstanding agreement. Those spectra that were rejected correlate to device breakdowns and were not related to any specific patient characteristics. An ongoing analysis of the differences in accepted and rejected spectra helped us develop quality assurance software that eventually allowed us to evaluate in real time whether spectra were acceptable or required repetition. [data not shown]

Development of an Algorithm for Optical Spectroscopy

The objective was to find the most effective classification algorithm for detecting "disease" based on the spectroscopic data and other data that were available at the same time. We evaluated numerous classification algorithms and selected the one which had the best estimate of performance. Essentially, it was a contest between all of the algorithms, with the winner having the highest specificity using an 80% sensitivity. The classification algorithms we tried were Bayesian variable selection, naïve Bayes, logistic regression with forward and backward variable selection, random forests, classification trees, neural nets, penalized logistic regression, linear discriminant analysis, nearest neighbor, linear support vector machines, kernel support vector machines, and others. There were several complications that needed to be addressed.[54−56]

One issue was whether the final output of the algorithm would be a prediction of whether the site of the measurement was diseased (per-site classification), or if a prediction needed to be made regarding whether the patient had disease anywhere in her cervix. All of the algorithms use the per-site data, so we decided to select the winner of our algorithm competition by using this data. Once the best algorithm was selected, we reported per-patient results based on assigning to each patient the worst biopsy from among all sites measured for that patient. The actual outcome at a site was of course the histologic reading of the biopsy from that site. The actual per-patient outcome was the worst histologic reading for all biopsies from the patient.

We chose to simplify the classification by dichotomizing the outcome as either "diseased" or "non-diseased." Our goal was to detect patients (or sites within patients) with high-grade squamous intraepithelial lesions or worse. Thus, "disease" (when applied to a measurement at a single site or to a patient) means "CIN 2 or worse," including a histology reading of CIN 2, CIN 3, CIS, and invasive squamous cancer. Patients (or sites) were classified as "non-diseased" if their histology reading was normal, atypia, inflammation, HPV-related changes, or CIN 1.

Most classification algorithms have tuning parameters that must be estimated from the data. The process of obtaining the parameters leads to biased performance estimation on the data on which it was trained. For example, some algorithms have many more tuning parameters than others; this makes them more susceptible to overtraining, wherein their performance looks good on the data on which they are trained, yet they perform poorly on new data. To obtain an unbiased estimate of the performance of an algorithm, the performance needs to be estimated on an independent set of data. Therefore we split the data into a training set and a test set, with 70% and 30% of available data in each set, respectively. We optimized the algorithm performance on the training data before applying it to the test data.

Classification algorithms require a sufficient number of cases of disease and non-disease. We were simultaneously evaluating potential screening devices in this trial, so spectroscopic measurements were taken on a screening population as well as the intended diagnostic population. Because we had access to the additional data, and due to the data's high-dimensionality, we used it to assist in the development of an algorithm for the diagnostic device.

Because we were considering multiple algorithms and were going to select the one with the best classification accuracy on the training data, we needed a reliable estimate of accuracy based only on the training data. Rather than splitting the training data into two sets (a training set and a validation set), we conducted 5-fold cross-validation within the training set for each algorithm. This allowed the algorithm to be trained on a larger subset and its performance estimated from the entire training set. The training set (70% of the total data set) was split into five subsets. Each of the algorithms was trained five times while omitting one of the subsets; thus 80% of the training data was used to train the algorithm each time. The algorithm was then applied to the remaining 20% of the training data. This produced a score for each of the remaining observations; the five sets of scores were then combined to evaluate the overall performance of the algorithm on the training set. Once we selected a final algorithm, it was trained on the entire training set and applied to the test set (the remaining 30% of the entire data) to obtain unbiased estimates of the sensitivity and specificity of the algorithm.

In creating the training set, test set, and five subsets of the training set, we employed the following sampling strategy. We assigned all measurements for a given patient to the same subset. We approximately stratified the sampling that was used for creating the subsets by

the demographic and biologic variables we thought might be important: menopausal status, the presence of high-grade histology (CIN 2 or worse), and the date of measurement (to account for possible time trends). The data within each stratum were ordered by diagnosis ("≥ CIN 2" or "< CIN 2"), then by menopausal status within diagnosis, and then by study identification number, for which the highest digits designate population (screening or diagnostic), the second highest digits designate the clinic, and the remaining digits give the order of recruitment within the clinic. The algorithm took successive blocks of ten patients and randomly assigned seven patients from each block to the training set and three patients from each block to the test set. The patients in the training set (in the same order) were separated into blocks of five patients and were then randomly assigned to each of the five subsets of training data (one patient from the block to each of the five subsets). This method guaranteed that the training and test sets would be approximately balanced for each of the variables known to affect the spectra in previous studies of tissue biology.[56−58]

The per-site data are complex high-dimensional arrays; therefore dimension reduction was needed to analyze them. We primarily used principal components analysis for the purpose of dimension reduction, although we also investigated fast Fourier transforms, B-splines, and an inverse model.[56] We tried various dimension reduction methods as inputs for each of the algorithms. Two strategies were employed for principal component analysis: concatenating all of the measured intensities into a single vector, or using principal components of the emission spectra for each excitation wavelength separately. The principal components were computed from the covariance matrix of the emission spectra from the individual excitation wavelengths. We kept only the principal components that accounted for at least 95% of the total variation. For the principal components computed for individual excitation wavelengths, this reduced the dimension of the measurements to between 1 and 3 principal components per excitation wavelength.

We developed the classification algorithms using the features obtained from the methods of dimension reduction, the colposcopic impression, and the biographical variables: age, menopausal status (pre-menopausal, peri-menopausal, and post-menopausal), hormone use (use of any of oral contraceptive pills, hormone replacement therapy, or depoprovera), colposcopic tissue type (columnar or squamous), and bleeding (during measurement). We included these variables because this information was available at the time of the clinical visit and because several studies have shown that biological variables influence spectroscopy measurements. We provide details on a logistic regression approach, as this demonstrated the most effective performance of all data reduction methods and all algorithms.[53−55] The variable selection was applied to the biographical variables (menopausal status, hormonal use, colposcopic tissue type, and age), the colposcopic impression, and to the principal components of the spectroscopic data computed for individual excitation wavelengths. We then fit a logistic regression model to the principal components and biographical variables and used Akaike's information criterion to perform a backwards stepwise selection of variables.

We used receiver operator characteristic (ROC) analysis to summarize the performance of the selected algorithm. To account for differences in the prevalence of the diagnostic and screening populations, covariate-adjusted ROC curves were computed.[55] These can be viewed as a weighted average of the screening and diagnostic population ROC curves. To compare the covariate-adjusted ROC curves, bootstrap confidence intervals were calculated at a fixed sensitivity of 80%. We wished to compare the specificities of colposcopy and spectroscopy when the sensitivity was set at 0.80. A challenge is that some patients in the study were evaluated by both technologies while others were evaluated by colposcopy only (due to drop out, faulty readings, etc.) The *p* value was computed by modeling the subjects who received both measurements using a multinomial model and the subjects who received

colposcopy using a binomial model. We were then able to compute a likelihood ratio chi-square to determine whether the differences were statistically significant.

Statistical analysis was performed using the statistical packages R version 2.6.2 (R Foundation for Statistical Computing, Vienna, Austria), Matlab® (The MathWorks, Inc., Natick, MA), Mathematica (Wolfram Research, Inc., Champaign, IL), and Stata Statistical Software Release 10.1 (StataCorp LP, College Station, TX). Exact binomial confidence intervals were calculated for sensitivity and specificity. Statistical significance was set at 0.05.

RESULTS

Figure 2 shows a flow diagram detailing how the initial enrollment resulted in the final sample size after data quality control. The demographic and diagnostic information concerning the 1850 patients can be found in Tables 2 and 3. The 1850 patients yielded 4686 biopsies and spectra. Although the total number of patients entering our clinical trials was 1850, we had usable spectroscopy and biopsies on 735 diagnostic patients (115 eliminated) and on 707 screening patients (293 eliminated). Approximately 30% of spectra, including repeated measures, were judged inadequate for analysis during the quality assurance review.

As noted, the patients ranged in age from 18 to 85 years; the mean age for the entire sample of patients was 39 years. The majority of patients were born in the United States or Canada, had a college education, and were married. We were successful in recruiting a racially and ethnically diverse sample of women, reflective of the populations of Houston, Texas, and Vancouver, British Columbia. There were no adverse events throughout the course of the trial; specifically, no patients returned for bleeding or infection after the measurements and cervical biopsies. All of the Canadian patients have been seen again by their physicians and long-term follow-up shows no adverse effects of the spectroscopy on subsequent Papanicolaou smears. Approximately 70% of the diagnostic patients were followed for two years, and no adverse effects of the spectroscopy were noted.

According to tissue histology, 201 of 735 patients had biopsies that had CIN 2, CIN 3, CIS, or cancer in the diagnostic group (27%) compared to 12 of 707 patients (2%) in the screening group. HCII test results were high-risk type positive in 46% of patients in the diagnostic group, compared to 10% of the screening group.

As stated earlier, the best algorithm results as estimated from five-fold cross-validation with training data were obtained using logistic regression with variable selection applied to the biographical variables (menopausal status, hormonal use, colposcopic tissue type, and age), colposcopic call, and principal components of the fluorescence data only. The final variables in the model included all four biographical variables, the colposcopic impression, and 23 of the principal component variables from the spectroscopic data. Not only did this algorithm perform the best as determined by ROC curve analysis, but the algorithm would be easy to implement in the diagnostic device.

Figure 3 presents boxplots of the predicted scores from the classification algorithm. Using the point on the ROC curve closest to perfect classification (0.80 sensitivity, 0.84 specificity), we identified a cut point of 0.221 on the spectroscopy score in the training set. This yielded a by-patient sensitivity (correctly identifying those patients who had histology reading CIN 2 or worse) of 1.00 [one-sided 95% confidence interval (CI) = 0.92 1.00] and a specificity (correctly identifying those patients who had histology reading CIN 1 or better) of 0.71 [95% CI = 0.62 – 0.79] in the test set using the second-generation device. Given the prevalence of 10% for CIN 2 or worse in our combined screening and diagnostic data, the positive predictive value was 0.58 and the negative predictive value was 1.00.

The purpose of showing the data in a ROC curve format is to demonstrate how the device performs over a range of sensitivity and specificity pairs, emphasizing the trade-off between sensitivity and specificity. Figures 4A and 4B show the diagnostic classification algorithm applied to the training and test sets, respectively. The area under the ROC curve (AUC) for spectroscopy when applied to the test set for the second-generation device was 0.85 [95% CI = 0.81 – 0.89].

Figure 5 shows the results of the ROC curve analysis for both the screening and diagnostic populations. Figure 5A shows the results of a per-patient analysis and 5B shows the results of a per-site analysis. The per-patient AUC for the diagnostic study in Figure 5A is 0.83 (95% CI 0.77 – 0.88), and for the screening study the AUC is 0.58 (95% CI 0.47 – 0.69). The per-site AUC in Figure 5B for the diagnostic study in Figure 5A is 0.78 (95% CI 0.73 – 0.83), and for the screening study the AUC is 0.60 (95% CI 0.46 – 0.74). We assume, as with colposcopy, that the poor performance in the screening study is due to the lower prevalence of disease. The per-patient and per-site analyses are quite similar.

To provide some basis for showing that spectroscopy can add value even when colposcopically-directed, we performed two further analyses shown in Figures 6A and 6B. In the first analysis, we controlled for the differences in prevalence between the screening and diagnostic trials by using whether the patient was enrolled in the screening or diagnostic study as a covariate in the analysis. This allowed us to compare spectroscopy in the combined population in the test set, allowing for the lower prevalence in the screening setting, to colposcopy in the whole data set. In the second analysis, we compared diagnostic spectroscopy with diagnostic colposcopy, excluding the screening patients. This analysis was intended to show how colposcopy and spectroscopy perform in the intended diagnostic setting. These analyses model what we hope to find in a randomized trial of colposcopy versus colposcopy plus spectroscopy. Figure 6A shows covariate-adjusted ROC curves comparing spectroscopy with colposcopy. At a sensitivity of 0.80, the specificity was 0.73 (CI = 0.64 – 0.82) for spectroscopy and 0.46 (CI = 0.42 – 0.5) for colposcopy, a statistically significant difference ($p < 0.05$). The classification performance of the tests in the high-sensitivity region of the ROC curve decreases some, mostly due to the expected small number of cases in the screening population. For Figure 6B, we compared the accuracy of spectroscopy in the test set to the accuracy of colposcopy in the whole data set. We chose to use the whole data set for colposcopy because it did not require training of an algorithm; therefore we could use more data to obtain a more precise unbiased estimate of its accuracy. Using the whole data for spectroscopy would have biased our spectroscopy accuracy estimates upwards because the training data was used to select the best model. Figure 6B shows ROC curves comparing diagnostic spectroscopy with diagnostic colposcopy. At a sensitivity of 0.80, the specificity was 0.76 (CI = 0.69, 0.82) for spectroscopy and 0.68 (CI = 0.64, 0.72) for colposcopy. Using the likelihood ratio chi-square test [as described in the Material and Methods section], we determined that the difference in specificity between spectroscopy and colposcopy was not statistically significant ($p = 0.5$). For both the diagnostic and screening populations, we observed an increase in accuracy similar to that in the covariate-adjusted ROC curves. The colposcopy-naïve spectroscopy and colposcopy ROC curves are very similar to each other. We recognize that the colposcopy-naïve spectroscopy described here is, in fact, colposcopically-directed. However, we included these ROC curves to provide some basis for comparison. The areas under the covariate-adjusted ROC curves for diagnostic spectroscopy and diagnostic colposcopy were 0.77 and 0.78, respectively.

DISCUSSION

We developed an algorithm for point-probe optical spectroscopy that yielded operating characteristics with reasonable performance and that has the potential for use in real time. The data show that in a diagnostic setting, research-grade point-probe devices using colposcopically-directed optical spectroscopy perform similarly to colposcopy in expert hands. The role of this technology was to be an adjunct to colposcopy so that one could avoid biopsies of inflammatory lesions and see and treat with confidence that disease would be in the specimen. We continue to develop a multispectral digital colposcope (MDC) for screening and for eventual combination with the probe technology. The MDC has been through two pilot trials, each of which demonstrated sensitivities of 85% and specificities of 90% in automated algorithms on few patients. We will begin testing a combined device based on this work.

In the original statistical plan of the protocol, we used a sensitivity of 84% and a specificity of 76% as parameters to calculate the sample size. These data were based on work using three wavelengths of light in 104 patients, many of whom had high-grade squamous intraepithelial lesions. We estimated that 200 patients with high-grade squamous intraepithelial lesions would allow us to study the diagnostic algorithm, and we hoped to stratify the diagnostic group for several demographic categories: premenopausal on oral contraceptive pills (OCP), premenopausal not on OCP, postmenopausal on hormone replacement therapy (HRT), and postmenopausal not on HRT. Then the results of the Women's Health Study were published and fewer women took OCP or HRT. In the end, we were able to examine these demographic variables in the analysis, but 200 patients alone were too few to develop a classifier in the multidimensional data.

The high-dimensional data we obtained required cross-validation for the algorithm development. We learned that developing a classifier would force us to combine the screening and diagnostic populations. We calculated that the development of a classifier in a screening population with a point probe would require 16,000 patients, for which funding would be a near impossible task. We also thought of the MDC as the instrument that would accompany the point probe in the screening setting. This was true of our studies of quantitative cytology, for which we also needed a large dataset to develop a classifier.

While we expected the per-site analysis to yield a higher AUC and higher sensitivities and specificities than the per-patient analysis, it did not. This may be because the spectroscopy assesses the epithelial/stromal interaction. Thus, one would expect that if any area of the cervix has high-grade dysplasia, there is a field effect such that the entire cervical epithelial-stromal interface is different from that of a patient with no disease. Biomarker data supports that normal areas of a diseased cervix are not as genetically stable as normal areas in a normal cervix. We are actively investigating the stromal biology of these types of lesions.

How do our results compare to the literature? Table 4 shows the spectroscopic approach and modality as well as the sensitivity and specificity obtained in the trials.[22, 25, 59–82] Our results compare favorably with those of other investigators.

The main strengths of this study are that 1) each patient had several biopsies that underwent multiple blinded reviews and thus provided an excellent gold standard upon which to judge all the technologies under study, 2) few registration problems occurred with the biopsied tissues, 3) robust analysis of the multiple algorithms yielded similar results using different approaches for both data reduction and data analysis, and 4) attention was paid to all aspects of technology assessment. Previous trials by Alvarez[82] and DeSantis[77] used multispectral technologies to view the whole cervix and compared the multispectral readings to both biopsies and areas of loop electrosurgical excision procedure specimens from the cervix.[77,]

82 However, registration, or linking, of the optical image to the area of histopathologic reading was difficult in their study designs. In our studies, each 2 mm area that was measured was biopsied, thus registration was not an issue. Further, because we had spectroscopic measurements from both colposcopically-positive (if any were present) and colposcopically-negative sites on all patients, we were able to find false positive and false negative colposcopic lesions, eliminating the problem of verification bias and something that was not done in other studies.

The weaknesses of the study are 1) the number of scrapes to the cervix prior to measurement, 2) the probe placement possibly not being precisely over the area biopsied, 3) the discarding of approximately 30% of spectrographs, and 4) the use of a cut off point for the classifier. The scrapes may have affected the epithelium, but we believe we sampled the epithelial stromal interface; however, in future studies of the probe we will not take any scrapes. Also, the probe placement may never be precisely over the area imaged. However, the microenvironment in a region may be similar over the microns that are measured. Discarding spectrographs comes with the territory of studying emerging technologies, as does the development of first- and second-generation devices. We resisted changes to the devices but instead quantified how the devices performed differently for our own understanding of how changes might impact the data. Finally, choosing a cut point is a complicated process. We have not focused on this in this paper, but the cut point changes the trade-off between the number of false negatives and the number of false positives. This will be the subject of a future cost-effectiveness analysis. Because we subjected the algorithmic analyses to a robust number of methods of data reduction and then analysis using the cut point of CIN 2 and above, we are certain the results reflect what is contained in the dataset.

The endocervical canal presents challenges for two reasons: first, there may be squamous lesions high in the canal, and second, there is an increasing incidence of adenocarcinoma in situ and adenocarcinoma arising in the columnar epithelium. In general, patients referred for colposcopy with abnormal Papanicolaou smears (diagnostic patients) usually receive an endocervical curettage (scraping of the endocervix) along with their cervical biopsies, whereas those patients with a history of normal Papanicolaou smears (screening patients) receive only a cytologic evaluation of the endocervical canal. Many gynecologic oncologists have found an invasive cancer or lesions of adenocarcinoma in situ or adenocarcinoma in a patient for whom the only abnormality was a positive endocervical curettage or suspicious endocervical cytology. This is a particular concern in older patients, as the squamo-columnar junction moves farther up the endocervical canal with age and endocervical lesions could be missed without a vigilant approach. At present, we would still recommend an evaluation of the endocervical canal apart from or in addition to optical spectroscopy. This is a major limitation of this spectroscopy, but also of visual inspection of the cervix with acetic acid and other existing screening devices. We are researching ways to use spectroscopy to detect abnormalities in the canal and perhaps will have other solutions in the future. For now, the Papanicolaou smear and/or endocervical curettage remain critically important in the clinical evaluation of patients to diagnose all cervical cancers.

The device developed in this study would be an adjunct to colposcopy and probably would not be commercially viable by itself. As we found, a point probe cannot be used for screening; for that purpose we are developing the MDC,[83⁻86] a device that sees the whole cervix. The combined MDC and point probe would be made for the developed world, where it is important to save heath care dollars by eliminating unnecessary biopsies and treating only those patients with CIN 2 to cancer. For the developing world, the lack of sufficient electricity led us to develop a portable battery-powered device that we hope to test with our collaborators in developing nations.

Acknowledgments

References

1. Parkin D, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin. 2005; 55:74–108. [PubMed: 15761078]

2. Sherris J, Herdman C, Elias C. Cervical cancer in the developing world. West J Med. 2001; 175:231–33. [PubMed: 11577044]

3. Anderson GH, Boyes DA, Benedet JL, Le Riche JC, Matisic JP, Suen KC, Worth AJ, Millner A, Bennett OM. Organisation and results of the cervical cytology screening programme in British Columbia, 1955–85. Br Med J (Clin Res Ed). 1988; 296:975–8.

4. Benedet JL, Anderson GH, Matisic JP, Miller DM. A quality-control program for colposcopic practice. Obstet Gynecol. 1991; 78:872–5. [PubMed: 1923214]

5. American Cancer Society Guidelines for the Early Detection of Cancer. Available at: http://www.cancer.org/docroot/ped/content/ped_2_3x_acs_cancer_detection_guidelines_36.asp

6. Kurman R, Henson D, Herbst A, Noller K, Schiffman M. Interim guidelines for management of abnormal cervical cytology. The 1992 National Cancer Institute Workshop. JAMA. 1994; 271:1866–9. [PubMed: 8196145]

7. Sokolov K, Follen M, Richards-Kortum R. Optical spectroscopy for detection of neoplasia. Curr Opin Chem Biol. 2002; 6:651–8. [PubMed: 12413550]

8. Drezek RA, Richards-Kortum R, Brewer MA, Feld MS, Pitris C, Ferenczy A, Faupel ML, Follen M. Optical imaging of the cervix. Cancer. 2003; 98:2015–27. [PubMed: 14603538]

9. Sankaranarayanan R, Nene BM, Shastri SS, Jayant K, Muwonge R, Budukh AM, Hingmire S, Malvi SG, Thorat R, Kothari A, Chinoy R, Kelkar R, et al. HPV screening for cervical cancer in rural India. N Engl J Med. 2009; 360:1385–94. [PubMed: 19339719]

10. Qiao YL, Sellors JW, Eder PS, Bao YP, Lim JM, Zhao FH, Weigl B, Zhang WH, Peck RB, Li L, Chen F, Pan QJ, et al. A new HPV-DNA test for cervical-cancer screening in developing regions: a cross-sectional study of clinical accuracy in rural China. Lancet Oncol. 2008; 9:929–36. [PubMed: 18805733]

11. Muñoz N, Bosch F, Castellsagué X, Díaz M, de Sanjose S, Hammouda D, Shah K, Meijer C. Against which human papillomavirus types shall we vaccinate and screen? The international perspective. Int J Cancer. 2004; 111:278–85. [PubMed: 15197783]

12. Franco EL, Harper DM. Vaccination against human papillomavirus infection: a new paradigm in cervical cancer control. Vaccine. 2005; 23:2388–94. [PubMed: 15755633]

13. Littenberg B. Technology assessment in medicine. Acad Med. 1992; 67:424–8. [PubMed: 1616553]

14. Wright TC Jr, Massad LS, Dunton CJ, Spitzer M, Wilkinson EJ, Solomon D. 2006 consensus guidelines for the management of women with cervical intraepithelial neoplasia or adenocarcinoma in situ. Am J Obstet Gynecol. 2007; 197:340–5. [PubMed: 17904956]

15. Brookner C, Utzinger U, Follen M, Richards-Kortum R, Cox D, Atkinson EN. Effects of biographical variables on cervical fluorescence emission spectra. J Biomed Opt. 2003; 8:479–83. [PubMed: 12880354]

16. Brookner C, Follen M, Boiko I, Galvan J, Thomsen S, Malpica A, Suzuki S, Lotan R, Richards-Kortum R. Autofluorescence patterns in short-term cultures of normal cervical tissue. Photochem Photobiol. 2000; 71:730–36. [PubMed: 10857369]

17. Ramanujam N, Richards-Kortum R, Thomsen S, Mahadevan-Jansen A, Follen M, Chance B. Low temperature fluorescence imaging of freeze-trapped human cervical tissues. Opt Express. 2001; 8:335–43. [PubMed: 19417824]

18. Drezek R, Brookner C, Pavlova I, Bioko I, Malpica A, Lotan R, Follen M, Richards-Kortum R. Autofluorescence microscopy of fresh cervical-tissue sections reveals alterations in tissue biochmistry with dysplasia. Photochem Photobiol. 2001; 73:636–41. [PubMed: 11421069]

19. Drezek R, Sokolov K, Utzinger U, Boiko I, Malpica A, Lotan R, Follen M, Richards-Kortum R. Understanding the contributions of NADH and collagen to cervical tissue fluorescence spectra: modeling, measurements, and implication. J Biomed Opt. 2001; 6:385–96. [PubMed: 11728196]

20. Arifler D, Pavlova I, Gillenwater A, Richards-Kortum R. Light scattering from collagen fiber networks: micro-optical properties of normal and neoplastic stroma. Biophys J. 2007; 92:3260–74. [PubMed: 17307834]

21. Drezek R, Guillaud M, Collier T, Boiko I, Malpica A, Macaulay C, Follen M, Richards-Kortum R. Light scattering from cervical cells throughout neoplastic progression: influence of nuclear morphology, DNA content, and chromatin texture. J Biomed Opt. 2003; 8:7–16. [PubMed: 12542374]

22. Chang SK, Follen M, Malpica A, Utzinger U, Staerkel GA, Cox D. Optimal excitation wavelengths for discrimination of cervical neoplasia. IEEE Trans Biomed Eng. 2002; 49:1102–11. [PubMed: 12374334]

23. Pavlova I, Sokolov K, Drezek R, Malpica A, Follen M, Richards-Kortum R. Microanatomical and biochemical origins of normal and precancerous cervical autofluorescence using laser-scanning fluorescence confocal microscopy. Photochem Photobiol. 2003; 77:550–5. [PubMed: 12812299]

24. Chang SK, Arifler D, Drezek R, Follen M, Richards-Kortum R. Analytical model to describe fluorescence spectra of normal and preneoplastic epithelial tissue: comparison with Monte Carlo simulations and clinical measurements. J Biomed Opt. 2004; 9:511–22. [PubMed: 15189089]

25. Chang SK, Mirabal YN, Atkinson EN, Cox D, Malpica A, Follen M, Richards-Kortum R. Combined reflectance and fluorescence spectroscopy for in vivo detection of cervical pre-cancer. J Biomed Opt. 2005; 10:024031. [PubMed: 15910104]

26. Arifler D, MacAulay C, Follen M, Richards-Kortum R. Spatially resolved reflectance spectroscopy for diagnosis of cervical precancer: Monte Carlo modeling and comparison to clinical measurements. J Biomed Opt. 2006; 11:064027. [PubMed: 17212550]

27. Chang SK, Marin N, Follen M, Richards-Kortum R. Model-based analysis of clinical fluorescence spectroscopy for in vivo detection of cervical intraepithelial dysplasia. J Biomed Opt. 2006; 11:024008. [PubMed: 16674198]

28. Brookner C, Agrawal A, Trujillo E, Mitchell M, Richards-Kortum R. Safety analysis: relative risks of ultraviolet exposure from fluorescence spectroscopy and colposcopy are comparable. Photochem Photobiol. 1997; 65:1020–5. [PubMed: 9188282]

29. Lee JS, Shuhatovich O, Price R, Pikkula B, Follen M, McKinnon N, Macaulay C, Knight B, Richards-Kortum R, Cox DD. Design and preliminary analysis of a study to assess intra-device and inter-device variability of fluorescence spectroscopy instruments for detecting cervical neoplasia. Gynecol Oncol. 2005; 99:S98–111. [PubMed: 16188298]

30. Pikkula BM, Shuhatovich O, Price RL, Serachitopol DM, Follen M, McKinnon N, MacAulay C, Richards-Kortum R, Lee JS, Atkinson EN, Cox DD. Instrumentation as a source of variability in the application of fluorescence spectroscopic devices for detecting cervical neoplasia. J Biomed Opt. 2007; 12:034014. [PubMed: 17614722]

31. Pikkula B, Serachitopol D, MacAulay C, Mackinnon N, Lee JS, Cox D, Atkinson E, Follen M, Richards-Kortum R. Multicenter clinical trials of in vivo fluorescence: are the measurements equivalent? Proc SPIE. 2007; 6430:64301Q.

32. Nath A, Rivoire K, Chang S, Cox D, Atkinson EN, Richards-Kortum R. Effect of probe pressure on cervical fluorescence spectroscopy measurements. J Biomed Opt. 2004; 9:523–33. [PubMed: 15189090]

33. Rivoire K, Nath A, Cox D, Atkinson EN, Richards-Kortum R, Follen M. The effects of repeated spectroscopic pressure measurements on fluorescence intensity in the cervix. Am J Obstet Gynecol. 2004; 191:1606–17. [PubMed: 15547531]

34. Freeberg J, Serachitopol D, McKinnon N, Price R, Atkinson E, Cox D, MacAulay C, Richards-Kortum R, Follen M, Pikkula B. Fluorescence and reflectance device variability throughout the progression of a phase II clinical trial to detect and screen for cervical neoplasia using a fiber optic probe. J Biomed Opt. 2007; 12:034015. [PubMed: 17614723]

35. Marin N, MacKinnon N, MacAulay C, Chang SK, Atkinson EN, Cox D, Serachitopol D, Pikkula B, Follen M, Richards-Kortum R. Calibration standards for multicenter clinical trials of fluorescence spectroscopy for in vivo diagnosis. J Biomed Opt. 2006; 11:014010. [PubMed: 16526887]

36. Pham B, Rhodes H, Milbourne A, Adler-Storthz K, Follen M, Scheurer ME. Epidemiologic differentiation of diagnostic and screening populations for the assessment of cervical dysplasia using optical technologies. DrexelMed J. 2010 In press.

37. Pham B, Earle N, Rabel K, Follen M, Scheurer ME. Maximizing the diversity of participants in a phase II clinical trial of optical technologies to detect cervical neoplasia. Gynecol Oncol. 2007; 107:S208–14. [PubMed: 17822750]

38. Chang SK, Dawood M, Staerkel G, Utzinger U, Atkinson EN, Richards-Kortum R, Follen M. Fluorescence spectroscopy for cervical precancer detection: is there variance across the menstrual cycle? J Biomed Opt. 2002; 7:595–602. [PubMed: 12421126]

39. Cox D, Chang SK, Dawood M, Staerkel G, Utzinger U, Richards-Kortum R, Follen M. Detecting the signal of the menstrual cycle in fluorescence spectroscopy of the cervix. Appl Spectrosc. 2003; 57:67–72. [PubMed: 14610938]

40. MacAulay C, Richards-Kortum R, Utzinger U, Fedyk A, Atkinson E, Cox D, Follen M. Variation of fluorescence spectroscopy during the menstrual cycle. Opt Express. 2002; 10:493–504. [PubMed: 19436387]

41. Malpica A, Matisic JP, Niekirk DV, Crum CP, Staerkel GA, Yamal JM, Guillaud MH, Cox DD, Atkinson EN, Adler-Storthz K, Poulin NM, Macaulay CA, et al. Kappa statistics to measure interrater and intrarater agreement for 1790 cervical biopsy specimens among twelve pathologists: qualitative histopathologic analysis and methodologic issues. Gynecol Oncol. 2005; 99:S38–52. [PubMed: 16183106]

42. Basen-Engquist K, Shinn EH, Warneke C, de Moor C, Le T, Richards-Kortum R, Follen M. Patient distress and satisfaction with optical spectroscopy in cervical dysplasia detection. Am J Obstet Gynecol. 2003; 189:1136–42. [PubMed: 14586367]

43. Shinn EH. Provider satisfaction regarding cervical fluorescence spectroscopy. DrexelMed J. 2010 In press.

44. Cantor SB, Mitchell MF, Tortolero-Luna G, Bratka CS, Bodurka DC, Richards-Kortum R. Cost-effectiveness analysis of diagnosis and management of cervical squamous intraepithelial lesions. Obstet Gynecol. 1998; 91:270–7. [PubMed: 9469288]

45. Cárdenas-Turanzas M, Freeberg JA, Benedet JL, Atkinson EN, Cox DD, Richards-Kortum R, MacAulay C, Follen M, Cantor SB. The clinical effectiveness of optical spectroscopy for the in vivo diagnosis of cervical intraepithelial neoplasia: where are we? Gynecol Oncol. 2007; 107:S138–46. [PubMed: 17908588]

46. Freeberg JA, Benedet JL, MacAulay C, West LA, Follen M. The performance of fluorescence and reflectance spectroscopy for the in vivo diagnosis of cervical neoplasia; point probe versus multispectral approaches. Gynecol Oncol. 2007; 107:S248–55. [PubMed: 17825399]

47. Shuhatovich OM, Sharman MP, Mirabal YN, Earle NR, Follen M, Basen-Engquist K. Participant recruitment and motivation for participation in optical technology for cervical cancer screening research trials. Gynecol Oncol. 2005; 99:S226–31. [PubMed: 16143374]

48. Manos M, Ting Y, Wright D, Lewis A, Broker T, Wolinsky S. The use of polymerase chain reaction amplification for the detection of genital human papillomaviruses. Cancer Cells. 1989; 7:209–14.

49. Scheurer M, Dillon L, Chen Z, Follen M, Adler-Storthz K. Absolute quantitative real-time polymerase chain reaction for the measurement of human papillomavirus E7 mRNA in cervical cytobrush specimens. Infect Agent Cancer. 2007; 2:8. [PubMed: 17407544]

50. Cárdenas-Turanzas M, Nogueras-Gonzalez G, Scheurer M, Adler-Storthz K, Benedet J, Beck J, Follen M, Cantor S. The performance of human papillomavirus high-risk DNA testing in the screening and diagnostic settings. Cancer Epidemiol Biomarkers Prev. 2008; 17:2865–71. [PubMed: 18843032]

51. Cárdenas-Turanzas M, Follen M, Nogueras-Gonzalez G, Benedet J, Beck J, Cantor S. The accuracy of the Papanicolaou smear in the screening and diagnostic settings. J Low Genital Tract Dis. 2008; 12:269–75.

52. Guillaud M, Benedet J, Cantor S, Staerkel G, Follen M, MacAulay C. DNA ploidy compared with human papilloma virus testing (Hybrid Capture II) and conventional cervical cytology as a primary screening test for cervical high–grade lesions and cancer in 1555 patients with biopsy confirmation. Cancer. 2006; 107:309–18. [PubMed: 16773634]

53. Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2. New York: Springer; 2009.

54. Pepe, M. The Statistical Evaluation of Medical Tests for Classification and Prediction. 1. New York: Oxford University Press; 2004.

55. Janes, H.; Pepe, M. University of Washington Biostatistics Working Paper Series. 2007. Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: an old concept in a new setting.

56. Redden Weber C, Schwarz RA, Atkinson EN, Cox DD, Macaulay C, Follen M, Richards-Kortum R. Model-based analysis of reflectance and fluorescence spectra for in vivo detection of cervical dysplasia and cancer. J Biomed Opt. 2008; 13:064016. [PubMed: 19123662]

57. Zhu H, Vannucci M, Cox DD. A Bayesian hierarchical model for classification with selection of functional predictors. Biometrics. 2010; 66:463–73. [PubMed: 19508236]

58. Zhu, H.; Cox, D. A functional generalized linear model with curve selection in cervical pre-cancer diagnosis using fluorescence spectroscopy. In: Rojo, J., editor. Optimality: The Third Erich L Lehmann Symposium. Beachwood, Ohio: Institute of Mathematical Statistics; 2009. p. 173-89.

59. Weingandt H, Stepp H, Baumgartner R, Diebold J, Xiang W, Hillemanns P. Autofluorescence spectroscopy for the diagnosis of cervical intraepithelial neoplasia. BJOG. 2002; 109:947–51. [PubMed: 12197377]

60. Ramanujam N, Mitchell MF, Mahadevan A, Thomsen S, Silva E, Richards-Kortum R. Fluorescence spectroscopy: a diagnostic tool for cervical intraepithelial neoplasia (CIN). Gynecol Oncol. 1994; 52:31–8. [PubMed: 8307499]

61. Ramanujam N, Mitchell MF, Mahadevan A, Thomsen S, Malpica A, Wright T, Atkinson N, Richards-Kortum R. Development of a multivariate statistical algorithm to analyze human cervical tissue fluorescence spectra acquired in vivo. Lasers Surg Med. 1996; 19:46–62. [PubMed: 8836996]

62. Ramanujam N, Mitchell MF, Mahadevan A, Thomsen S, Malpica A, Wright T, Atkinson N, Richards-Kortum R. Spectroscopic diagnosis of cervical intraepithelial neoplasia (CIN) in vivo

using laser-induced fluorescence spectra at multiple excitation wavelengths. Lasers Surg Med. 1996; 19:63–74. [PubMed: 8836997]

63. Mitchell MF, Cantor SB, Ramanujam N, Tortolero-Luna G, Richards-Kortum R. Fluorescence spectroscopy for diagnosis of squamous intraepithelial lesions of the cervix. Obstet Gynecol. 1999; 93:462–70. [PubMed: 10075001]

64. Mitchell MF, Cantor SB, Brookner C, Utzinger U, Schottenfeld D, Richards-Kortum R. Screening for squamous intraepithelial lesions with fluorescence spectroscopy. Obstet Gynecol. 1999; 94:889–96. [PubMed: 10546779]

65. Belinson J, Qiao YL, Pretorius R, Zhang WH, Elson P, Li L, Pan QJ, Fischer C, Lorincz A, Zahniser D. Shanxi Province Cervical Cancer Screening Study: a cross-sectional comparative trial of multiple techniques to detect cervical neoplasia. Gynecol Oncol. 2001; 83:439–44. [PubMed: 11606114]

66. Parker M, Mooradian G, Karins J, O'Connor D, Speer B, Owensby P. Hyperspectral diagnostic imaging of the cervix: report on a new investigational device. J Low Genit Tract Dis. 2000; 4:119–24.

67. Parker MF, Mooradian GC, Okimoto GS, O'Connor DM, Miyazawa K, Saggese SJ. Initial neural net construction for the detection of cervical intraepithelial neoplasia by fluorescence imaging. Am J Obstet Gynecol. 2002; 187:398–402. [PubMed: 12193932]

68. Wright T, Ferenczy A, Wray S, Christianson R, Ganguly D. Abstracts presented for the thirtieth annual meeting of the Society of Gynecologic Oncologists: Detection of cervical squamous intraepithelial lesions using evoked tissue fluorescence. Gynecol Oncol. 1999; 72:458.

69. Dattamajumdar, AK.; Wells, D.; Parnell, J.; Lewis, JT.; Ganguly, D.; Wright, TC, Jr. Preliminary experimental results from multi-center clinical trials for detection of cervical precancerous lesions using the Cerviscan system: a novel full-field evoked tissue fluorescence based imaging instrument. 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2001. p. 3150-52.

70. Coppleson M, Reid BL, Skladnev VN, Dalrymple JC. An electronic approach to the detection of pre-cancer and cancer of the uterine cervix: a preliminary evaluation of Polarprobe. Int J Gynecol Cancer. 1994; 4:79–83. [PubMed: 11578389]

71. Singer A, Coppleson M, Canfell K, Skladnev V, Mackellar G, Pisal N, Deery A. A real time optoelectronic device as an adjunct to the Pap smear for cervical screening: a multicenter evaluation. Int J Gynecol Cancer. 2003; 13:804–11. [PubMed: 14675317]

72. Mirabal YN, Chang SK, Atkinson EN, Malpica A, Follen M, Richards-Kortum R. Reflectance spectroscopy for in vivo detection of cervical precancer. J Biomed Opt. 2002; 7:587–94. [PubMed: 12421125]

73. Mourant J, Bocklage T, Powers T, Greene H, Bullock K, Marr-Lyon L. In vivo light scattering measurements for detection of precancerous conditions of the cervix. Gynecol Oncol. 2007; 105:439–45. [PubMed: 17303229]

74. Orfanoudaki I, Themelis G, Sifakis S, Fragouli D, Panayiotides J, Vazgiouraki E. A clinical study of optical biopsy of the uterine cervix using a multi-spectral imaging system. Gynecol Oncol. 2005; 96:119–31. [PubMed: 15589590]

75. Georgakoudi I, Sheets EE, Muller MG, Backman V, Crum CP, Badizadegan K, Dasari RR, Feld MS. Trimodal spectroscopy for the detection and characterization of cervical precancers in vivo. Am J Obstet Gynecol. 2002; 186:374–82. [PubMed: 11904594]

76. Ferris D, Lawhead R, Dicjman E, Holtzapple N, Miller J, Grogan S, Bambot S, Agrawal A, Faupel M. Multimodal hyperspectral imaging of the noninvasive diagnosis of cervical neoplasia. J Low Genit Tract Dis. 2001; 5:65–72. [PubMed: 17043578]

77. DeSantis T, Chakhtoura N, Twiggs L, Ferris D, Lashgari M, Flowers L. Spectroscopic imaging as a triage test for cervical disease: a prospective multicenter clinical trial. J Low Genit Tract Dis. 2007; 11:18–24. [PubMed: 17194946]

78. Burke L, Niloff J, Kobelin M, Abu-Jawdeh G, Zelenchuk A, Modell M. Use of autofluorescence of cells to evaluate cervical neoplasia. J Gynecol Tech. 1996; 2:90.

79. Burke L, Modell M, Niloff J, Kobelin M, Abu-Jawdeh G, Zelenchuk A. Identification of squamous intraepithelial lesions: fluorescence of cervical tissue during colposcopy. J Low Genit Tract Dis. 1999; 3:159–62.

80. Nordstrom R, Burke L, Niloff J, Myrtle J. Identification of cervical intraepithelial neoplasia (CIN) using UV-excited fluorescence and diffuse-reflectance tissue spectroscopy. Lasers Surg Med. 2001; 29:118–27. [PubMed: 11553898]

81. Huh WK, Cestero RM, Garcia FA, Gold MA, Guido RS, McIntyre-Seltman K, Harper DM, Burke L, Sum ST, Flewelling RF, Alvarez RD. Optical detection of high-grade cervical intraepithelial neoplasia in vivo: results of a 604-patient study. Am J Obstet Gynecol. 2004; 190:1249–57. [PubMed: 15167826]

82. Alvarez RD, Wright T. Effective cervical neoplasia detection with a novel optical detection system: a randomized trial. Gynecol Oncol. 2007; 104:281–9. [PubMed: 17173959]

83. Benavides J, Chang S, Park S, Richards-Kortum R, Mackinnon N, MacAulay C, Milbourne A, Malpica A, Follen M. Multispectral digital colposcopy for in vivo detection of cervical cancer. Opt Express. 2003; 11:1223–36. [PubMed: 19465988]

84. Milbourne A, Park SY, Benedet JL, Miller D, Ehlen T, Rhodes H, Malpica A, Matisic J, Van Niekirk D, Atkinson EN, Hadad N, Mackinnon N, et al. Results of a pilot study of multispectral digital colposcopy for the in vivo detection of cervical intraepithelial neoplasia. Gynecol Oncol. 2005; 99:S67–75. [PubMed: 16202444]

85. Park SY, Follen M, Milbourne A, Rhodes H, Malpica A, MacKinnon N, MacAulay C, Markey M, Richards-Kortum R. Automated image analysis of digital colposcopy for the detection of cervical neoplasia. J Biomed Opt. 2008; 13:014029. [PubMed: 18315387]

86. Nakappan S, Park SY, Serachitopol D, Price R, Cardeno M, Au S, Mackinnon N, MacAulay C, Follen M, Pikkula BM. Methodology of a real-time quality control for the multispectral digital colposcope (MDC). Gynecol Oncol. 2007; 107:S215–22. [PubMed: 17825397]
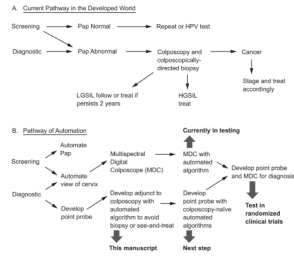
Figure 1.
A and B. Current paradigm of clinical care and demonstration of optical technologies in the process.
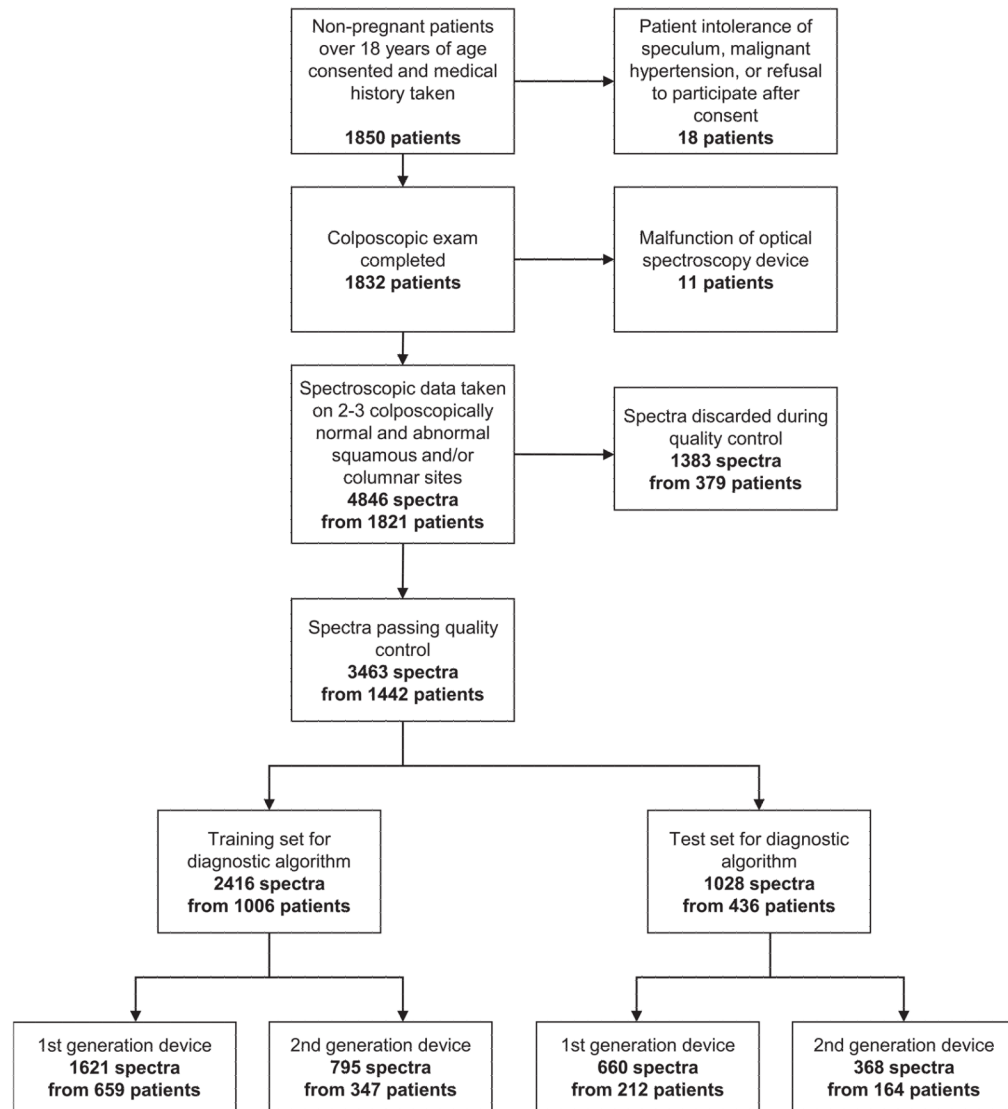
Figure 2.
Patient enrollment and study flow diagram showing final post-quality control sample size.
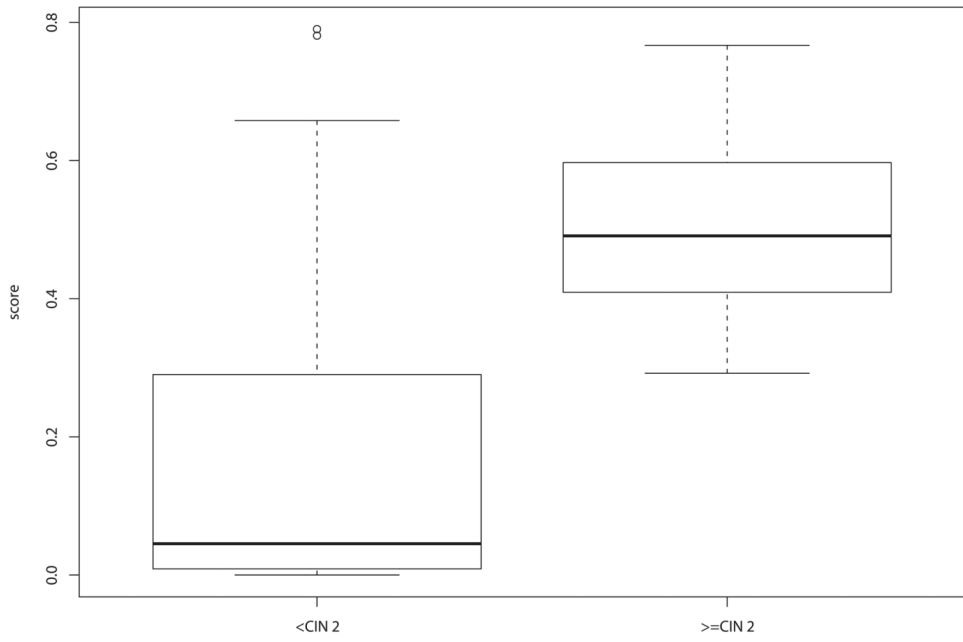
Figure 3.
Boxplot of scores in test set by histologic grade.
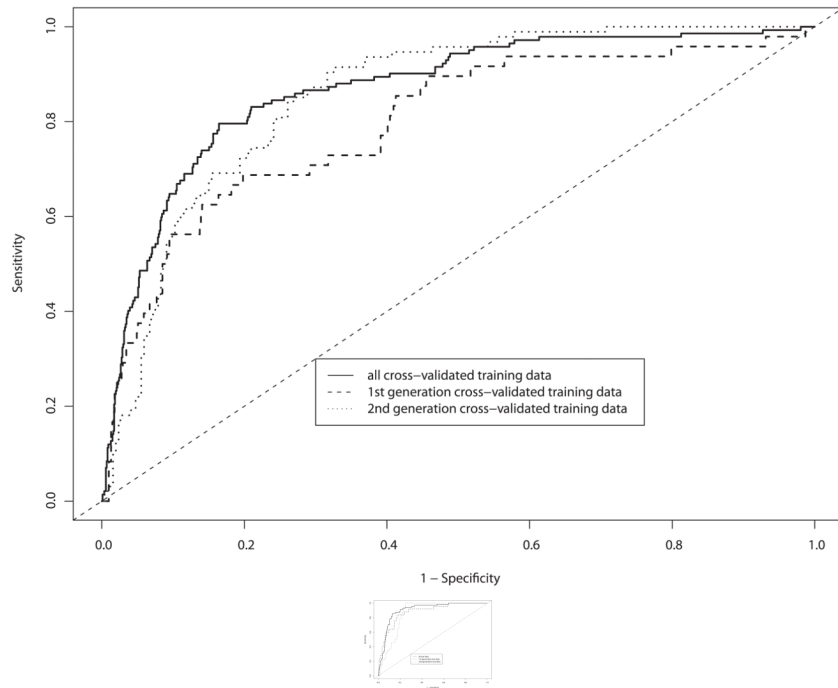
Figure 4.
A and B. By-patient receiver operating characteristic (ROC) curve analysis for each device, and combined data of: (a) logistic regression cross-validated on training and validation data, and (b) final test set.
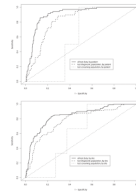
Figure 5.
A and B. Receiver operating characteristic (ROC) test characteristics of the per-patient and per-site analysis of the data. Each graph shows the performance of the whole data set, the diagnostic trial, and the screening trial.
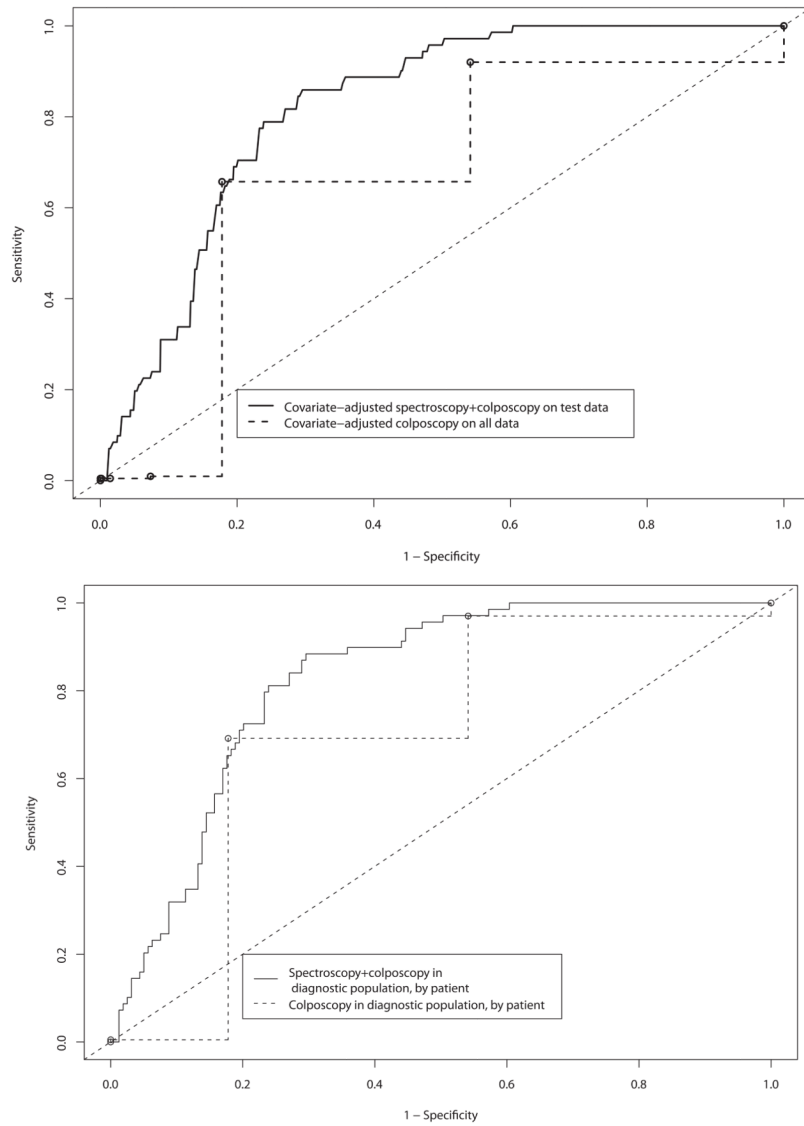
Figure 6.
A and B. Receiver operating characteristic (ROC) test characteristic comparisons of spectroscopy to diagnostic colposcopy using by-patient covariate-adjusted ROC curve analysis on the test set. The second curve compares diagnostic colposcopy to diagnostic spectroscopy.

Table 1

Technologies tested against the gold standard of clinical histopathology with blinded review in this clinical trial

Technology
Clinical cytology taken at the visit
Quantitative cytology using different algorithms: ploidy only, using multiple variables
Hybrid Capture II
Polymerase chain reaction (PCR) for human papillomavirus (HPV)
Messenger ribonucleic acid (mRNA) for HPV
Quantitative histopathology
Fluorescence spectroscopy using a point probe and a colposcope in both screening and diagnostic populations
Reflectance spectroscopy using a point probe and a colposcope in both screening and diagnostic populations

Table 2

Demographic, socioeconomic, and health-related factors for diagnostic and screening populations in Houston, TX and Vancouver, BC

	Diagnostic Population	Screening Population
Race	**850**	**1000**
White	**543** (63.9%)	**491** (49.1%)
Black	**94** (11.1%)	**154** (15.4%)
Hispanic	**112** (13.2%)	**276** (27.6%)
Asian	**63** (7.4%)	**67** (6.7%)
Native American	**8** (0.9%)	**3** (0.3%)
Other	**30** (3.5%)	**9** (0.9%)
Age	**850**	**1000**
Mean ± Std Dev (Yrs)	36.6 ± 11.8	44.1 ± 12.1
Range (Yrs)	18–85	18–80
Birthplace	**820**	**999**
United States/Canada	**663** (80.9%)	**676** (67.7%)
Mexico	**27** (3.3%)	**74** (7.4%)
Central America	**11** (1.3%)	**33** (3.3%)
South America	**12** (1.5%)	**78** (7.8%)
Puerto Rico	**3** (0.4%)	**6** (0.6%)
Vietnam	**4** (0.5%)	**5** (0.5%)
China	**13** (1.6%)	**11** (1.1%)
Other	**87** (10.6%)	**116** (11.6%)
Annual Household Income[1]	**816**	**998**
Low ($0–$19,999)	**161** (19.7%)	**177** (17.7%)
Medium ($20,000–$39,999)	**197** (24.1%)	**285** (28.6%)
High (> $40,000)	**352** (43.1%)	**464** (46.5%)
Don't Know/Refused	**106** (13.0%)	**72** (7.2%)
Education Level	**844**	**999**
High School/GED or Less	**239** (28.3%)	**242** (24.2%)
College	**501** (59.4%)	**622** (62.3%)
Graduate School	**104** (12.3%)	**135** (13.5%)
Employment Status	**834**	**999**
Full/Part Time	**565** (67.7%)	**679** (68%)
Unemployed/Retired/Housewife	**185** (22.2%)	**293** (29.3%)
Student	**83** (10.0%)	**25** (2.5%)
Refused	**1** (0.1%)	**2** (0.2%)
Marital Status	**849**	**1000**
Single (Never Married)	**244** (28.7%)	**198** (19.8%)
Married/Married-Like Situation	**410** (48.3%)	**595** (59.5%)

	Diagnostic Population	Screening Population
Divorced/Separated	**175** (20.6%)	**181** (18.1%)
Widowed	**18** (2.1%)	**26** (2.6%)
Refused	**2** (0.2%)	**0** (0.0%)

[1] Stratification into *Low*, *Medium*, and *High* is with respect to the currency of the country where each patient was seen (United States or Canada).

Table 3

Histology, HPV infection, menopausal status, oral contraceptive use, and age for diagnostic and screening populations in Houston, TX and Vancouver, BC

	Diagnostic Population	Screening Population
Worst Histology	**N=735**	**N=707**
Negative for Dysplasia	**219** (29.8%)	**489** (69.2%)
Atypia	**135** (18.4%)	**129** (18.2%)
HPV Associated Changes	**88** (12.0%)	**59** (8.3%)
CIN I (Mild Dysplasia	**92** (12.5%)	**18** (2.5%)
CIN II (Moderate Dysplasia)	**89** (12.1%)	**7** (1.0%)
CIN III (Severe Dysplasia)	**70** (9.5%)	**5** (0.7%)
CIS	**41** (5.6%)	**0** (0.0%)
Cancer	**1** (0.1%)	**0** (0.0%)
Hybrid Capture II		
Negative	**371** (50.5%)	**615** (87.0%)
Low-risk	**22** (3.0%)	**16** (2.3%)
High-risk	**284** (38.6%)	**59** (8.3%)
Both	**54** (7.3%)	**11** (1.6%)
Unascertainable	**4** (0.5%)	**6** (0.8%)
Menopausal Status		
Premenopausal	**615** (83.7%)	**384** (54.3%)
Peri-menopausal	**21** (2.9%)	**93** (13.2%)
Postmenopausal	**99** (13.5%)	**230** (32.5%)
Oral Contraceptive Use		
No	**260** (35.4%)	**125** (17.7%)
Yes	**475** (64.6%)	**582** (82.3%)
Age		
Mean ± Std Dev (Yrs)	36.6 ± 11.7	44.9 ± 12.0
Range (Yrs)	18–85	18–80

Note: Table includes only those patients included in the analysis, i.e., with useable spectroscopy and biopsies and who passed all quality control measures.

Table 4

Performance of instrumentation and spectral approach from the literature

First Author	Source of Funding	Population	Fluorescence Description	Fluorescence Excitation Wavelength (nm)	Fluorescence Emission Wavelength (nm)	Reflectance Description	Reflectance Illumination Wavelength (nm)	Sensitivity (%) by Site (*pt)	Specificity (%) by Site (*pt)
Fluorescence spectroscopy probe									
Weingandt 2002	University + BIOMED grant	diagnostic	broadband illumination	375–440	520	none	none	*88	*53
Ramanujam 1994	Whitaker	diagnostic	single excitation three	337	420	none	none	87	73
Ramanujam 1996	Whitaker	diagnostic	excitations three	337,380,460	300–700	none	none	91	76
Ramanujam 1996	Whitaker	diagnostic	excitations three	337,380,460	300–700	none	none	88	70
Mitchell 1999 (a)	University + NCI	diagnostic	excitations three	337,380,460	300–700	none	none	88	70
Mitchell 1999 (b)	University + NCI	screening	excitations multiple	337,380,460	300–700	none	none	76	81
Chang 2002	University + NCI	diagnostic	excitations	330–480	330–700	none	none	71	77
Belinson 2001	Optical Biopsy Technologies	screening	not stated	not stated	not stated	none	none	*94	*9
multispectral									
Parker 2000	STI + U.S. Gov't	diagnostic	one excitation	365	400–800	none	none	80	81
Parker 2002	STI + U.S. Gov't	diagnostic	one excitation	365	400–800	none	none	*98	*99
Wright 1999	Lifespex	diagnostic	multiple excitations	not stated	not stated	none	none	99	95
Dattamajumdar 2001	Lifespex	diagnostic	multiple excitations	not stated	not stated	none	none	84	93
Reflectance spectroscopy probe									
Coppleson 1994	Polartechnics	diagnostic	none	none	none	two wavelengths four	not stated	90	90
Singer 2003	Polartechnics	diagnostic	none	none	none	wavelengths*	not stated	*70	*77
Mirabal 2002	University + NCI	diagnostic	none	none	none	white light	355–655	72	83
Mourant 2007	NIH + NCI	diagnostic	none	none	none	white light	not stated	100	80
multispectral									
Orfanoudaki 2005	University + F.O.R.T.H.	diagnostic	none	none	none	white light time course	not stated	*91	*100
Combined fluorescence and reflectance spectroscopy probe									
Georgakoudi 2005	University + NIH	diagnostic	multiple excitations	337–610	350–700	white light	350–700	92	71
Chang 2005	University +NCI	diagnostic	multiple excitations	330–480	300–700	white light	355–655	*83	*80
multispectral									

First Author	Source of Funding	Population	Fluorescence Description	Fluorescence Excitation Wavelength (nm)	Fluorescence Emission Wavelength (nm)	Reflectance Description	Reflectance Illumination Wavelength (nm)	Sensitivity (%) by Site (*pt)	Specificity (%) by Site (*pt)
Ferris 2001	SpectRx, Inc.	diagnostic	multiple excitations	300–500	not stated	white light illumination	350–900	100 *95	95 *83
Desantis 2006	SpectRx, Inc.	diagnostic	multiple excitations	300–500	not stated	white light illumination	350–900	*95	*55
Burke 1996	Medispectra	diagnostic	single excitation	337	370–620	white light full field illumination	not stated	81 *96	82 *71
Burke 1999 – Group I	Medispectra	diagnostic	single excitation	337	not stated	white light full field illumination	not stated	93	94
Burke 1999 – Group II	Medispectra	diagnostic						89	93
Nordstrom 2001 - Fluorescence	Medispectra	diagnostic	single excitation	355	390–590	white light full field illumination	360–740	91	93
Nordstrom 2001 – Reflectance	Medispectra	diagnostic						82	67
Huh 2004	Medispectra	diagnostic	single excitation	337	not stated	white light full field illumination	360–740	*92	*50
Alvarez 2007	Medispectra	diagnostic	single excitation	337	360–590	white light full field illumination	360–740	56 *52	87 *90