

Journal of Biomedical Optics

SPIEDigitalLibrary.org/jbo

Accuracy of optical spectroscopy for the detection of cervical intraepithelial neoplasia without colposcopic tissue information; a step toward automation for low resource settings

Jose-Miguel Yamal
Getie A. Zewdie
Dennis D. Cox
E. Neely Atkinson
Scott B. Cantor
Calum MacAulay
Kalatu Davies
Isaac Adewole
Timon P. H. Buys
Michele Follen

Accuracy of optical spectroscopy for the detection of cervical intraepithelial neoplasia without colposcopic tissue information; a step toward automation for low resource settings

Jose-Miguel Yamal,^a Getie A. Zewdie,^a Dennis D. Cox,^b E. Neely Atkinson,^c Scott B. Cantor,^c Calum MacAulay,^d Kalatu Davies,^e Isaac Adewole,^f Timon P. H. Buys,^d and Michele Follen^e

^aThe University of Texas Health Science Center at Houston, Division of Biostatistics, School of Public Health, 1200 Herman Pressler, RAS W928 Houston, Texas 77030

^bRice University, Department of Statistics, P.O. Box 1892, Houston, Texas 77251-1892

^cThe University of Texas MD Anderson Cancer Center, Department of Biostatistics, P.O. Box 301402, Houston, Texas 77230-1402

^dBritish Columbia Cancer Research Centre, Integrative Oncology, 675 West 10th Avenue, Vancouver, B.C. V5Z 1L3

^eDrexel University College of Medicine, Department of Obstetrics and Gynecology, 245 North 15th Street, Philadelphia, Pennsylvania 19102

^fUniversity of Ibadan, Department of Obstetrics and Gynecology, College of Medicine, Ibadan, Nigeria

Abstract. Optical spectroscopy has been proposed as an accurate and low-cost alternative for detection of cervical intraepithelial neoplasia. We previously published an algorithm using optical spectroscopy as an adjunct to colposcopy and found good accuracy (sensitivity = 1.00 [95% confidence interval (CI) = 0.92 to 1.00], specificity = 0.71 [95% CI = 0.62 to 0.79]). Those results used measurements taken by expert colposcopists as well as the colposcopy diagnosis. In this study, we trained and tested an algorithm for the detection of cervical intraepithelial neoplasia (i.e., identifying those patients who had histology reading CIN 2 or worse) that did not include the colposcopic diagnosis. Furthermore, we explored the interaction between spectroscopy and colposcopy, examining the importance of probe placement expertise. The colposcopic diagnosis-independent spectroscopy algorithm had a sensitivity of 0.98 (95% CI = 0.89 to 1.00) and a specificity of 0.62 (95% CI = 0.52 to 0.71). The difference in the partial area under the ROC curves between spectroscopy with and without the colposcopic diagnosis was statistically significant at the patient level ($p = 0.05$) but not the site level ($p = 0.13$). The results suggest that the device has high accuracy over a wide range of provider accuracy and hence could plausibly be implemented by providers with limited training. © 2012 Society of Photo-Optical Instrumentation Engineers (SPIE). [DOI: [10.1117/1.JBO.17.4.047002](https://doi.org/10.1117/1.JBO.17.4.047002)]

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

Paper 11336 received Jun. 30, 2011; revised manuscript received Jan. 30, 2012; accepted for publication Feb. 17, 2012; published online Apr. 5, 2012.

1 Introduction

Cervical cancer is the second most common cancer in women worldwide and a leading cause of death, particularly in low-resource areas without screening programs.¹ GLOBOCAN data from 2008 show that 80% of new cervical cancer incidence and mortality will occur in low-resource settings.² Organized screening programs have decreased the morbidity and mortality of cervical cancer in all geographical regions since their introduction.³ To make an impact, screening programs appropriate for low-resource settings should be developed.^{4,5}

Organized screening programs in high-resource settings have been based on the use of the Papanicolaou smear, a cytologic sampling of the cervix.³ If abnormal, a second visit is required for a colposcopy in which the cervix is visualized under 3.5 to 15 fold magnification. Typically acetic acid at 5 to 6% is used as a contrast agent, and the cervix is viewed using white and green light. The green light enhances the recognition of abnormal vasculature.⁶ Colposcopically directed biopsies are

performed, and the patient returns for a third visit for treatment. While treatments have varied over the last 50 years, they can be summarized as either “ablative” in which no specimen is obtained or “excisional” in which the tissue is removed for histopathological analysis. Either treatment results in two-year cure rates for pre-cancers of 85 to 90%.⁷ Invasive cancers are treated with radical hysterectomy, radiation therapy, chemotherapy, or combinations. Detection in the precancerous phase results in tangible survival and little morbidity.

The most important advance in cervical cancer etio-pathogenesis is the establishment of a causal link to the human papillomavirus (HPV).^{8,9} HPV has found to be causal in the development of cervical cancer. This finding has resulted in two major advances in the field of cervical cancer control and prevention: HPV testing and an HPV vaccine. There are over 60 types of HPV that infect the cervix, and among them some types are associated with cervical cancer while other are associated with warts.^{10,11} Studies worldwide have demonstrated that types 16 and 18 account for 70% of invasive cancers. Two preventative vaccines have been developed and both cover types 16 and 18.^{12,13} HPV testing has been evaluated

Address all correspondence to: Jose-Miguel Yamal, The University of Texas Health Science Center at Houston, Division of Biostatistics, School of Public Health, 1200 Herman Pressler, RAS W928 Houston, TX 77030. Tel: 713 500 9566; Fax: 713 500 9530; E-mail: Jose-Miguel.Yamal@uth.tmc.edu.

for diagnosis and screening in the United States, and recent guidelines from the U.S. Preventive Services Task Force suggest that HPV testing is not cost-effective in the United States. A prospective cohort study is being conducted in the province of British Columbia to evaluate the use of HPV testing in predicting recurrence of pre-invasive lesions. Thus the role of HPV testing in the developed world, where abundant resources have been spent on cervical cancer screening, remains unclear.

The International Agency for Research in Cancer (IARC), part of the World Health Organization, has led the international effort to address cervical screening in low- and middle-income countries. It has published reviews addressing the subject, guidelines for care, and manuals for cervical cancer screening and treatment. They have analyzed the distribution of HPV types worldwide in order to assess the adequacy of the vaccine. Furthermore, they have conducted provocative studies of alternative strategies to Papanicolaou testing such as visual inspection with acetic acid with (VIAM) and without magnification (VIA), visual inspection with Lugol's Iodine (VILI), and HPV testing with the Qiagen test.

Sankaranarayanan, from the IARC, conducted a randomized trial of three strategies in rural India: no increased intervention, cytological screening, and VIA and demonstrated that both the cytology arm and VIA were superior to no increased intervention.¹⁴⁻¹⁶ Further analysis yielded results showing that the most cost-effective strategy was VIA.¹⁷⁻¹⁹ Sankaranarayanan later conducted a randomized trial of the Papanicolaou test, VIA, and HPV testing with the Qiagen test in rural India and concluded the HPV test was most effective in detecting cervical cancer and increasing lives saved.⁹ This study suggests there may be a role for HPV testing in the low- and middle-income country setting. At the present time, the Qiagen test may cost as much as \$30 while a more affordable version called "Care HPV" is undergoing pilot testing in selected countries. While the role and cost of HPV testing is a moving target, VIA followed by immediate treatment, if appropriate, is being disseminated in many countries.^{15,20} One issue with the studies conducted by Sankaranarayanan and others studying VIA, the Papanicolaou smear, and HPV testing is that only the patients with abnormalities (detected by the screening tests) received further diagnostic studies. This is referred to as "diagnostic bias" or verification bias in the epidemiologic literature. When every patient in the sample is not biopsied (i.e., doesn't have the gold standard test), there is no way to calculate a true sensitivity or specificity because the denominator contains only those patients with abnormalities rather than every patient in the trial. One can adjust for verification bias; however, one must assume that the positive and negative predictive values in people who had the gold standard procedure (those who screened positive) are the same as the positive and negative predictive values in people who did not have the gold standard procedure (those who did not screen positive). This is unlikely to hold true in the case of prescreening for HPV and only sending those positive for diagnostic procedures. Furthermore, VIA is likely to suffer from low specificity. Thus there remains a need for cost-effective see-and-treat strategies in the low-resource setting for cervical cancer that are more cost-effective, e.g., a test that is significantly more effective than VIA yet not much more expensive.

There are three major obstacles to controlling cervical cancer in the developing world: 1) the HPV vaccine is expensive

(though reductions in prices may make it affordable) and addresses 70% of the types that are known to cause cancer; 2) an increase in the quantity and quality of trained health care workers is needed to perform a cytology-based screening program, Colposcopy, VIA, or any technology that involves see and treat; and 3) there is a need for a revamped electrical and water infrastructure in many areas. Other important obstacles include: 1) dissemination of the vaccine, which has been slow in the United States;²¹ 2) lack of sufficient understanding of HPV types outside of the developed world; 3) lack of trained personnel in cytology, histopathology, and surgical oncology; and 4) absence of functional cancer registries and plans for follow-up of patients.

Figure 1 shows the screening and diagnostic paradigms for both the high- and low-resource settings that are clinically necessary to have an impact on cervical cancer. An automated pathway that could eliminate unnecessary biopsies and confirm and differentiate high-grade disease from cancer would be of great value in the high-resource setting because it would save valuable health care dollars. To be cost-effective in the low-resource setting, the strategy must be deployable in an environment without electricity, be as cost-effective as visual inspection with acetic acid, and be usable by nurses or trained health care workers.

A new generation of inexpensive, miniature optical sensors has been developed to measure the interaction of light with tissue. Optical spectroscopy has shown that tissue fluorescence spectra contain information about the metabolic activity of epithelial cells. As precancer and cancer develop in the epithelia of the colon,²² esophagus,²³ bladder,²⁴ bronchus,²⁵ and cervix,²⁶⁻⁴¹ there are resulting changes in the metabolic activity of the tissue. Specifically, NADH, FAD, and collagen are fluorophores that change in concentration as lesions become more dysplastic.⁴²⁻⁴⁴ Reflectance spectroscopy uses white light to probe tissue changes that correlate well with precancerous and cancerous changes in tissue morphology. Biologic plausibility studies have demonstrated that reflected white light scatters more as chromatin increases in density in the nucleus, and that the changes in vasculature are evident in reflectance spectra.⁴⁵⁻⁵⁰

We previously published the results of spectroscopy as an adjunct to colposcopy, using histopathology with consensus review as the gold standard, for the detection of cervical intraepithelial neoplasia (CIN).⁵¹ In that study, we created a classification algorithm to discriminate normal tissue from neoplastic and precancerous tissue. This algorithm used spectroscopy measurements as well as the colposcopic diagnosis, menopause status, age, oral contraceptive use, and colposcopic tissue type together as predictors in the model. The algorithm had an estimated sensitivity of 1.00 [95% CI = 0.92 to 1.00] and specificity of 0.71 [95% CI = 0.62 to 0.79] for detecting moderate cervical intraepithelial neoplasia (CIN 2) or worse. That work established spectroscopy could be successfully used as an adjunct to colposcopy. By comparison, in the diagnostic study sample, colposcopy had a sensitivity of 98% and specificity of 45%.⁵² A previous study by Cantor et al. showed that the increased specificity of spectroscopy has the potential to save unnecessary biopsies through a see-and-treat at one visit and be cost-effective.^{53,54} This new study investigates the performance of the device by a "colposcopic diagnosis naïve" provider.

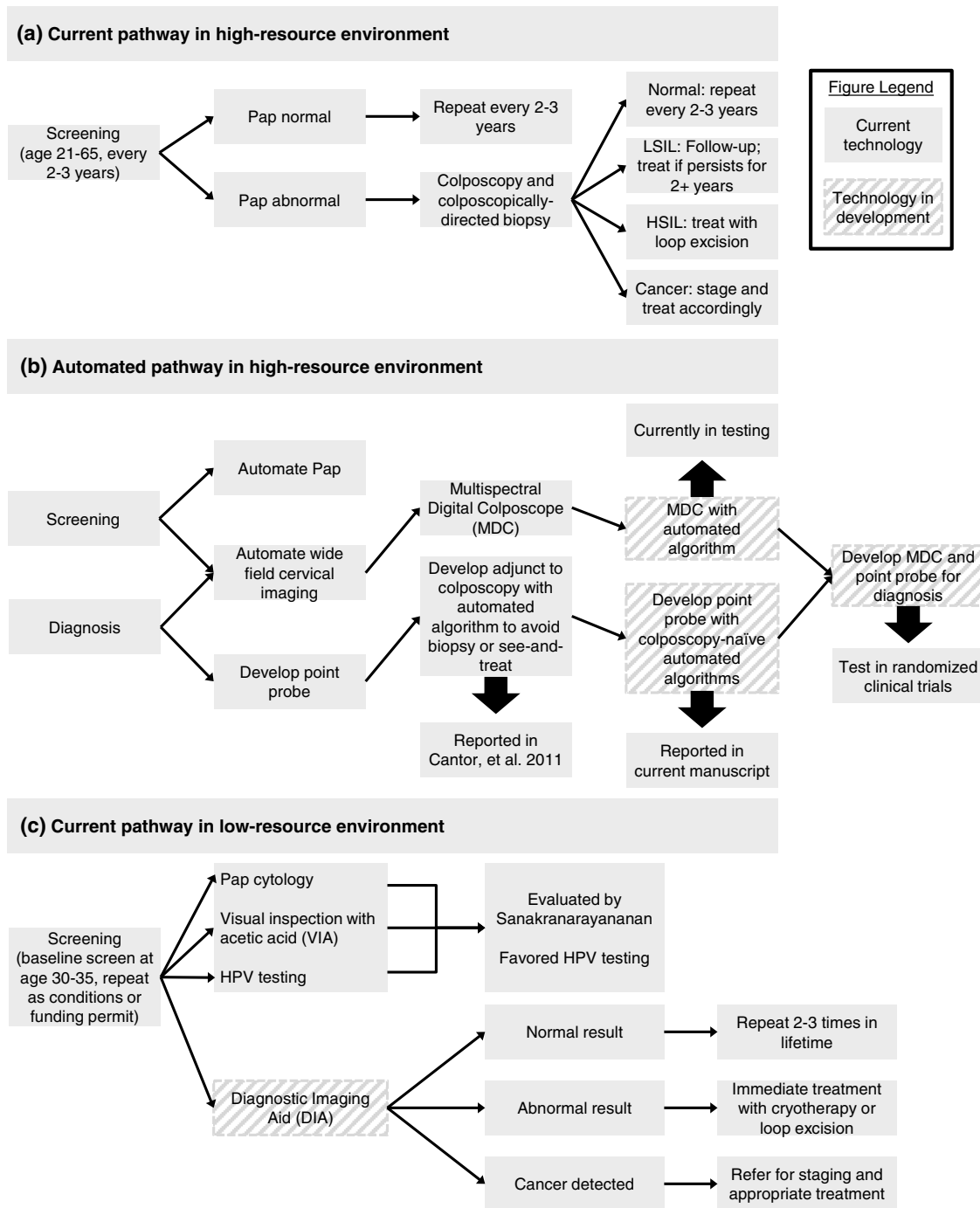


Fig. 1 Current and proposed pathways for screening and diagnosis of cervical cancer.

2 Material and Methods

2.1 Overview of Study Procedures

Patients were recruited either from a screening population with no self-reported history of abnormal Papanicolaou smears or a diagnostic population who were referred with an abnormal Papanicolaou smear or had previous treatment for CIN at three clinical locations. Each patient received a Papanicolaou smear and colposcopic examination of the cervix. Following colposcopic examination, a fiber-optic probe was placed in gentle contact with the cervix, and 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 who presented visible anomalies had a sampling of both abnormal and normal tissue areas.

The probe interrogated an area 2 mm in diameter on the cervix. Following spectroscopic measurements, all sites interrogated with the fiber-optic probe were biopsied with a biopsy forceps yielding specimens that were 2 mm long by 1 mm

wide by 1 mm deep, which closely approximated the measured area.

Details on the study design and the overview of study procedures can be found in Cantor et al.⁵¹

2.2 Gold Standard

Histopathologic sections were initially reviewed clinically by the pathologist at the respective institution, and a second time by a study histopathologist who was blinded to the results of the first review to colposcopy and to all other clinical tests including the spectroscopy.⁴⁹ When diagnoses were discrepant, the specimens were reviewed a third time by the study histopathologist to resolve the discrepancy. The histopathologic consensus diagnosis was used as the gold standard for the trial. The goal was to discriminate patients with no abnormalities to low-grade squamous intraepithelial lesions from those with moderate to high-grade lesions, carcinoma *in situ*, or cancer. Thus we identified those patients with “disease” as “CIN 2 or worse” (i.e., “ \geq CIN 2”) including those who had a histology reading of CIN 2, CIN 3, carcinoma *in situ* (CIS), and invasive squamous cancer. Patients were classified as “ $<$ CIN 2” if their histology reading was normal, atypia, inflammation, HPV-related changes, or CIN 1. We developed an algorithm to discriminate between the two groups using biographical variables and spectroscopic measurements. A subset of slides were reviewed a fourth time by the study pathologists at the MD Anderson Cancer Center, a fifth time at the British Columbia Cancer Agency, and a sixth time by a cervical pathologist at the Brigham and Women’s Hospital, Boston, Massachusetts. The results of these reviews can be found in Malpica et al.⁴⁹ The kappa statistics for the diagnosis of CIN 2 and worse were between 0.80 and 1.00, which represents the highest categories of agreement.

2.3 Spectroscopic Measurements

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

There were two generations of the device used during the seven years of the trial. The second-generation device improved over the first generation in that it was more affordable to construct and took less time to take measurements. More details of the devices used during the trial can be found in Freeberg et al.⁴⁸ and a full treatment of data processing from the devices can be found in Marin et al.⁵⁵

2.4 Biographical Variables

Four biographical variables were used as features in the classification algorithm: menopausal status, colposcopic tissue type, hormone use, and age. The study of biographical variables and their inclusion in algorithms in optical technologies is an advance in the field. These variables were included because, in biological plausibility studies of this technology, they were

found to be statistically significant variables that explain variability in the spectroscopic measurements and because this information was available at the time of the clinical visit.^{28,35,46,56–59} 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 comparison with the laboratory results to ensure consistency. The colposcopic tissue type, identified by the naked eye, does not require extensive training and was classified as either columnar or squamous. Hormone use was defined by the use of any of oral contraceptive pills, hormone replacement therapy, or depo-provera.

2.5 Development of an Algorithm for Optical Spectroscopy

The overall strategy of the algorithm development was to first split the data into training and test sets, perform data reduction on the EEMs, train several algorithms using the data-reduced EEMs and biographical variables, choose the algorithm with the highest accuracy, and compute its accuracy on the independent data (the test set).

All classification algorithms have tuning parameters that must be estimated from the data. The process of obtaining the parameters leads to biased performance toward the data on which it was trained. 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% in each set, respectively. All measurements within a subject were either in the training or in the test set. 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 nondisease. Therefore, the data were from a combination of the screening and diagnostic groups to provide a sufficient number of patients with disease.

Because we considered multiple algorithms and intended 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 split the training data into two datasets, fivefold cross-validation was conducted within the training set for each algorithm.⁵¹ Some algorithms have more tuning parameters than others. The inclusion of more parameters makes an algorithm more susceptible to overtraining, wherein its performance appears good on the data on which they are trained, yet they perform poorly on new data.

We explored several options for dimension reduction methods as inputs for each of the algorithms.⁵¹ 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 within that excitation wavelength. This reduced the dimension of the measurements to between one and three principal components per excitation wavelength.

We applied a variety of classification algorithms to the features obtained from the methods of EEM data reduction and the biographical variables: age, menopausal status, hormone use, and colposcopic tissue type. The various classification algorithms applied to the training data included various Bayesian methods, including Bayesian variable selection and 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, and kernel support vector machines.⁶⁰

Once the classifiers were optimized, we chose the algorithm that had the highest specificity for 80% sensitivity (as estimated by fivefold cross-validation on the training set). We used receiver operating characteristic (ROC) curve analysis to summarize the estimated performance of the selected algorithm.⁶¹

We provide details on the logistic regression approach, as this yielded the best algorithm results as estimated from fivefold cross-validation with training data. We fit a logistic regression model to the principal components of the spectroscopic data and biographical variables and used Akaike's information criterion to perform a backward stepwise selection of variables.⁶⁰

We report both per-site and per-patient results when appropriate. The per-patient results were computed by taking the patient spectroscopy score to be the maximum score among all sites within that patient. This was compared with the patient's disease status, defined as the worst histologic finding among all sites for the patient.

2.6 Interaction of Spectroscopy with Colposcopy

To examine the interaction of spectroscopy with colposcopy, we first compared the accuracy of the spectroscopy algorithm with the colposcopic diagnosis (the results of our previous study⁵¹) to the accuracy of spectroscopy without the colposcopic diagnosis. The colposcopist identified sites—both normal and abnormal—of interest upon which to place the probe to obtain spectroscopic measurements; therefore, the training of a completely colposcopy-independent algorithm for spectroscopy is impossible since even when the colposcopic diagnosis information is not used, the knowledge and experience of the colposcopist guides the placement of the probe. This placement could affect the quality of the resulting fluorescence measurements. To evaluate the impact of operator expertise on gathered data, we sought to compare the accuracy of measurements taken by highly experienced colposcopists to those gathered by nonexperts. The level of provider expertise was estimated by their clinical accuracy in the performance of standard-of-care colposcopy. This was compared with the accuracy of spectroscopy. A high degree of correlation between accuracy in colposcopy and in the results of spectroscopy measurements would suggest a more substantial impact of operator expertise on the quality of spectroscopic tissue data.

Each colposcopist's accuracy was estimated using the whole data set by computing the percentage of colposcopy diagnoses that were correct: either correctly identifying a site that had an abnormal biopsy or giving a negative clinical diagnosis for a site that had a normal biopsy. We used the percentage correct instead of ROC methods because the low number of "diseased" patients per colposcopist does not allow for a proper estimation of sensitivity. Spectroscopy used the colposcopists' expertise as to where to place the probe but not the actual colposcopy diagnosis; the accuracy of spectroscopy was similarly estimated by computing the percentage of measurements that correctly classified the sample.

To examine the interaction between spectroscopy and colposcopy, we computed the correlation between the spectroscopy algorithm scores and the ordinal colposcopic diagnosis. Per-patient accuracies were computed by comparing the maximum

spectroscopy score at the colposcopic negative or positive sites for each patient to the worst histologic diagnosis for all sites in that patient. An analysis by tissue site compared spectroscopy to the histologic diagnosis at that site. This analysis could additionally suggest diagnostic strategies to integrate spectroscopy into the clinical workflow.

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. Correlations were calculated using Spearman's rank correlation coefficient. The Wilcoxon signed-rank test was used to compare the average difference in paired continuous variables. The test set ROC curves were compared by constructing bootstrap confidence intervals of the difference between the smoothed partial areas under the ROC curves (pAUC).^{62,63}

3 Results

The study yielded usable spectroscopy and biopsies for 735 of the 850 diagnostic patients and 707 of the 1000 screening patients. According to tissue histology, 201 of the 735 (27%) patients had biopsies that showed CIN 2, CIN 3, CIS, or cancer in the diagnostic group, compared with 12 of the 707 patients (2%) in the screening group. The test set collected using the second-generation device had 164 patients (46 diseased and 118 nondiseased). This subset was used to obtain the unbiased estimates of the accuracy since the second-generation device is the basis for the design of alpha prototypes that are being used going forward.

The best algorithm results as estimated from fivefold cross-validation with training data were obtained using logistic regression with variable selection applied to the biographical variables and principal components of the fluorescence data. The final variables in the model, as chosen using ROC curve analysis, included all four biographical variables and 23 of the principal component variables from the spectroscopic data.

Using a by-site sensitivity of 0.80, we identified a cutpoint of 0.1615 for the spectroscopy score on the training set, yielding a by-patient sensitivity (correctly identifying those patients who had histology reading CIN 2 or worse) of 0.98 [95% confidence interval (CI) = 0.89 to 1.00] and a specificity (correctly identifying those patients who had histology reading CIN 1 or better) of 0.62 (95% CI = 0.52 to 0.71) on the test set using the second-generation device. Given the prevalence of 16.4% of CIN 2 or worse in our combined screening and diagnostic and combined training and test data, the positive predictive value was 0.50, and the negative predictive value was 0.99.

Figure 2 shows the by-patient boxplot of the spectroscopy scores on the test set by histologic diagnosis and for the subset of screening and diagnostic patients separately. The spectroscopy scores increase with the severity of disease. Figure 3 shows the results of the by-patient ROC curve analysis of the spectroscopy algorithm with and without the colposcopic diagnosis on the test set for the second-generation device on the test data. The area under the ROC curve (AUC) for the by-patient algorithm using spectroscopy and the colposcopic diagnosis was 0.86 (0.81 to 0.90, CI) (previously published in Cantor et al.⁵¹) as compared with an AUC of 0.82 (0.76 to 0.89, CI) for spectroscopy without the colposcopic diagnosis. Including the

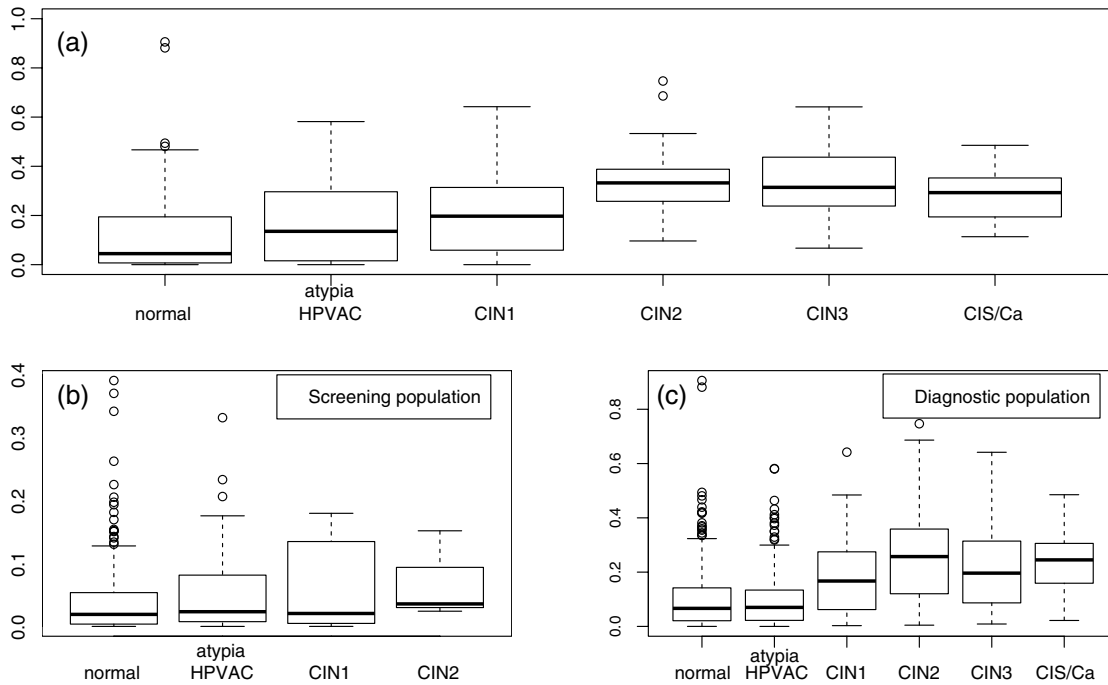


Fig. 2 By-patient boxplot of spectroscopy scores on the test set by histologic diagnosis for (a) combined screening and diagnostic populations (second-generation device); (b) screening population (first- and second-generation device); and (c) diagnostic population (first- and second-generation device).

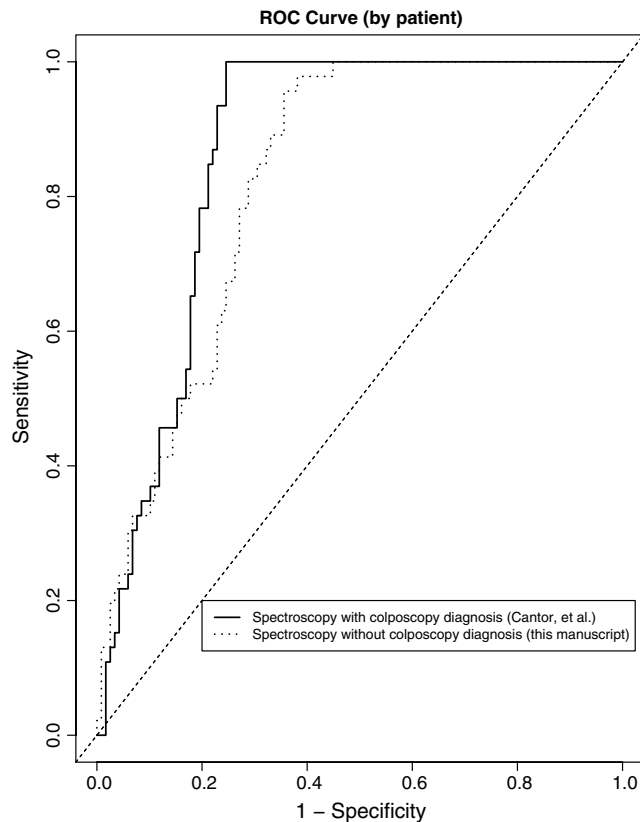


Fig. 3 By-patient receiver operating characteristic curve analysis of logistic regression algorithm using spectroscopy with and without the colposcopic diagnosis on the second-generation device test set. The partial AUCs for sensitivity between 0.8 and 1.0 were statistically significantly different ($p = 0.05$).

colposcopic diagnosis in the algorithm improves the sensitivity and specificity. The by-patient pAUCs in the range of 0.80 to 1.00 sensitivity were statistically significantly different (pAUC = 0.77 versus 0.65 with and without the colposcopic diagnosis, respectively, $p = 0.05$). Figure 4 shows the by-site results. The by-site AUCs were 0.85 (0.81 to 0.89, CI) and 0.83 (0.79 to 0.87, CI) for spectroscopy with and without the colposcopic diagnosis, respectively. The by-site pAUCs were not statistically different (pAUC = 0.66 versus 0.62, $p = 0.13$).

The by-site and by-patient results for the test set diagnostic and screening populations are shown in Figs. 5(a) and 5(b), respectively. We included the data from the first generation of device due to the very low prevalence of disease in the screening population. We observed a decrease in performance of spectroscopy in the diagnostic population. Furthermore, Fig. 6 shows the accuracy of spectroscopy when the algorithm was trained with and without the biographical variables. Adding the biographical variables increased the accuracy substantially.

3.1 Results of the Role of the Colposcopist's Expertise

There were measurements from nine different colposcopists in the test set. We used the combined first- and second-generation devices to get more precise estimates of accuracy. The percentage of correct classifications for colposcopy ranged from 0.69 to 0.90 with a median of 0.82. The accuracy percentages of spectroscopy ranged from 0.59 to 1.00 with a median of 0.88. The average difference, in favor of spectroscopy, was 0.06 ($p = 0.39$, Wilcoxon signed-rank test) with no statistical difference between the two. A scatter plot of the by-patient accuracies is shown in Fig. 7. The correlation coefficient between the two was 0.65 and marginally statistically significant ($p = 0.07$, Spearman's rank correlation coefficient). There is

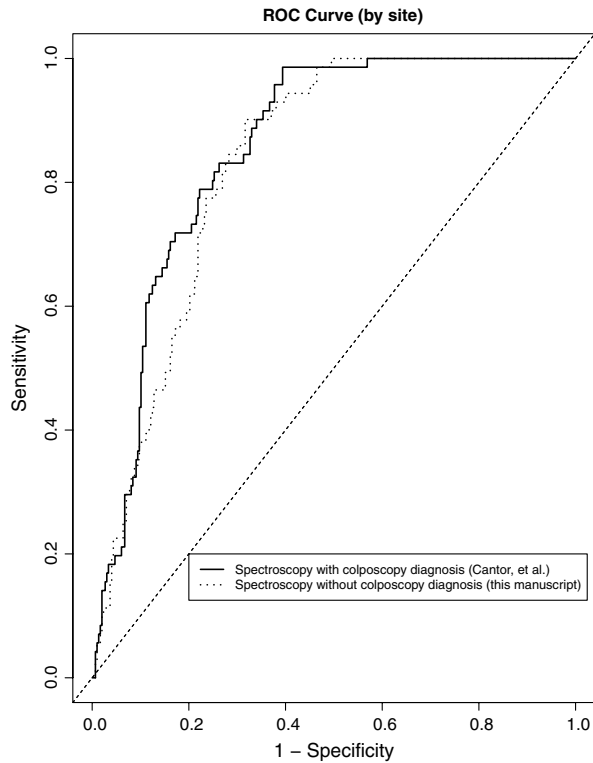


Fig. 4 By-site receiver operating characteristic curve analysis of logistic regression algorithm using spectroscopy with and without the colposcopic diagnosis on the second-generation device test set. The partial AUCs for sensitivity between 0.8 and 1.0 were not statistically significantly different ($p = 0.13$).

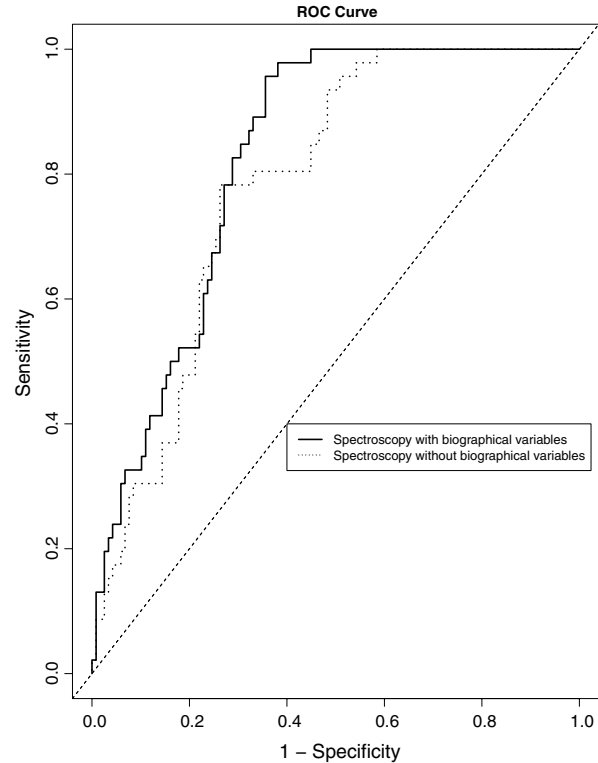


Fig. 6 By-patient receiver operating characteristic curve analysis of logistic regression algorithm using spectroscopy without the colposcopic diagnosis, with and without the biographical variables on the second-generation device test set.

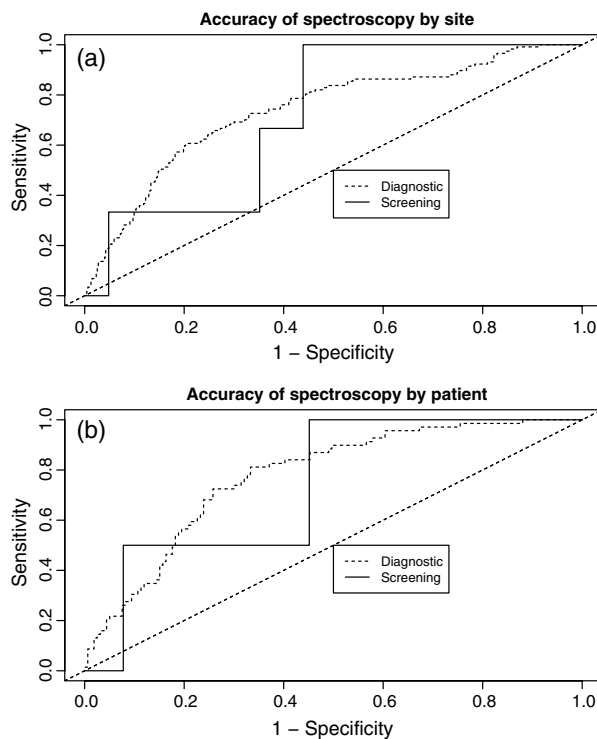


Fig. 5 (a) By-site and (b) by-patient receiver operating characteristic curve analysis of logistic regression algorithm using spectroscopy without the colposcopic diagnosis, using both first- and second-generation devices, on test set by study population.

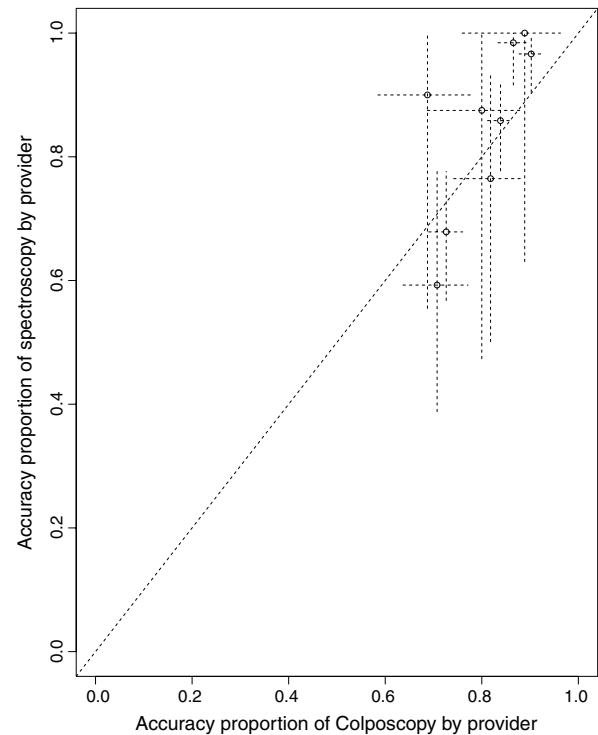


Fig. 7 Scatter plot of the accuracy of colposcopy (using the whole data set) compared with the accuracy of colposcopically directed but colposcopic diagnosis-independent spectroscopy (using the second-generation device test set). The crosshairs on each point give the 95% confidence interval—the horizontal line is the confidence interval for the accuracy of colposcopy and the vertical line is for spectroscopy.

Table 1 Frequencies of patients from the screening and diagnostic population evaluated using spectroscopy and colposcopy by each provider. The spectroscopy accuracies were calculated using both generations of device in the test set. The colposcopy accuracies were calculated using all available data in both the training and test sets. The provider's degree is coded as "NP" for nurse practitioners, "GynOnc" for gynecologic oncologists, and "MD" for gynecologists.

| Provider | Degree | Spectroscopy | | | | | Colposcopy | | | | |
|----------|--------|-----------------------|------------------------|-------------|-------------|----------|-----------------------|------------------------|-------------|-------------|----------|
| | | Screening sample size | Diagnostic sample size | Specificity | Sensitivity | Accuracy | Screening sample size | Diagnostic sample size | Specificity | Sensitivity | Accuracy |
| 1 | NP | 40 | 24 | 1.00 | 0.00 | 0.98 | 132 | 97 | 0.88 | 0.48 | 0.87 |
| 2 | NP | 53 | 53 | 0.99 | 0.00 | 0.86 | 175 | 165 | 0.84 | 0.79 | 0.84 |
| 3 | NP | 48 | 41 | 0.99 | 0.33 | 0.97 | 175 | 121 | 0.91 | 0.65 | 0.90 |
| 4 | NP | 8 | 0 | 1.00 | NA | 1.00 | 23 | 0 | 0.89 | NA | 0.89 |
| 5 | NP | 5 | 5 | 1.00 | 0.00 | 0.90 | 15 | 18 | 0.68 | 0.75 | 0.69 |
| 6 | GynOnc | 36 | 48 | 0.63 | 0.84 | 0.68 | 110 | 167 | 0.74 | 0.68 | 0.73 |
| 7 | GynOnc | 4 | 4 | 0.83 | 1.00 | 0.88 | 14 | 12 | 0.77 | 1.00 | 0.80 |
| 8 | MD | 3 | 24 | 0.42 | 0.73 | 0.59 | 15 | 58 | 0.68 | 0.80 | 0.71 |
| 9 | GynOnc | 2 | 15 | 0.63 | 0.89 | 0.76 | 9 | 59 | 0.80 | 0.88 | 0.82 |

significant overlap in the 95% confidence intervals. Table 1 gives the accuracy, sensitivity, specificity, and numbers of screening and diagnostic patients by provider.

4 Discussion

Our previous results of optical spectroscopy focused on spectroscopy as an adjunct to colposcopy, although expert colposcopists are unlikely to be available in low-resource settings. The classification algorithm used in the previous study was colposcopically directed (the colposcopists guided the placement of the probe onto the cervix), and the colposcopic diagnosis and colposcopic tissue type were also used as covariates in the classification model. In the current study we developed an algorithm for point-probe optical spectroscopy without the colposcopic diagnosis that yielded operating characteristics with good performance and that has the potential for use in real time. The data suggest that research-grade point-probe devices lost little accuracy without using the colposcopic diagnosis in the algorithm. Clearly, a device for screening or diagnosis for the low- and middle-income countries will need to be effective in the hands of health care workers with less training.

We observed a statistical difference between the by-patient accuracy of spectroscopy with and without the colposcopic diagnosis. The ROC curves were similar over the range of high specificity, and the ROC curve for spectroscopy with the colposcopic diagnosis had notable improvement in the range of high sensitivity (Fig. 2). Specifically, the omission of the colposcopic diagnosis had decreased specificity for the range of high sensitivity. The by-site ROC curves were not statistically significantly different (Fig. 3).

We observed a decrease in performance when we added the first-generation device and broke analysis down by screening and diagnostic populations. Future studies will explore this decrease in performance. We hypothesize that it could be due to differences in the devices or differences in the Vancouver/Houston populations.

Spectroscopy alone (without the colposcopic diagnosis) generally had similar accuracy to colposcopy alone. In Fig. 4, about half of the points are above/or to the left of the diagonal line, meaning that among the patients seen by each colposcopist, spectroscopy had similar accuracy to the colposcopic diagnosis. Assuming that a colposcopist's accuracy is correlated with accurate detection of regions of interest on the cervix and, hence, accurate placement of the probe on a diseased site, the colposcopist's expertise was expected to have some effect on the accuracy of spectroscopy. We observed a slight upward trend ($p = 0.07$) suggesting that the accuracy of the colposcopist might influence the accuracy of the probe through the probe placement. However, the accuracies of spectroscopy were relatively high (median 0.88, range 0.59 to 1.00), and the confidence intervals overlapped considerably. Further follow-up is needed, but this shows promise for the use of spectroscopy in low-resource settings where expert colposcopists are not widely available, and for the possible use of methods of probe placement that rely less on precise identification of abnormalities prior to spectroscopic measurement.

The colposcopic tissue type was used as a covariate in the algorithm and is easily identifiable with limited training. Basic training in performing a Papanicolaou smear involves 1) understanding that the cervix is composed of two tissues (squamous and columnar) and 2) obtaining a sample of both ectocervix and endocervix. The "transformation zone," the area in which the squamous epithelium is constantly, throughout life, growing over the columnar epithelium, is the area most at risk for the location of lesions. This is thought due to the continuous growth of tissue and the turnover of cells, leading to increased risk of mitotic events. The most adequate Papanicolaou specimens contain cells from both the endocervical canal and the ecto-cervical area of the transformation zone. Any health care worker who learns how to perform cervical screening with Papanicolaou smear or VIA is taught to be aware of both tissue types and the importance of endo- and ecto-cervical sampling.

We believe there is a role for optical technologies in the high-, middle-, and low-income settings. There remain large expenditures in the high-income setting for the evaluation of minimally abnormal lesions. The role in the high-income setting is to decrease costs. The role in low- and middle-income country is to replace infrastructure on two levels: the need for trained expertise and to help with low adherence to follow-up.

Costs can be decreased in high-income settings such as the United States, Canada, and Europe by implementing a few principal changes to our current screening and diagnostic system: 1) decreasing the number of unnecessary treatments; 2) improving the performance of the current standard of care tests; 3) improving adherence; 4) increasing the HPV vaccine uptake; and 5) performing screening in high-risk populations that have previously never been screened.

The high number of unnecessary treatments is due to over-evaluation of unimportant lesions that are not likely to progress to cancer. There is abundant data that atypical and low-grade Papanicolaou smears are overevaluated. Kurman estimated that \$6 billion were spent annually on the evaluation of atypical cells of uncertain significance (ASCUS) and low-grade squamous intraepithelial lesions. Improving the accuracy of discriminating between high-grade and ASCUS lesions can decrease costs substantially.

The second change is to improve the performance of our screening and diagnostic tests. There is a wide variation in the performance of colposcopy, cytology, and histopathology. The sensitivity of colposcopy in the ALTS multicenter study was 0.37 (95% CI 0.33 to 0.42) and specificity 0.90 (95% CI 0.89 to 0.91). Moreover, the Kappa statistic comparing colposcopists in the trial was 0.36 (95% CI 0.33 to 0.39). Benedet et al. documented the performance of colposcopy in Vancouver, British Columbia, Canada, and showed a sensitivity of 90% and specificity of 50% for the diagnosis of high-grade lesions.⁶⁴ Similarly, in our trials, the sensitivity for the diagnosis of high-grade lesions in the diagnostic population was found to be 98% and the specificity 45%.⁵² The ALTS trial also reported wide variation in the performance of cytology, colposcopy, and the reading of histopathologic biopsies (low agreement amongst providers and sites). Automation can improve the performance by making a test reproducible.

There is low adherence in the United States and Canada. The “no-show” rate in the colposcopy clinics in both countries is 50%. Though there are many reasons for this low adherence rate in the high-income countries, providers must spend time and resources contacting patients because of their ethical duty to follow-up abnormalities, established guidelines for care, and to mitigate the other cost of health care in the United States (malpractice and litigation).

Despite the introduction of an HPV vaccine in the United States, Canada, and Europe, screening is still recommended for prevention of cervical cancer since the vaccines do not confer protection against all types of HPV that cause cervical cancer.⁶⁵⁻⁶⁹ The current HPV vaccines cover 70 to 74% of all HPV types that cause cancer globally. The vaccine costs approximately \$360 in the United States, requires refrigeration, and is administered over three doses, making it difficult to deliver the vaccine in the developing world, where it could have the most impact. The vaccine should be given to men and women to have an impact on a sexually transmitted disease and ideally would cover the high-risk HPV types most prevalent in the population being vaccinated.⁷⁰ New data concerning

cross-reactivity shows that infection with HPV types 52 and 58 are not well covered by the current vaccine, meaning the current vaccine does not provide sufficient protection from these types.⁷¹

All these findings led the U.S. Preventive Services Task Force to conclude that the most cost-effective use of dollars in the United States would be used to screen women who have not previously been screened. To address these issues will require identifying these women and studying why they have not been screened. A cost-effective device with high accuracy for diagnosis would have a role in decreasing costs in this setting.

The role of optical technologies in low- and middle-income countries is to replace infrastructure on two levels: the need for trained expertise and to help with low adherence to follow-up. Creating an inexpensive infrastructure (due to the lack of trained health care workers, cancer registries, cytologic expertise, histopathologic expertise, and colposcopy expertise) and replacing the need for adherence since patients may be seen only once in their lifetime in the current paradigm of care.

A principal challenge to the implementation of large-scale cancer screening programs in the developing world is the shortage of highly trained medical practitioners. Effective cervical screening programs would benefit greatly from automation that reduce or obviate the need for expert colposcopists. The screening model that our future studies will evaluate involves the use of a multispectral digital colposcope (MDC) to create an automated or semi-automated method to direct the placement of a point probe. The spectroscopic probe would then be used to interrogate the selected area(s) of tissue. This study is a first step toward evaluating the effectiveness of optical spectroscopy without the use by an expert colposcopist by excluding the colposcopic diagnosis in the spectroscopy algorithm and by exploring the influence of the colposcopist accuracy and, hence, accuracy of probe placement on the accuracy of the spectroscopy measurements attained by that colposcopist. There remain challenges with the HPV vaccine for deployment in the low resource setting: coverage of HPV types, cost, refrigeration, and adherence to three visits.

In the developing world, patient adherence to follow-up monitoring is even more of a challenge than in the developed world. A see-and-treat strategy is feasible with a combined MDC-probe device. Additionally, improving the sensitivity of the screening and diagnostic process, while maintaining a relatively high specificity, would benefit these populations.

The goal of our future studies is to use a MDC to direct the probe placement on the cervix and then use a point probe to predict the histology of the patient. The results of this study suggest that the skilled placement of the probe does not have a significant effect on the accuracy of spectroscopy and, hence, shows promise for the feasibility of being able to implement spectroscopy using an automated or semi-automated MDC algorithm for the selection of tissue sites to measure using the point probe.

A major strength of this study is that we have performed biopsies of all sites, and thus we have the gold standard for all patients' diagnoses, which alleviates the problem of verification bias. A study limitation is the difficulty to quantify the importance of the probe placement in the spectroscopy algorithm as we still do not have automated algorithms for probe placement. Each provider's probe placement accuracy was estimated based on his colposcopic diagnosis accuracy,

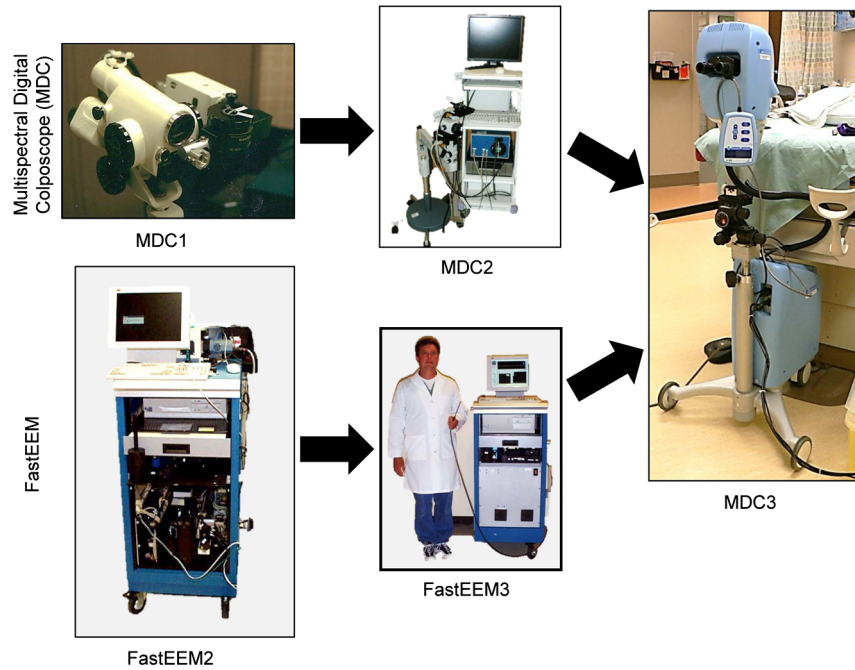


Fig. 8 Evolution of cancer imaging devices for high-resource settings: point-probe spectroscopy (FastEEM) and multispectral digital colposcope (MDC). The FastEEM2 and FastEEM3 are the first- and second-generation devices, respectively.

though the precision of these estimates varied depending on the number of patients seen and the population being evaluated. Future planned trials will address this. The weakness for deployment in the low- and middle-income setting is that the study was conducted in the United States and Canada by expert colposcopists.

Further research is needed to evaluate the possible clinical applications of this new technology. For example, a negative spectroscopy result could lead to the elimination of an unneeded biopsy. Alternatively, a positive spectroscopy result that confirms a positive clinical test, if used in a see-and-treat strategy, would also lead to the elimination of a biopsy and avoid a further clinical visit. The spectroscopic device described in this paper could be used in a clinic to replace biopsies and permit diagnosis and treatment in a single visit. We hope this model could be a great benefit to patients in developing countries or underserved areas of the developed world and reduce costs in the developed countries.

We plan on training an algorithm for the MDC and correlating it with the spectroscopic probe to develop a joint algorithm for the two devices. We will then use this information to construct a lower-cost device that uses both technologies. The combination of the MDC with the spectroscopic point probe could potentially be used by nonexperts, thus addressing that limitation of our study.

The device of choice in Nigeria and other resource-challenged settings of Africa, Asia, Oceania, and Latin America should be battery-powered, low cost, be able to distinguish infection from neoplasia and view the entire cervix. We are developing a Diagnostic Imaging Aid (DIA) optimized for use in low-resource settings based on the MDC.⁷²⁻⁷⁵ Figures 8 and 9 show the evolution of cancer imaging devices for high-resource settings and proposed devices for low-resource settings, respectively.

A commercial device that had both multispectral images of the cervix and point probe measurement of specific sites would

provide the best of both approaches. Continuing development of classification algorithms for use in wide-field and point-probe devices could produce a combined screening and diagnostic system with the potential to change the way we control cervical cancer. Research has suggested that patients prefer screening and diagnosis using optical spectroscopy when compared with conventional cytology.⁷⁶ We hope that forthcoming developments could bring about a cost-effective and clinically advantageous cervical cancer control model using a combination of vaccination and minimally invasive optical technologies.

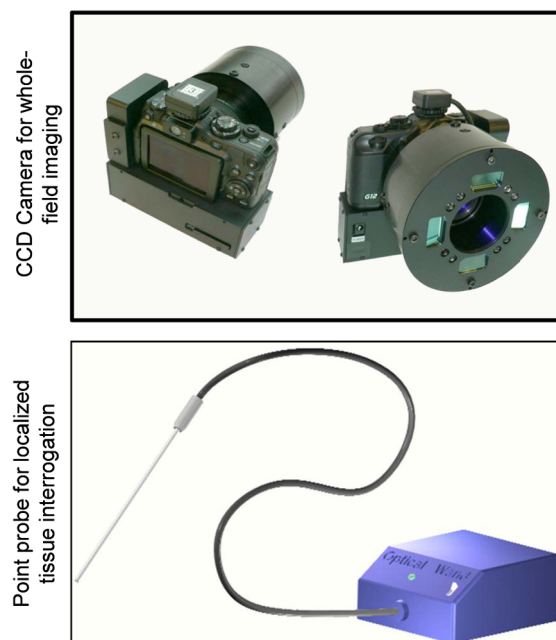


Fig. 9 Proposed cancer-imaging devices for low-resource settings.

Acknowledgments

Financial support for this study was provided entirely by grant number P01-CA-82710-09 from the National Cancer Institute and the National Biomedical Imaging Branch. The funding agreement ensured the authors' independence in designing the study, interpreting the data, and writing and publishing the report. Calum MacAulay, Dennis Cox, Michele Follen, and E. Neely Atkinson all hold some patents that Remicalm, L.L.C., has licensed from their respective past and present universities, including The British Columbia Cancer Research Centre, Rice University, and The University of Texas MD Anderson Cancer Center. Drs. Yamal, Cantor, Davies, Adewole, Buys, and Mr. Zewdie have no agreements, jointly held patents, nor other potential conflicts of interest with regard to this research.

References

- D. M. Parkin et al., "Global cancer statistics, 2002," *CA. Cancer J. Clin.* **55**(2), 74–108 (2005).
- J. Ferlay et al., "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008," *Int. J. Cancer* **127**(12), 2893–2917 (2010).
- J. Sherris, C. Herdman, and C. Elias, "Cervical cancer in the developing world," *West J. Med.* **175**(4), 231–233 (2001).
- R. Sankaranarayanan, A. M. Budukh, and R. Rajkumar, "Effective screening programmes for cervical cancer in low- and middle-income developing countries," *Bull. World Health Organ.* **79**(10), 954–962 (2001).
- World Health Organization, "Sexual and reproductive health—cancer of the cervix," <http://www.who.int/reproductivehealth/topics/cancers/en/> (2011).
- P. Sparen et al., "Improved control of invasive cervical cancer in Sweden over six decades by earlier clinical detection and better treatment," *J. Clin. Oncol.* **13**(3), 715–725 (1995).
- F. B. Stehman et al., "Innovations in the treatment of invasive cervical cancer," *Cancer* **98**(S9), 2052–2063 (2003).
- J. Cuzick et al., "Overview of human papillomavirus-based and other novel options for cervical cancer screening in developed and developing countries," *Vaccine* **26**(Suppl 10) K29–41 (2008).
- R. Sankaranarayanan et al., "HPV screening for cervical cancer in rural India," *N. Engl. J. Med.* **360**(14), 1385–1394 (2009).
- G. M. Clifford et al., "Worldwide distribution of human papillomavirus types in cytologically normal women in the International Agency for Research on Cancer HPV prevalence surveys: a pooled analysis," *Lancet* **366**(9490), 991–998 (2005).
- G. M. Clifford et al., "Human papillomavirus types in invasive cervical cancer worldwide: a meta-analysis," *Br. J. Cancer* **88**(1), 63–73 (2003).
- B. Lu et al., "Efficacy and safety of prophylactic vaccines against cervical HPV infection and diseases among women: a systematic review & meta-analysis," *BMC Infect. Dis.* **11**(13) (2011).
- J. H. No et al., "Human papillomavirus vaccine: widening the scope for cancer prevention," *Mol. Carcinog.* **50**(4), 244–253 (2011).
- R. Sankaranarayanan et al., "Effect of visual screening on cervical cancer incidence and mortality in Tamil Nadu, India: a cluster-randomised trial," *Lancet* **370**(9585), 398–406 (2007).
- R. Sankaranarayanan et al., "Visual inspection of the uterine cervix after the application of acetic acid in the detection of cervical carcinoma and its precursors," *Cancer* **83**(10), 2150–2156 (1998).
- R. Sankaranarayanan et al., "Test characteristics of visual inspection with 4% acetic acid (VIA) and Lugol's iodine (VILI) in cervical cancer screening in Kerala, India," *Int. J. Cancer* **106**(3), 404–408 (2003).
- S. J. Goldie et al., "Cost-effectiveness of cervical-cancer screening in five developing countries," *N. Engl. J. Med.* **353**(20), 2158–2168 (2005).
- R. Sankaranarayanan et al., "A critical assessment of screening methods for cervical neoplasia," *Int. J. Gynaecol. Obstet.* **89**(Suppl 2), S4–S12 (2005).
- R. Legood et al., "Screening for cervical cancer in India: how much will it cost? A trial based analysis of the cost per case detected," *Int. J. Cancer* **117**(6), 981–987 (2005).
- N. Thekkekk and R. Richards-Kortum, "Optical imaging for cervical cancer detection: solutions for a continuing global problem," *Nat. Rev. Cancer* **8**(9), 725–731 (2008).
- R. L. Cook et al., "Factors associated with initiation and completion of human papillomavirus vaccine series among young women enrolled in Medicaid," *J. Adolesc. Health* **47**(6), 596–599 (2010).
- K. T. Schomacker et al., "Ultraviolet laser-induced fluorescence of colonic polyps," *Gastroenterology* **102**(4 Pt 1), 1155–1160 (1992).
- T. Vo-Dinh et al., "In vivo cancer diagnosis of the esophagus using differential normalized fluorescence (DNF) indices," *Lasers Surg. Med.* **16**(1), 41–47 (1995).
- M. Anidjar et al., "The role of laser-induced autofluorescence spectroscopy in bladder tumor detection. Dependence on the excitation wavelength," *Ann. N. Y. Acad. Sci.* **838**(1), 130–142 (1998).
- S. Lam et al., "Detection of dysplasia and carcinoma in situ by ratio fluorometry," *Am. Rev. Respir. Dis.* **146**(6), 1458–1461 (1992).
- L. Burke et al., "Use of autofluorescence of cells to evaluate cervical neoplasia," *J. Gynecol. Tech.* **2**(90), 187–190 (1996).
- L. Burke et al., "Identification of squamous intraepithelial lesions: fluorescence of cervical tissue during colposcopy," *J. Low. Genit. Tract. Dis.* **3**(3), 159–162 (1999).
- S. K. Chang et al., "Fluorescence spectroscopy for cervical precancer detection: is there variance across the menstrual cycle?," *J. Biomed. Opt.* **7**(4), 595–602 (2002).
- S. K. Chang et al., "Combined reflectance and fluorescence spectroscopy for in vivo detection of cervical pre-cancer," *J. Biomed. Opt.* **10**(2), 024031 (2005).
- T. DeSantis et al., "Spectroscopic imaging as a triage test for cervical disease: a prospective multicenter clinical trial," *J. Low. Genit. Tract. Dis.* **11**(1), 18–24 (2007).
- D. G. Ferris et al., "Multimodal hyperspectral imaging for the noninvasive diagnosis of cervical neoplasia," *J. Low. Genit. Tract. Dis.* **5**(2), 65–72 (2001).
- J. A. Freeberg et al., "The performance of fluorescence and reflectance spectroscopy for the in vivo diagnosis of cervical neoplasia; point probe versus multispectral approaches," *Gynecol. Oncol.* **107**(1 Suppl 1), S248–255 (2007).
- I. Georgakoudi et al., "NAD(P)H and collagen as in vivo quantitative fluorescent biomarkers of epithelial precancerous changes," *Cancer Res.* **62**(3), 682–687 (2002).
- W. K. Huh et al., "Optical detection of high-grade cervical intraepithelial neoplasia in vivo: results of a 604-patient study," *Am. J. Obstet. Gynecol.* **190**(5), 1249–1257 (2004).
- Y. N. Mirabal et al., "Reflectance spectroscopy for in vivo detection of cervical precancer," *J. Biomed. Opt.* **7**(4), 587–594 (2002).
- J. R. Mourant et al., "In vivo light scattering measurements for detection of precancerous conditions of the cervix," *Gynecol. Oncol.* **105**(2), 439–445 (2007).
- R. J. Nordstrom et al., "Identification of cervical intraepithelial neoplasia (CIN) using UV-excited fluorescence and diffuse-reflectance tissue spectroscopy," *Lasers Surg. Med.* **29**(2), 118–127 (2001).
- I. M. Orfanoudaki et al., "A clinical study of optical biopsy of the uterine cervix using a multispectral imaging system," *Gynecol. Oncol.* **96**(1), 119–131 (2005).
- M. Parker et al., "Hyperspectral diagnostic imaging of the cervix: report on a new investigational device," *J. Low. Genit. Tract. Dis.* **4**(3), 119–124 (2000).
- N. Ramanujam et al., "Fluorescence spectroscopy: a diagnostic tool for cervical intraepithelial neoplasia (CIN)," *Gynecol. Oncol.* **52**(1), 31–38 (1994).
- N. Ramanujam et al., "Cervical precancer detection using a multivariate statistical algorithm based on laser-induced fluorescence spectra at multiple excitation wavelengths," *Photochem. Photobiol.* **64**(4), 720–735 (1996).
- R. Drezek et al., "Autofluorescence microscopy of fresh cervical-tissue sections reveals alterations in tissue biochemistry with dysplasia," *Photochem. Photobiol.* **73**(6), 636–641 (2001).
- R. Drezek et al., "Understanding the contributions of NADH and collagen to cervical tissue fluorescence spectra: modeling, measurements, and implications," *J. Biomed. Opt.* **6**(4), 385–396 (2001).
- I. Georgakoudi et al., "Trimodal spectroscopy for the detection and characterization of cervical precancers in vivo," *Am. J. Obstet. Gynecol.* **186**(3), 374–382 (2002).
- D. Arifler et al., "Light scattering from collagen fiber networks: micro-optical properties of normal and neoplastic stroma," *Biophys. J.* **92**(9), 3260–3274 (2007).

46. C. K. Brookner et al., "Safety analysis: relative risks of ultraviolet exposure from fluorescence spectroscopy and colposcopy are comparable," *Photochem. Photobiol.* **65**(6), 1020–1025 (1997).
47. R. Drezek et al., "Light scattering from cervical cells throughout neoplastic progression: influence of nuclear morphology, DNA content, and chromatin texture," *J. Biomed. Opt.* **8**(1), 7–16 (2003).
48. J. A. Freeberg et al., "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.* **12**(3), 034015 (2007).
49. A. Malpica 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.* **99**(3 Suppl 1), S38–52 (2005).
50. J. M. Levitt et al., "Intrinsic fluorescence and redox changes associated with apoptosis of primary human epithelial cells," *J. Biomed. Opt.* **11**(6), 064012 (2006).
51. S. B. Cantor et al., "Accuracy of optical spectroscopy for the detection of cervical intraepithelial neoplasia: testing a device as an adjunct to colposcopy," *Int. J. Cancer* **128**(5), 1151–1168 (2011).
52. S. B. Cantor et al., "Accuracy of colposcopy in the diagnostic setting compared with the screening setting," *Obstet. Gynecol.* **111**(1), 7–14 (2008).
53. S. B. Cantor et al., "Cost-effectiveness analysis of diagnosis and management of cervical squamous intraepithelial lesions," *Obstet. Gynecol.* **91**(2), 270–277 (1998).
54. M. Cardenas-Turanzas et al., "See-and-treat strategy for diagnosis and management of cervical squamous intraepithelial lesions," *Lancet Oncol.* **6**(1), 43–50 (2005).
55. N. M. Marin et al., "Calibration standards for multicenter clinical trials of fluorescence spectroscopy for in vivo diagnosis," *J. Biomed. Opt.* **11**(1), 014010 (2006).
56. C. K. Brookner et al., "Autofluorescence patterns in short-term cultures of normal cervical tissue," *Photochem. Photobiol.* **71**(6), 730–736 (2000).
57. D. D. Cox et al., "Detecting the signal of the menstrual cycle in fluorescence spectroscopy of the cervix," *Appl. Spectrosc.* **57**(1), 67–72 (2003).
58. C. Macaulay et al., "Variation of fluorescence spectroscopy during the menstrual cycle," *Opt. Express* **10**(12), 493–504 (2002).
59. N. Ramanujam et al., "Low temperature fluorescence imaging of freeze-trapped human cervical tissues," *Opt. Express* **8**(6), 335–343 (2001).
60. T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Science +Business Media, New York, NY (2009).
61. M. S. Pepe, *The Statistical Evaluation of Medical Tests for Classification and Prediction*, Oxford University Press, New York (2004).
62. X. Robin et al., "pROC: an open-source package for R and S+ to analyze and compare ROC curves," *BMC Bioinformatics* **12**, 77 (2011).
63. Y. Jiang, C. Metz, and R. Nishikawa, "A receiver operating characteristic partial area index for highly sensitive diagnostic tests," *Radiology* **201**(3), 745–750.
64. J. L. Benedet, J. P. Maticic, and M. A. Bertrand, "The quality of community colposcopic practice," *Obstet. Gynecol.* **103**(1), 92–100 (2004).
65. J. M. Agosti and S. J. Goldie, "Introducing HPV vaccine in developing countries—key challenges and issues," *N. Engl. J. Med.* **356**(19), 1908–1910 (2007).
66. L. R. Baden et al., "Human papillomavirus vaccine—opportunity and challenge," *N. Engl. J. Med.* **356**(19), 1990–1991 (2007).
67. C. Dailard, "HPV in the United States and Developing Nations: A Problem of Public Health and Politics," *The Guttmacher Report on Public Policy* **6**(3), 4–15 (2003).
68. S. J. Goldie et al., "Projected clinical benefits and cost-effectiveness of a human papillomavirus 16/18 vaccine," *J. Natl. Cancer Inst.* **96**(8), 604–615 (2004).
69. G. D. Sanders and A. V. Taira, "Cost-effectiveness of a potential vaccine for human papillomavirus," *Emerg. Infect. Dis.* **9**(1), 37–48 (2003).
70. "Cheaper HPV vaccines needed," *Lancet* **371**(9625), 1638 (2008).
71. M. van den Hende et al., "Evaluation of immunological cross-reactivity between clade A9 high-risk human papillomavirus types on the basis of E6-Specific CD4+ memory T cell responses," *J. Infect. Dis.* **202**(8), 1200–1211 (2010).
72. J. Benavides et al., "Multispectral digital colposcopy for in vivo detection of cervical cancer," *Opt. Express* **11**(10), 1223–1236 (2003).
73. A. Milbourne et al., "Results of a pilot study of multispectral digital colposcopy for the in vivo detection of cervical intraepithelial neoplasia," *Gynecol. Oncol.* **99**(3 Suppl 1), S67–75 (2005).
74. S. Nakappan et al., "Methodology of a real-time quality control for the multispectral digital colposcope (MDC)," *Gynecol. Oncol.* **107**(1 Suppl 1), S215–222 (2007).
75. S. Y. Park et al., "Automated image analysis of digital colposcopy for the detection of cervical neoplasia," *J. Biomed. Opt.* **13**(1), 014029 (2008).
76. K. Basen-Engquist et al., "Patient distress and satisfaction with optical spectroscopy in cervical dysplasia detection," *Am. J. Obstet. Gynecol.* **189**(4), 1136–1142 (2003).