Optimal Excitation Wavelengths for *In Vivo* Detection of Oral Neoplasia Using Fluorescence Spectroscopy¶

Douglas L. Heintzelman[1], Urs Utzinger[1], Holger Fuchs[1], Andres Zuluaga[1], Kirk Gossage[1], Ann M. Gillenwater[2], Rhonda Jacob[2], Bonnie Kemp[3] and Rebecca R. Richards-Kortum*[1]

[1]Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX and Departments of [2]Head and Neck Surgery and [3]Pathology, University of Texas M. D. Anderson Cancer Center, Houston, TX

ABSTRACT

There is no satisfactory mechanism to detect premalignant lesions in the upper aero-digestive tract. Fluorescence spectroscopy has potential to bridge the gap between clinical examination and invasive biopsy; however, optimal excitation wavelengths have not yet been determined. The goals of this study were to determine optimal excitation–emission wavelength combinations to discriminate normal and precancerous/cancerous tissue, and estimate the performance of algorithms based on fluorescence. Fluorescence excitation–emission matrices (EEM) were measured *in vivo* from 62 sites in nine normal volunteers and 11 patients with a known or suspected premalignant or malignant oral cavity lesion. Using these data as a training set, algorithms were developed based on combinations of emission spectra at various excitation wavelengths to determine which excitation wavelengths contained the most diagnostic information. A second validation set of fluorescence EEM was measured *in vivo* from 281 sites in 56 normal volunteers and three patients with a known or suspected premalignant or malignant oral cavity lesion. Algorithms developed in the training set were applied without change to data from the validation set to obtain an unbiased estimate of algorithm performance. Optimal excitation wavelengths for detection of oral neoplasia were 350, 380 and 400 nm. Using only a single emission wavelength of 472 nm, and 350 and 400 nm excitation, algorithm performance in the training set was 90% sensitivity and 88% specificity and in the validation set was 100% sensitivity, 98% specificity. These results suggest that fluorescence spectroscopy can provide a simple, objective tool to improve *in vivo* identification of oral cavity neoplasia.

INTRODUCTION

Approximately 30 000 new cases of oral cancer occur annually in the United States, resulting in over 8000 deaths each year (1). Eighty-five percent of malignancies occurring in this anatomic region are squamous cell carcinomas (SCC)† (2). Despite tremendous advances in ablative and reconstructive techniques for treating SCC of the oral cavity, there has been little increase in survival rates over the past 30 years. Patients with cancer of the oral cavity usually present when their disease is already advanced. Overall, 5-year survival rates for patients with oral cavity SCC remain around 50% and are even less for patients with advanced disease (3). Treatment for patients with advanced disease is more disfiguring and debilitating, more expensive and more prone to failure. Those patients who do survive their initial cancer are at high risk to develop a second primary tumor (4). These patients need continuous, close follow-up for early detection of second primaries.

Early detection of neoplastic changes in the oral cavity may be the best method to improve patient quality of life and survival rates. Certain discreet lesions have been identified clinically that have potential for malignant conversion. These include oral leukoplakia (white plaques) (5–7) and erythroplakia (velvety, reddish mucosal lesions) (8). Despite the easy accessibility of the oral cavity to examination, there is no satisfactory mechanism to adequately screen and detect premalignant changes and early lesions in the upper aero-digestive tract. Several factors contribute to the difficulty in developing effective early detection tools. (1) Inexperienced practitioners often fail to recognize the subtle changes indicative of early dysplastic or neoplastic transformation. It is difficult to distinguish premalignant lesions from more common benign inflammatory conditions in the general population. (2) Practitioners and patients are often reluctant to perform invasive biopsies or repeated biopsies of oral lesions in this situation where the expected yield is low; and (3) in patients at high risk for malignancy, often the whole lining of the oral cavity is at risk or has premalignant changes. Even for experienced clinicians, it is difficult to know when and where to biopsy and the entire mucosal lining would have to be removed to biopsy all areas at risk.

¶Posted on the web 17 May 2000.

*To whom correspondence should be addressed at: Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712, USA. Fax: 512-475-8854; e-mail: kortum@mail.utexas.edu

†*Abbreviations:* DMBA, 7,12-dimethylbenz(α)anthracene; EEM, excitation–emission matrix; ESL, eigenvector significance level; FOM, floor of mouth; LIFE, light-induced fluorescence endoscopy; MRA, mucosal reactive atypia; SCC, squamous cell carcinoma; WLE, white-light endoscopy.

The standard method for oral cavity screening relies heavily on the practitioners' clinical experience in the recognition of suspicious lesions during physical examination. Several studies have been conducted to assess the ability of vital staining with agents such as toluidine blue and Lugol's iodine to improve diagnostic accuracy (9–12). Although sensitivity‡ approximates 90% or greater in many of these trials, specificity is lower. Importantly, most of these studies were conducted by clinicians who are experts in the diagnosis of oral cavity malignancies, and may not reflect the diagnostic performance of these agents in the hands of less experienced personnel.

The development of a noninvasive and accurate method for real-time screening and diagnosis of oral cavity lesions would have great potential to improve early detection of neoplastic changes, and thereby improve the quality of life and survival rates for persons developing SCC of the oral cavity. Fluorescence spectroscopy is a new diagnostic modality with the potential to bridge the gap between clinical examination and invasive biopsy. Tissue architecture and biochemical composition can be evaluated in near real-time using optical spectroscopy without the need for tissue removal (13,14). This technique has shown promise in the detection of intraepithelial neoplasias in the cervix (15), the colon (16) and the oral cavity (17–26).

Fluorescence spectra have been measured from normal and neoplastic areas of oral mucosa *in vivo* using fiber optic probes in several small clinical series (17,18,23,24). Schantz *et al.* (17) measured fluorescence excitation spectra at 450 nm emission from lesions and contralateral normal sites in 35 patients with untreated oral neoplasia using a fiber optic probe. They found that the average maximum fluorescence intensity of poorly differentiated tumors was significantly lower than those of well or moderately differentiated tumors, although there was considerable overlap in the distributions of fluorescence intensities associated with these diagnostic categories.

It is well known that the choice of excitation and emission wavelengths determines which tissue chromophores contribute to the resulting spectrum and strongly affect the shape and intensity of fluorescence spectra (13,14). A number of groups have compared the diagnostic performance of various combinations of excitation and emission wavelengths to determine which combination provides the most effective diagnostic performance.

Kolli *et al.* (18) measured fluorescence excitation spectra at 380 and 450 nm emission and fluorescence emission spectra at 300 and 340 nm excitation from oral cavity lesions and contralateral normal sites in 31 patients with untreated oral neoplasia using a fiber probe. They found that differences in the mean ratios of intensities in the fluorescence of oral cavity neoplasia and contralateral normal sites were statistically significant for emission spectra at 300 and 340 nm excitation and excitation spectra at 380 nm emission but not at 450 nm emission.

Other groups have measured fluorescence emission spectra at many excitation wavelengths from biopsy specimens

‡Sensitivity is defined as the proportion of people with disease who have a positive test; specificity is defined as the proportion of people without disease who have a negative test.

in vitro to determine which excitation wavelengths contain the most diagnostic information. Chen *et al.* (19) measured fluorescence emission spectra at 270–400 nm excitation in 10 nm steps *in vitro* from normal and neoplastic oral tissues. At 300 nm excitation, emission spectra exhibited peaks at 330 and 470 nm emission. The average ratio of fluorescence intensities at 330 nm emission to that at 470 nm emission for malignant and premalignant oral samples was significantly greater than that for the normal oral cavity ($P < 0.01$).

In two series, Roy *et al.* (20) and Ingrams *et al.* (21) measured fluorescence emission spectra at multiple excitation wavelengths from biopsies of normal, dysplastic and malignant oral cavity sites. In both series, differences were most marked at 410 nm excitation, with abnormal samples showing enhanced red fluorescence at 635 nm emission (20,21). Using this wavelength, fluorescence could be used to correctly diagnose 20 of 22 specimens (21). Based on these results, this group measured fluorescence spectra *in vivo* from an animal model (22) as well as *in vivo* in human subjects (23). Fluorescence emission spectra were measured at 410 nm excitation from 7,12-dimethylbenz(α)anthracene (DMBA)-induced precancers and early cancers in the hamster cheek pouch *in vivo*. Neoplastic lesions showed characteristic fluorescence between 630 and 640 nm emission (22). Using this as a diagnostic criterion, 45 of 49 lesions were correctly diagnosed, including early dysplasias (22). Fluorescence emission spectra were measured from 19 untreated lesions and contralateral normal sites in 13 patients and 10 normal volunteers at 370 and 410 nm excitation (23). Differences in the fluorescence of normal and neoplastic sites were more obvious at 410 nm excitation. Again using the increase in red fluorescence as a diagnostic algorithm, 17 of 19 lesions could be correctly diagnosed with two false positive results.

Our group measured fluorescence emission spectra *in vivo* using a fiber optic probe at 337, 365 and 410 nm excitation from 95 sites in eight normal volunteers and 45 sites in 15 patients with premalignant or malignant oral cavity lesions (24). At 337 nm excitation, the fluorescence intensity of contralateral normal sites was greater than that of abnormal sites. At 410 nm excitation the ratio of red- to blue fluorescence was greater in abnormal areas than in contralateral normal areas. A diagnostic algorithm based on these two differences achieved a sensitivity of 94% and a specificity of 100%, compared to 76% sensitivity and 100% specificity for clinical impression.

Other groups have demonstrated that imaging systems that record the spatial distribution of fluorescence intensity at specific excitation–emission wavelength combinations can be used to survey large areas of oral cavity mucosa for neoplastic changes. Kulapaditharom and Boonkitticharoen (25) used the light-induced fluorescence endoscopy (LIFE) system to image the red- to green fluorescence intensity ratio at 442 nm excitation to identify areas of oral cavity neoplasia; results were compared to traditional white-light endoscopy (WLE) in 25 patients suspected to have oral cavity malignancy. Oral cavity lesions exhibited an increased red/green fluorescence intensity ratio. Using the LIFE system resulted in a detection rate of 100% with a specificity of 87.5%. WLE achieved a lower detection rate of 87.5% and a lower specificity of 50%. Similarly, Onizawa *et al.* (26)

showed that fluorescence photography at 360 nm excitation, with emission above 480 nm could be used to separate benign and malignant oral cavity tumors. Of the 16 malignant tumors 14 exhibited increased orange fluorescence, while only one of the 16 benign lesions showed increased orange fluorescence, resulting in 88% sensitivity and 94% specificity.

These studies indicate that fluorescence spectroscopy has the potential to improve screening and detection of early oral cavity neoplasia. However, an important limitation of this previous work is that the choice of excitation and emission wavelengths has not been fully optimized. In general, studies that have surveyed the fluorescence of normal and neoplastic tissues over large wavelength regions have been carried out *in vitro* using small biopsy specimens. There are significant differences between fluorescence measurements made *in vitro* and *in vivo*. These differences arise from the oxidation of electron carriers (*e.g.* reduced nicotinamide adenine dinucleotide, reduced flavin adenine dinucleotide) (16,27), lack of blood flow to the specimen (28) and the small size of biopsies (29). The loss of perfusion alters the contributions of hemoglobin absorption to fluorescence spectra. Due to multiple scattering in tissue, some fluorescence may escape from the sides and bottom of small tissue specimens such as biopsies; these losses can be significant, and will affect the fluorescence lineshape measured from small samples (29). Once optimal wavelengths are determined, inexpensive imaging systems that record fluorescence at a small number of excitation and emission wavelength combinations could be used to survey large areas of oral cavity mucosa for early neoplastic changes.

The goal of the clinical study described in this paper was to address these limitations and determine the fluorescence excitation–emission wavelength combinations which result in data that contain the most diagnostic information for the detection of neoplasia in the oral cavity. We measured fluorescence excitation–emission matrices (EEM) at 18 excitation wavelengths ranging from 330 to 500 nm and fluorescence emission wavelengths from 340 to 700 nm from 62 sites in 11 patients and nine normal volunteers. We developed a method to analyze fluorescence EEM to determine which excitation and emission wavelengths contain the most diagnostic information and to estimate the expected performance of diagnostic algorithms based on this information. We then measured fluorescence EEM in a second validation group of three patients and 53 normal volunteers. Algorithms developed using data from the first group of subjects (training set) were applied without change to the validation data set, yielding an unbiased estimate of algorithm performance. In this study, we found that the combination of full emission spectra from three excitation wavelengths resulted in a training set performance of 100% sensitivity and 88% specificity, while an increase to four excitation wavelengths did not significantly increase diagnostic performance. Using only a single emission wavelength of 472 nm, which is common to two excitation wavelengths (350 and 400 nm), yields a performance of 90% sensitivity, 88% specificity in the training set and 100% sensitivity, 98% specificity in the validation set.

MATERIALS AND METHODS

Study subjects. In the training phase of the study, nine normal volunteers and 11 patients with a known or suspected premalignant or malignant oral cavity lesion were recruited to participate in the study at the Head and Neck Surgery Clinic at The University of Texas M.D. Anderson Cancer Center. In the validation phase of the study, 53 normal volunteers and three patients with a known or suspected premalignant or malignant oral cavity lesion were recruited to participate. The study was reviewed and approved by the Institutional Review Board at The University of Texas at Austin and the Surveillance Committee at The M.D. Anderson Cancer Center. Written informed consent was obtained from each person in the study.

Instrument. The spectroscopic system used to measure fluorescence excitation–emission matrices has been described in detail previously (30). Briefly, the system measures fluorescence emission spectra at 18 excitation wavelengths, ranging from 330 to 500 nm in 10 nm increments with a spectral resolution of 7 nm. The system incorporates a fiber optic probe, a Xenon arc lamp coupled to a monochromator to provide excitation light and a polychromator and thermo-electrically cooled charge-coupled device camera to record fluorescence intensity as a function of emission wavelength. Data in the training phase of the study were obtained between June 1997 and January 1998; minor modifications were made to the spectroscopic instrumentation in January 1998 (31) and data for the validation phase were obtained from January 1998 to January 2000.

Calibration. As a negative control a background EEM was obtained with the probe immersed in a nonfluorescent bottle filled with distilled water at the beginning of each measurement day. Then a fluorescence EEM was measured with the probe placed on the surface of a quartz cuvette containing a solution of Rhodamine 610 (Exciton, Dayton, OH) dissolved in ethylene glycol (2 mg/mL) at the beginning of each patient measurement.

To correct for the nonuniform spectral response of the detection system, the spectra of two calibrated sources were measured at the beginning of the training and validation phases of the study; in the visible a National Institute of Standards and Technology (NIST)-traceable calibrated tungsten ribbon filament lamp was used and in the UV a deuterium lamp was used (550C and 45D, Optronic Laboratories Inc, Orlando, FL). Correction factors were derived from these spectra. Dark current–subtracted EEM from patients were then corrected for the nonuniform spectral response of the detection system. Variations in the intensity of the fluorescence excitation light source at different excitation wavelengths were corrected using measurements of the intensity at each excitation wavelength at the probe tip made using a calibrated photodiode (818-UV, Newport Research Corp., Irvine, CA). Finally, corrected fluorescence intensities from each site were divided by the fluorescence emission intensity of the Rhodamine standard at 460 nm excitation, 580 nm emission. Thus, data illustrated in this paper are not the absolute fluorescence intensities of tissue but rather given in calibrated intensity units relative to the Rhodamine standard.

Data acquisition. Before the probe was used it was disinfected with Metricide (Metrex Research Corp., Orange, CA) for 20 min. The probe was then rinsed with water and dried with sterile gauze. The disinfected probe was guided into the oral cavity and its tip positioned flush with the mucosa. Then the fluorescence EEM was measured. Measurement of each EEM required approximately 2 min.

In the training set, fluorescence EEM was measured from nine volunteers with no history of oral cavity neoplasia, at 41 clinically normal sites in the oral cavity. In the validation set, fluorescence EEM was measured from 53 volunteers with no history of oral cavity neoplasia, at 274 clinically normal sites in the oral cavity. No biopsies were obtained from volunteers. In the training set, fluorescence EEM was measured following visual screening from 47 sites in 17 patients with a known or suspected premalignant or malignant oral cavity lesion. In the validation set, fluorescence EEM was measured from seven sites in three patients. The examiner placed the fiber optic probe on a lesion or suspected lesion and the fluorescence of that site was measured. In addition to the three to five visually abnormal sites, fluorescence EEM was measured from one to three contralateral normal sites. After spectroscopy, abnormal sites were tattooed with India Ink, where the probe measured the spectra.

Table 1. Summary of sites in training and validation datasets and corresponding diagnosis

	No. of sites	No. of patients	Tongue	FOM	Buccal mucosa	Gingiva	Palate	Lip
Training set								
Total sites	62	20	37	8	7	4	1	5
Normals	52	20	32	5	7	3	0	5
Normal	35	9	25	1	4	—	—	5
Adjacent	17	11	7	4	3	3	—	—
Abnormals	10	9	5	3	0	1	1	0
Dysplasia								
Mild	1	1	—	1	—	—	—	—
Moderate	1	1	1	—	—	—	—	—
Severe	2	2	1	—	—	—	1	—
Cancer								
Invasive	4	3	2	1	—	1	—	—
Submucosal	1	1	1	—	—	—	—	—
Verrucus	1	1	—	1	—	—	—	—
Validation set								
Total sites	281	56	204	0	75	2	0	0
Normals	277	56	201	0	75	1	0	0
Normal	274	53	199	—	75	—	—	—
Adjacent	3	3	2	—	—	1	—	—
Abnormals	4	3	3	0	0	1	0	0
Dysplasia								
Mild	2	1	2	—	—	—	—	—
Moderate	0	0	—	—	—	—	—	—
Severe	0	0	—	—	—	—	—	—
Cancer								
Invasive	2	2	1	—	—	1	—	—
Submucosal	0	0	—	—	—	—	—	—
Verrucus	0	0	—	—	—	—	—	—

A clinical diagnosis of each lesion as normal, abnormal (not dysplastic), abnormal (dysplastic) or cancerous was recorded by an experienced head and neck surgeon (A.M.G.) or dental oncologist (R.J.). During subsequent surgery, a 2–4 mm biopsy of the tissue was taken from the tattooed area. These specimens were evaluated by an experienced pathologist using light microscopy and classified as normal, mucosal reactive atypia (MRA), dysplasia or cancer using standard diagnostic criterion. Biopsy specimens with multiple diagnoses were classified according to the most severe pathological diagnosis. The pathologist and clinicians were blinded to the results of the spectroscopic analyses.

Data review. In the training set, a total of 88 sites were measured from 26 subjects. All spectra were reviewed by a single investigator (D.L.H.) blinded to the pathologic results. Spectra were discarded if files were not saved properly due to software error (eight sites), instrument error (two sites), operator error (four sites), probe movement (three sites) and the presence of room light artifacts at wavelengths below 600 nm (three sites) in at least one of the emission spectra. From the remaining sites, spectra from six sites were excluded because the tattoo could not be located and consequently reliable histologic diagnosis was not available for these sites. Therefore, in the training set, fluorescence EEM from 62 sites from 20 subjects was available for further analysis (Table 1).

In the validation set, a total of 325 sites were measured from 56 subjects. All spectra were reviewed by a single investigator (K.G.) blinded to the pathologic results. Spectra were discarded if files were not saved properly due to operator error (five sites), probe movement or the presence of room light artifacts at wavelengths below 600 nm

(39 sites) in at least one of the emission spectra. Therefore, in the validation set, fluorescence EEM from 281 sites from 56 subjects was available for analysis (Table 1).

Data analysis. Fluorescence data in the training set were analyzed to determine which excitation and emission wavelengths contained the most diagnostically useful information and to estimate the performance of diagnostic algorithms based on this information. We considered algorithms based on multivariate discriminant analysis (15,32). First, we developed algorithms based on combinations of emission spectra at various excitation wavelengths in order to determine which excitation wavelengths contained the most diagnostic information. Then, at those excitation wavelengths, we evaluated spectra with reduced numbers of emission wavelengths to determine whether complete emission spectra were required or whether accurate diagnosis could be made using multispectral measurements at a few excitation–emission wavelength combinations. In each case, the algorithm-development process, described in detail below, consisted of the following major steps: (1) data preprocessing to reduce interpatient variations; (2) data reduction to reduce the dimensionality of the data set; (3) feature selection and classification to develop algorithms which maximized diagnostic performance and minimized the likelihood of overtraining in a training set; and (4) evaluation of these algorithms using the technique of cross-validation.

Multivariate discriminant algorithms were sought to separate two histologic tissue categories: normal and abnormal. The abnormal class contained sites with dysplasia, carcinoma *in situ* and squamous cell carcinoma; the normal class contained sites that were clinically

and/or histologically normal as well as benign changes such as inflammation and MRA.

Fluorescence data from a single measurement site is represented as a matrix containing calibrated fluorescence intensity as a function of excitation and emission wavelength. Columns of this matrix correspond to emission spectra at a particular excitation wavelength; rows of this matrix correspond to excitation spectra at a particular emission wavelength. Each excitation spectrum contains 18 intensity measurements; each emission spectrum contains between 50 and 130 intensity measurements depending on excitation wavelength. Finally, emission spectra were truncated at 600 nm emission to eliminate the highly variable background due to room light present above 600 nm. Most multivariate data analysis techniques require vector input, so the column vectors containing the emission spectra at excitation wavelengths selected for evaluation were concatenated into a single vector.

Our previous work illustrates that spectra of oral cavity obtained *in vivo* show large patient to patient variations in intensity that can be greater than the intercategory differences (24). Therefore, we explored preprocessing methods to reduce the interpatient variations, while preserving intercategory differences. Two methods were selected for evaluation here: (1) normalization of all emission spectra in a concatenated vector by the largest emission intensity contained within that vector; and (2) normalization of each emission spectra to its maximum intensity. Because emission spectra were truncated to a maximum wavlength of 600 nm, the diagnostic capability of spectral information above 600 nm was evaluated with a different method described later.

In this study, fluorescence emission spectra were measured at 18 different excitation wavelengths. One goal of the data analysis was to determine which combination of excitation wavelengths contained the most diagnostic information. Combinations of emission spectra from up to four excitation wavelengths were considered. Limiting the device to four wavelengths allows for construction of a reasonably cost-effective clinical spectroscopy system (33). Two strategies were considered to identify the optimal excitation wavelength combination. The first was to identify the single wavelength that gave the best diagnostic performance. The wavelength that most improved diagnostic performance was then identified from the remaining wavelengths. This process was continued until the performance was no longer improved, or four wavelengths had been selected. The second strategy was to evaluate all possible combinations of up to four wavelengths chosen from the 18 possible excitation wavelengths. This equated to 18 combinations of one, 153 combinations of two, 816 combinations of three and 3060 combinations of four excitation wavelengths, for a total of 4047 combinations. While the first strategy required less computational time, it was only appropriate for normalization methods that removed relative intensity information. Otherwise, the best single wavelength may not be part of the best wavelength pair that exploits differences in relative intensity. The second strategy could be used with either normalization scheme. In addition, it provided a tool to rank the top wavelength combinations, rather than identifying the single best wavelength combination. Thus, the second strategy was pursued.

For each of the 4047 combinations of one to four excitation wavelengths, spectra from the training set were used to develop multivariate algorithms to separate normal and abnormal tissues based on their fluorescence emission spectra at all possible wavelength combinations. The algorithm development consisted of three steps: (1) preprocessing; (2) data reduction; and (3) development of a classification algorithm that maximized diagnostic performance. Data were preprocessed using the two normalization schemes described above. For each normalization, principal component analysis was performed using the entire dataset, and eigenvectors accounting for 65, 75, 85 and 95% of the total variance were retained. Principal component scores associated with these eigenvectors were calculated for each sample (32). Discriminant functions were then formed to classify each sample as normal or abnormal. The classification was based on the Mahalanobis distance, which is a multivariate measure of the separation of a point from the mean of a dataset in n-dimensional space (34). The sample was classified to the group from which it had the shorter Mahalonobis distance. The sensitivity and specificity of the algorithm were then evaluated relative to diagnoses based on histopathology (in patients suspected to have oral cavity

malignancy) or clinical impression (in normal volunteers). Overall diagnostic performance was evaluated as the sum of the sensitivity and the specificity, thus minimizing the number of misclassifications (when prevalence of disease and normal are approximately equal). The performance of the diagnostic algorithm depended on the principal component scores that were included. Four different diagnostic algorithms were developed using principal component scores derived from eigenvectors accounting for increasing amounts of total variance. From the available pool of principal component scores, the single principal component score yielding the best initial performance was identified, and then the principal component score that most improved this performance was selected. This process was repeated until performance is no longer improved by the addition of principal component scores, or all available scores were selected. The pool of available eigenvectors is specified by a variance criterion, eigenvector significance level (ESL), which represents the minimum variance fraction accounted for by the sum of the n largest eigenvalues. In this work we examined four ESL, corresponding to 65, 75, 85 and 95% of the total variance.

The presence of room light in the measurements at emission wavelengths greater than 600 nm precluded their incorporation into the multivariate statistical algorithm development. Yet, the literature indicates that fluorescence associated with endogenous porphyrins at 410 nm excitation, 635 nm emission may be diagnostically useful. To evaluate the diagnostic capability of the red fluorescence, we performed a least squares fit of our data to the sum of a straight line with variable slope and intercept, and a Gaussian (μ = 639 nm, σ = 7.5 nm) with variable intensity to the emission spectra obtained at 410 nm excitation from 600 to 650 nm. The line approximated the sloping background contributed by the room light while the Gaussian peak described the fluorescence contribution of endogenous porphyrin. We investigated whether combining the peak intensity with the optimal wavelength combination identified earlier could further improve algorithm performance. The peak intensity of the porphyrin peak was then incorporated into the algorithm as if it were an additional principal component score and the algorithm performance determined.

At each ESL, algorithm performance was noted for each wavelength combination, using the sum of sensitivity and specificity as a metric of performance. The 25 combinations of excitation wavelengths with the highest performance were then identified. However, as the ESL approaches 100%, overtraining becomes more likely, since the available pool of eigenvectors will account for nearly 100% of the variance, including variance due to noise. The magnitude of diagnostically important variances is unknown.

The risk of overtraining was assessed at the top-25 wavelength combinations of two, three and four excitation wavelengths, by comparing the training set performance to the performance of an algorithm developed from the same data after the diagnoses corresponding to each measurement site had been randomized. This provides a dataset with the same variance structure as the original dataset, but where the diagnostic performance is not expected to exceed that of chance. In order to make equivalent comparisons, the disease prevalence in the real sample was maintained in the randomly assigned diagnoses. Diagnostic algorithms were then developed again which minimized the number of misclassified samples at a specified ESL. Random diagnoses were assigned 50 times for each wavelength combination and the average and standard deviation of the sum of the sensitivity and specificity were calculated. Ideally, for completely normally distributed data, the sum of the sensitivity and specificity should be 1 for the randomized diagnosis at all levels of training significance. However, if overtraining occurs, this sum will be greater than one.

The top-25 wavelength combinations were then ranked based in order of the increase in performance between the training set performance with the correct histopathologic diagnoses and the training set performance with random assignment of diagnosis. This method allows the top wavelength combinations to be ranked in order of their robustness, or lack of propensity to overtrain. For a given number of wavelengths per combination, the differences were ranked across all four eigenvector significance levels. The largest difference, usually seen at ESL values of 65%, was selected as the optimal wavelength combination. This criterion selects the wavelength combination that is least prone to overtraining.
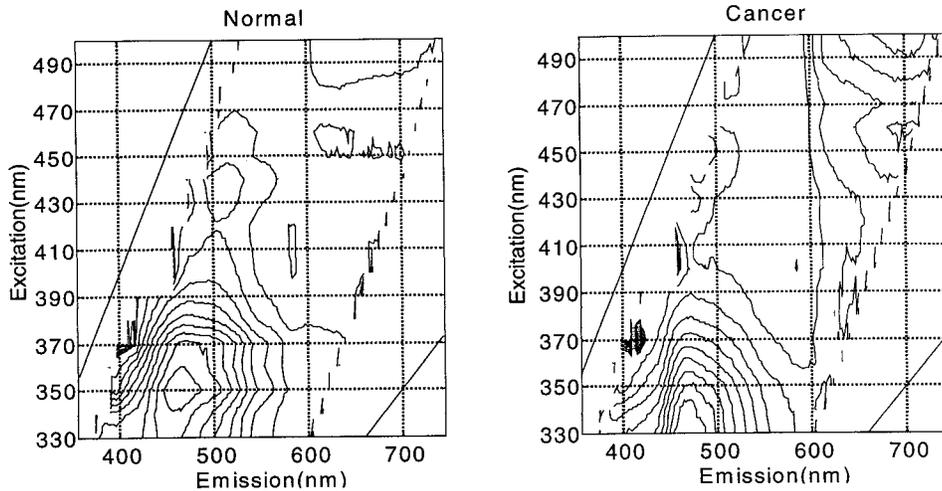
Normal

Cancer

Figure 1. Fluorescence EEM from a cancerous site on the tongue and a contralateral normal site. Contour lines connect points of equal fluorescence intensity.

Although the optimal wavelength combination has been identified based upon comparison of its performance to that which can be achieved when the tissue diagnoses have been randomized, our estimates of algorithm performance are still biased since they are based on the training set used to develop the algorithm. An unbiased performance estimate must be made to assess the true potential of this wavelength combination. The effects of overtraining in performance estimation can be minimized by using separate training and validations sets, or by using the method of cross-validation (35). Here we pursued both. Initially, we used the cross-validation method. In this method, all data from one patient are temporarily removed from the training data set, the algorithm is developed using the remaining data set, and then the new algorithm is applied to the left out sites. This is repeated until data from each patient has been left out once. Cross-validation was used to provide an estimate of the performance of the top-three combinations of excitation wavelengths for each normalization method.

Using the training set, we investigated whether effective diagnostic algorithms could be developed using reduced numbers of emission wavelengths at the top-performing excitation wavelength combinations. We calculated the component loadings associated with the eigenvectors corresponding to the principal component scores selected in these algorithms (15,33). A component loading represents the correlation between each principal component and the original preprocessed fluorescence emission spectra at each excitation wavelength. The component loadings at each excitation wavelength were evaluated to select fluorescence intensities at a minimum number of excitation–emission wavelength pairs required for the algorithms to perform with a minimal decrease in classification accuracy. Portions of the component loadings most highly correlated (correlation >0.5 or <−0.5) with corresponding emission spectra at each excitation wavelength were selected and the reduced data matrix was then used to regenerate and evaluate the algorithms. In addition, simple algorithms utilizing ratios of intensities at these reduced excitation–emission wavelength combinations were explored. The most successful of these simple algorithms were applied without change to data from the validation set to provide an unbiased estimate of algorithm performance.

RESULTS

In the training set, fluorescence EEM from 62 sites from 20 subjects was available for further analysis (Table 1). Of these 62 sites, 37 were measured from the tongue, eight from the floor of mouth (FOM), seven from the buccal mucosa, four from the gingiva, one from the palate and five from the lip. There were 52 normal, four dysplastic and six cancerous sites. The data set consisted of two types of normal sites: adjacent normals and normals from a population without suspected oral cancer. Adjacent normals are the visually nor-

mal sites taken from patients that have suspected lesions elsewhere in the oral cavity. In this data set there were 17 adjacent normal (histologically normal) sites from 11 patients, and 35 visually normal sites taken from nine normal volunteers. In the validation set, fluorescence EEM from 281 sites from 56 subjects was available for further analysis (Table 1). Of these 281 sites, 204 were measured from the tongue, 75 from the buccal mucosa and two from the gingiva. There were 274 normal, three adjacent normal, two dysplastic and two cancerous sites.

The visual screening accuracy of the head and neck specialists for the training data set was 100% sensitivity and 83% specificity and 100% sensitivity and 100% specificity for the validation set. This performance was determined by comparing the visual impressions of the clinicians to the histologic findings upon excision. Typical EEM for a cancerous lesion and its contralateral normal site from a single patient are depicted in Fig. 1. Results of the multivariate analysis of the spectroscopic data are presented according to the normalization method used.

Normalization by peak emission intensity of the concatenated vector

We ranked the 25 top-performing combinations of one to four excitation wavelengths in order of the largest difference in the sum of the sensitivity and the specificity in the training set with the actual histologic diagnoses and the average performance with randomly assigned diagnoses. The top-three combinations correspond to the following excitation wavelength combinations: (350, 380, 400, 480 nm), (350, 380, 400, 490 nm) and (350, 380, 400 nm). All of these combinations demonstrate approximately the same performance with the correct diagnosis, with 100% sensitivity and 90% specificity. These combinations have three wavelengths in common. Since no performance benefit was observed when a fourth wavelength was added for the top-performing combinations, combinations of four wavelengths were not pursued any further.

Table 2 shows the top-25 combinations of three excitation wavelengths. A bar graph depicting the number of times each wavelength appeared in the top-25 combinations from Table 2 is shown in Fig. 2 at various ESL. At low ESL

Table 2.   Combinations of three excitation wavelengths with the top performance increase in the sum of sensitivity and specificity using the correct diagnosis and a randomly assigned diagnosis (ESL = 65%)

| Rank | Wavelength combination | | | Sum of sensitivity and specificity | | Performance increase between correct and randomly assigned diagnosis |
				Correct diagnosis	Randomly assigned diagnosis	
1	350	380	400	1.90	1.13 ± 0.17	0.77
2	350	380	390	1.78	1.13 ± 0.14	0.66
3	350	380	460	1.87	1.25 ± 0.12	0.62
4	350	380	450	1.85	1.25 ± 0.12	0.59
5	360	370	400	1.77	1.20 ± 0.13	0.57
6	380	400	430	1.78	1.23 ± 0.12	0.56
7	380	400	410	1.78	1.23 ± 0.12	0.55
8	350	400	500	1.78	1.24 ± 0.12	0.54
9	330	410	430	1.77	1.23 ± 0.13	0.54
10	360	400	440	1.80	1.26 ± 0.14	0.54
11	330	410	440	1.77	1.23 ± 0.11	0.54
12	370	380	400	1.78	1.25 ± 0.14	0.53
13	330	400	410	1.77	1.23 ± 0.14	0.53
14	330	350	380	1.77	1.24 ± 0.13	0.53
15	350	380	430	1.78	1.26 ± 0.13	0.52
16	420	490	500	1.78	1.26 ± 0.12	0.52
17	330	350	400	1.77	1.25 ± 0.14	0.52
18	330	400	440	1.77	1.26 ± 0.14	0.50
19	330	400	430	1.77	1.27 ± 0.13	0.50
20	390	400	420	1.78	1.30 ± 0.13	0.49
21	350	370	430	1.79	1.31 ± 0.12	0.48
22	350	360	400	1.77	1.30 ± 0.11	0.47
23	350	360	370	1.78	1.34 ± 0.14	0.44
24	330	340	400	1.77	1.32 ± 0.14	0.44
25	330	360	400	1.77	1.33 ± 0.14	0.44
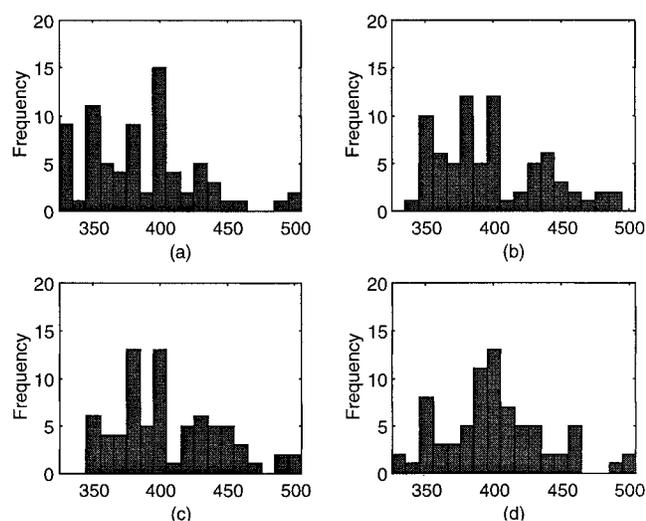
Figure 2. Frequency of occurrence for each emission wavelength in top-25 combinations of three wavelengths ranked according to greatest increase in the sum of sensitivity and specificity for the correct and randomly assigned diagnoses (a) ESL = 65%, (b) ESL = 75%, (c) ESL = 85% and (d) ESL = 95%.

values of 65, 75 and 85% the diagnostic importance of excitation at 350, 380 and 400 nm is evident. The same trends are seen for wavelength combinations of two and four excitation wavelengths, indicating the importance of these three excitation wavelengths.

To provide a less biased estimate of performance of these algorithms, the diagnostic performance of the top wavelength combinations was evaluated by using the method of cross-validation on the full data set. The wavelength combination (350, 380, 400 nm) demonstrated a cross-validation performance of 100% sensitivity and 88% specificity. The other two combinations (350, 380, 400, 480 nm) (350, 380, 400, 490 nm) demonstrated nearly identical performance upon cross-validation with a sensitivity of 100% and a specificity of 90%.

The emission spectra corresponding to all 62 sites in the training set at the three excitation wavelengths common to these combinations are shown in Fig. 3, where the concatenated emission vector has been normalized to a maximum of one. The peak intensity of this concatenated vector was nearly always observed at 350 nm excitation. Visual examination of Fig. 3 confirms the diagnostic potential of this wavelength combination. With this normalization, the normal sites demonstrate greater fluorescence intensity at 380 nm excitation, 450 nm emission than the abnormal sites.
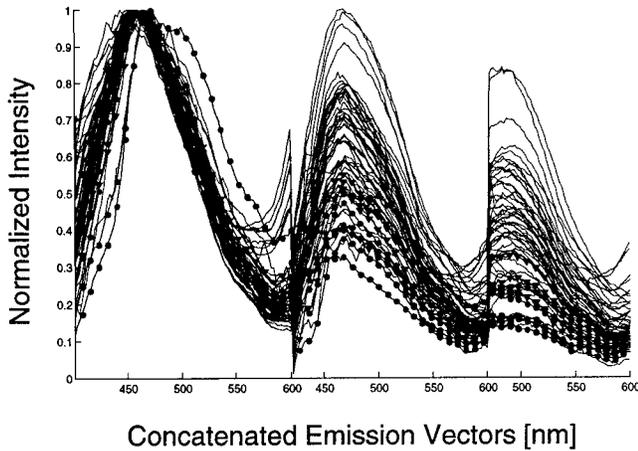
Figure 3. Fluorescence emission spectra normalized by the peak intensity of the concatenated vector for all 62 sites in the training set at 350, 380 and 400 nm excitation (-●- indicates histologically cancerous, -▼- indicates histologically dysplastic and lines indicate visually and/or histologically normal sites).

Figure 4. Plot of the only eigenvector of diagnostic importance at ESL = 65% for wavelength combination (350, 380, 400 nm) (solid line) and the corresponding component loading (dashed line with shaded circles). Shaded circles indicate wavelengths used in the reduced algorithm.

Additionally, the remaining emission peaks tend to be more intense in normal sites than for abnormal sites in most instances. Histologically, normal sites misclassified as abnormal demonstrated increased vascularity, suggesting that the increased hemoglobin absorption is one cause of the reduced relative fluorescence intensity from these sites.

The algorithm based on the combination of 350, 380 and 400 nm excitation wavelengths selected only a single principal component score associated with the eigenvector that accounted for most of the total variance. Figure 4 shows this eigenvector and the associated component loading as a function of emission wavelength for each of the three excitation wavelengths. The eigenvector depicts the general lineshape of the normalized spectra shown in Fig. 3. The component loading shows that the principal component score for this eigenvector is highly correlated to approximately four regions of the concatenated emission vector. Single emission intensities within these ranges were selected arbitrarily and are denoted as circles in Fig. 4. These points correspond to the emission intensities of 418 and 470 nm at 350 nm excitation, 448 nm emission at 380 nm excitation and 502 nm emission at 400 nm excitation. An algorithm was developed using the same data reduction and classification methods as above based upon this reduced data set. The training performance of the reduced algorithm is 100% sensitivity and 90% specificity, and the cross-validated performance is 90% sensitivity and 90% specificity compared to 100% sensitivity and 88% specificity for the algorithm based on the entire emission spectra. Motivated by the desire to construct a simple device that could interrogate or image large areas of tissue, a reduced algorithm based upon a single emission wavelength was evaluated. The emission wavelength chosen was common to all three emission spectra, 472 nm. The training performance of this reduced algorithm was 100% sensitivity, 88% specificity, and upon cross-validation it was 90% sensitivity and 88% specificity.

Normalization of each emission spectra by its peak emission intensity prior to concatenation

The analysis was repeated using concatenated vectors in which each emission spectrum was normalized to its peak
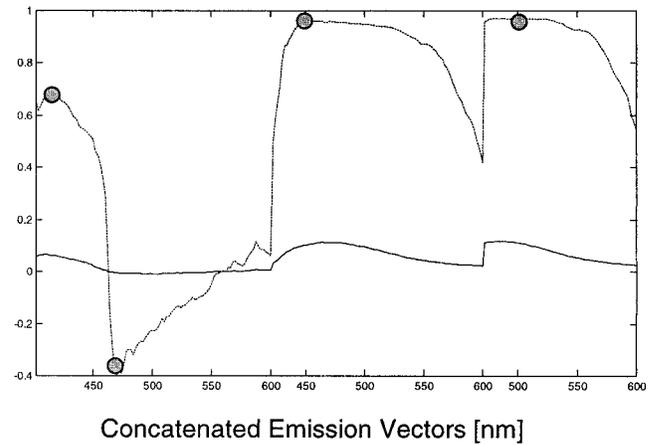
intensity. Figure 5 shows the emission spectra corresponding to all 62 sites in the training set at 350, 380 and 400 nm excitation, where each emission spectrum has been normalized to a maximum of 1. This method removes relative intensity information and relies on differences in fluorescence lineshape. The maximum difference between training performance and the performance after random diagnosis assignment was 0.58 compared to 0.82 using the previously described normalization method. Consequently, the top wavelength combination identified (350, 380, 400, 430 nm) showed poor performance upon cross-validation with a sensitivity of 50% and a specificity of 88%. It is interesting to note that the previously identified wavelengths (350, 380, 400 nm) are also a part of this combination, indicating that the lineshape at these wavelengths contains some diagnostic information.
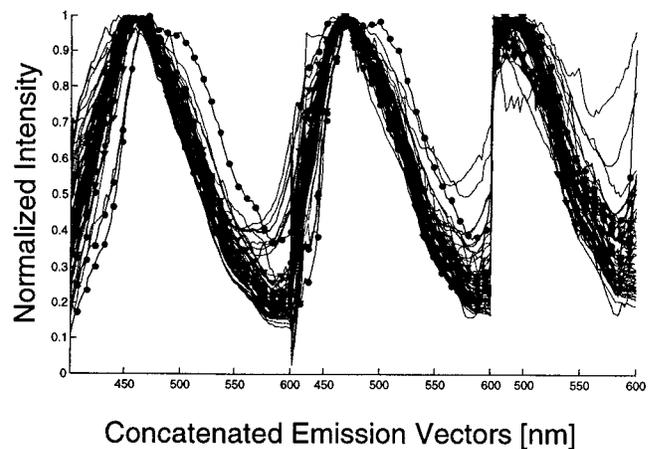

Figure 5. Plot of emission vector for a wavelength combination of three excitation wavelengths (350, 380, 400 nm) normalized by the peak intensity of each emission spectra (-●- indicates histologically cancerous, -▼- indicates histologically dysplastic and lines indicate visually and/or histologically normal sites).
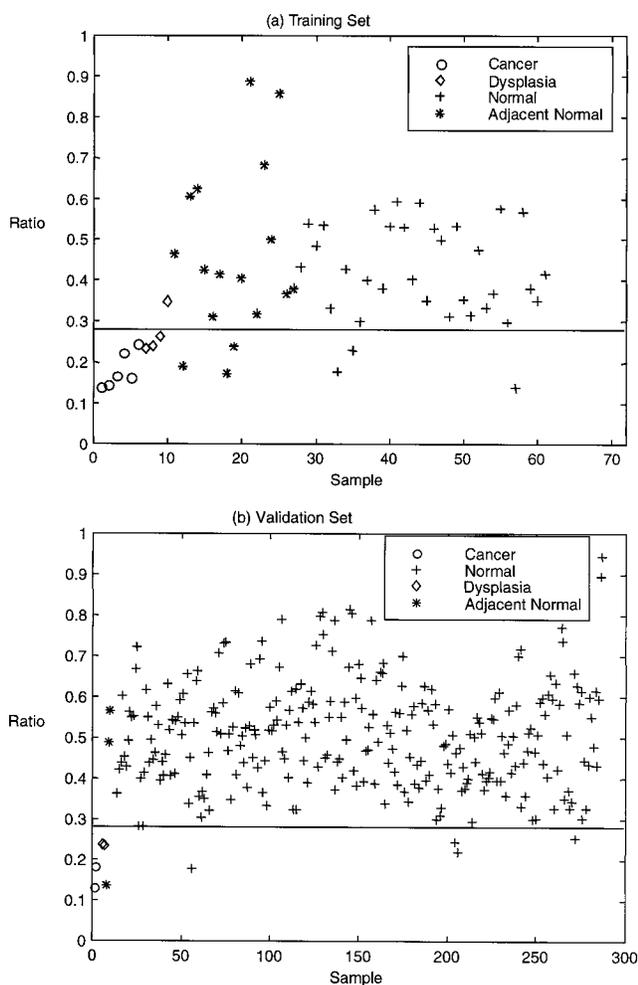
Figure 6. Ratio of fluorescence intensities at 400 nm excitation, 472 nm emission to that at 350 nm excitation, 472 nm emission for all samples in (a) the training set and (b) the validation set. The solid line at a ratio of 0.28 separates normal samples from dysplasias and cancers with a sensitivity of 90% and a specificity of 88% in the training set, and a sensitivity of 100% and a specificity of 98% in the validation set.

Inclusion of red fluorescence with optimal wavelength combination

Peaks at 639 nm emission at 410 nm excitation were present in 13 of 38 normal sites, six of 16 adjacent normals, three of four dysplastic sites and three of six cancerous sites in the training set. The inclusion of the red fluorescence intensity normalized by the peak of the concatenated vector did not offer any improvement in performance. This is mostly attributable to the inconsistent incidence of this peak in abnormal and normal sites.

Comparison of algorithm performance in training and validation datasets

The performance of a further simplified algorithm was compared in the training and validation datasets. Figure 6 shows a simple algorithm, based on the ratio of the fluorescence intensity at 400 nm excitation, 472 nm emission to that at 350 nm excitation, 472 nm emission. Data from the training

set are shown in Fig. 6a and the horizontal line at an intensity ratio of 0.28 separates normal oral cavity from dysplasia and cancer with a sensitivity of 90% and a specificity of 88%. The same algorithm was then applied to data from the validation set; results are shown in Fig. 6b with a sensitivity of 100% and a specificity of 98%.

DISCUSSION AND CONCLUSIONS

This study identified the optimal excitation wavelengths for *in vivo* detection of oral cancers with fluorescence spectroscopy. The optimal excitation wavelengths were found to be 350, 380 and 400 nm. Using data from the training set with cross-validation methods yielded an estimate of algorithm performance based on the entire emission spectra at these excitation wavelengths with a sensitivity of 100% and specificity of 88%. Increasing the number of excitation wavelengths did not improve algorithm performance. Better algorithm performance was obtained when data were normalized to the peak emission intensity of the concatenated vector than when each emission spectrum was normalized to its own peak emission wavelength. The discriminating ability of this wavelength combination is due to differences in both relative intensity and spectral line shape. The number of emission wavelengths could be significantly reduced as well without compromising algorithm performance. An algorithm based on four emission intensities: 418 and 470 nm at 350 nm excitation, 448 nm emission at 380 nm excitation and 502 nm emission at 400 nm excitation yielded 90% sensitivity and 90% specificity upon cross-validation. When only a single emission wavelength of 472 nm, common to all three excitation wavelengths, was used algorithm performance on cross-validation was 90% sensitivity and 88% specificity.

Further simplification is possible. An algorithm based on the ratio of fluorescence intensities at two excitation wavelengths (400 and 350 nm) and a single emission wavelength (472 nm) showed similar performance in both the training and validation sets, with an average sensitivity of 95 ± 7% and specificity of 93 ± 6% for the separate training and validation sets. Performance in the validation set slightly exceeded that in the training set, indicating that results of this analysis are not subject to overtraining or instrument drift, since collection of data in the training and validation sets were separated in time and minor instrument modifications were made just before this interval. While similar methodology can easily be applied to fluorescence EEM from other organ sites, our recent work indicates that different optimal excitation wavelength combinations exist even for epithelial tissues with similar histologic appearance such as the cervix (36).

Changes in fluorescence at these excitation–emission wavelength combinations likely result from a combination of differences in both the autofluorescence and absorption properties of normal and neoplastic tissue. Sacks and colleagues (37) reported differences in the autofluorescence of normal oral epithelial cells, which correlated with cell differentiation. At 480 nm emission, the ratio of autofluorescence at 390 nm excitation to that at 350 nm excitation was 0.28 for the least differentiated cells and was 0.87 for the most differentiated cells; these ratios are quite similar to

Table 3. Comparison of diagnostic performance based on clinical impression and fluorescence spectroscopy using the ratio of fluorescence intensities at 400 nm excitation, 472 nm emission and 350 nm excitation, 472 nm emission in the training and validation sets

	Training set		Validation set		Average ± standard deviation	
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
Clinical impression	100	83	100	100	100	92 ± 9
Fluorescence spectroscopy	90	88	100	98	95 ± 5	93 ± 5

those reported in Fig. 6 for neoplastic and normal oral cavity, respectively. Hemoglobin absorption, which is strongest at 420 nm, also contributes to fluorescence spectra. It is interesting to note that emission spectra obtained at 400 nm excitation are included in a majority of the top combinations, suggesting that differences in absorption due to perfusion may offer diagnostic information. This suggests that the combinations of reflectance and fluorescence spectroscopy may offer improved diagnostic performance and will be the subject of future work (38).

The inclusion of red fluorescence intensity information with the optimal wavelength combination did not improve algorithm performance. Ghadially *et al.* (39) identified the red fluorescence at approximately 640 nm emission to be from protoporphyrin. Their studies of necrotic tumors indicated this red fluorescence to be bacterial in origin rather than a product of the tumor. Although protoporphyrin was associated with a large proportion of our abnormal measurements (60%), it was not found to be diagnostically useful.

The unbiased performance estimate for the diagnostic algorithms based on fluorescence spectroscopy has a higher sensitivity than current visual screening by experts. Visual screening has been reported to have a sensitivity of 74% and specificity of 99% (40). The performance of visual screening by experts in this study was 100% sensitivity and 92 ± 9% specificity averaged for the separate training and validation sets. Table 3 shows that this compares favorably to that of fluorescence with an average performance of 95 ± 5% sensitivity and 93 ± 5% specificity for an algorithm based on the ratio of two fluorescence intensities. Thus, simple, but quantitative optical measurements can provide accurate tools to differentiate normal oral cavity from premalignant and malignant regions. This study shows that these tools have the potential to provide inexperienced practitioners with a screening tool with the sensitivity and specificity of experienced practitioners. Furthermore, because optical methods do not require tissue removal, they may provide better tools for experts to direct diagnostic biopsies and determine tumor margins.

REFERENCES

1. American Cancer Society (1993) *Cancer Facts and Figures,* Publication 93-400M, No. 5008-03. American Cancer Society, Washington, DC.
2. Boring, C. C., T. S Squires, T. Tong and S. Montgomery (1994) Cancer statistics, 1994. *CA: Cancer J. Clin.* **44**, 7–26.
3. Blair, E. A. and D. L. Callendar (1994) Head and neck cancer—the problem. *Clin. Plast. Surg.* **21**, 1–7.
4. Strong M. S., J. Incze and C. N. Vaughan (1984) Field cancerization in the aerodigestive tract—its etiology, manifestation, and significance. *J. Otolaryngol.* **13**, 1–6.
5. WHO Collaborating Centre for Oral Precancerous Lesions (1978) Definition of leukoplakia and related lesions: an aid to studies on oral precancer. *Oral Surg. Oral Med. Oral Pathol.* **46**, 518–539.
6. Shafer, W. G., M. K. Hine, B. M. Levy and C. E. Tomich (1983) *A Textbook of Oral Pathology.* Saunders, Philadelphia.
7. Roed-Peterson, B. (1971) Cancer development in oral leukoplakia: follow up of 331 patients. *J. Dent. Res.* **50**, 711.
8. Silverman, S., M. Gorsky and F. Lozada (1984) Oral leukoplakia and malignant transformation. A follow up study of 257 patients. *Cancer* **53**, 563–568.
9. Silverman, S., C. Migliorati and J. Barbarosa (1984) Toluidine blue staining in the detection of oral precancerous and malignant lesions. *Oral Surg. Oral Med. Oral Pathol.* **57**, 379–382.
10. Reddy, C. R. M., C. Ramilu, B. Sundareshwar, M. V. S. Raju, R. Gopal and R. Sarma (1973) Toluidine blue staining of oral cancer and precancerous lesions. *Indian J. Med. Res.* **61**, 1161–1164.
11. Rosenberg, D. and S. Cretin (1987) Use of meta-analysis to evaluate tolniun chloride in oral cancer screening. *Oral Surg. Oral Med. Oral Pathol.* **67**, 621–627.
12. Epstein, J. B. and C. Scully (1997) Assessing the patient at risk for oral squamous cell carcinoma. *Spec. Care Dent.* **17**, 120–128.
13. Sevick-Muraca, E. and R. Richards-Kortum (1996) Quantitative optical spectroscopy for tissue diagnosis. *Annu. Rev. Phys. Chem.* **47**, 555–606.
14. Wagnieres, G. A., W. M. Star and B. C. Wilson (1998) In vivo fluorescence spectroscopy and imaging for oncological applications. *Photochem. Photobiol.* **68**, 603–632.
15. Ramanujam, N., M. Follen-Mitchell, A. Mahadevan-Jansen, S. Thomsen, G. Staerkel, A. Malpica, T. Wright, N. Atkinson and R. Richards-Kortum (1996) Cervical precancer detection using multivariate statistical algorithm based on laser-induced fluorescence spectra at multiple excitation wavelengths. *Photochem. Photobiol.* **64**, 720–735.
16. Cothren, R., R. Richards-Kortum, M. Sivak, M. Fitzmaurice, R. Rava, G. Boyce, G. Hayes, M. Doxtader, R. Blackman, T. Ivanc, M. Feld and R. Petras (1990) Gastrointestinal tissue diagnosis by laser induced fluorescence spectroscopy at endoscopy. *Gastrointest. Endosc.* **36**, 105–111.
17. Schantz, S. P., V. Kolli, H. E. Savage, G. Yu, J. P. Shah, D. E. Harris, A. Katz, R. R. Alfan and A. G. Huvos (1998) *In vivo* native cellular fluorescence and histological characteristics of head and neck cancer. *Clin. Cancer Res.* **4**, 1177–1182.
18. Kolli, V., H. E. Savage, T. J. Yao and S. P. Schantz (1995) Native cellular fluorescence of neoplastic upper aerodigestive mucosa. *Arch. Otolaryngol. Head Neck Surg.* **121**, 1287–1292.
19. Chen, C. T., C. Y. Wang, Y. S. Kuo, H. H. Chiang, S. N. Chow, I. Y. Hsiao and C. P. Chiang (1996) Light-induced fluorescence

spectroscopy: a potential diagnostic tool for oral neoplasia. *Proc. Natl. Sci. Counc. Repub. China, Part B, Life Sci.* **20**, 123–130.

20. Roy, K., I. D. Bottrill, D. R. Ingrams, M. M. Pankratov, E. E. Rebeiz, P. Woo, S. Kabani, S. M. Shapshey, R. Manoharan, I. Itzkan and M. S. Feld (1995) Diagnostic fluorescence spectroscopy of oral mucosa. *SPIE* **2395**, 135–142.

21. Ingrams, D. R., J. K. Dhingra, K. Roy, D. F. Perrault Jr., I. D. Bottrill, S. Kabani, E. E. Rebeiz, M. M. Pankratov, S. M. Shapshay, R. Manoharan, I. Itzkan and M. S. Feld (1997) Autofluorescence characteristics of oral mucosa. *Head Neck* **19**, 27–32.

22. Dhingra, J. K., X. Zahng, K. McMillan, S. Kabani, R. Manoharan, I. Itzkan, M. S. Feld and S. M. Shapshay (1998) Diagnosis of head and neck precancerous lesions in an animal model using fluorescence spectroscopy. *Laryngoscope* **108**, 471–475.

23. Dhingra, J. K., D. F. Perrault Jr., K. McMillan, E. E. Rebeiz, S. Kabani, R. Manoharan, I. Itzkan, M. S. Feld and S. M. Shapshay (1996) Early diagnosis of upper aerodigestive tract cancer by autofluorescence. *Arch. Otolaryng. Head Neck Surg.* **122**, 1181–1186.

24. Gillenwater, A., R. Jacob, R. Ganeshappa, B. Kemp, A. K. El-Naggar, J. L. Palmer, G. Clayman, M. F. Mitchell and R. Richards-Kortum (1998) Noninvasive diagnosis of oral neoplasia based on fluorescence spectroscopy and native tissue autofluorescence. *Arch. Otolaryngol. Head Neck Surg.* **124**, 1251–1258.

25. Kulapaditharom, B. and V. Boonkitticharoen (1998) Laser-induced fluorescence imaging in localization of head and neck cancers. *Ann. Otol. Rhinol. Laryngol.* **107**, 241–246.

26. Onizawa, K., H. Saginoya, Y. Furuya and H. Yoshida (1996) Fluorescence photography as a diagnostic method for oral cancer. *Cancer Lett.* **108**, 61–66.

27. Lakowicz, J. (1983) *Principles of Fluorescence Spectroscopy.* Plenum Press, New York.

28. Schomacker, K., J. Frisoli, C. Compton, T. Flotte, J. Richter, N. Nishioka and T. Deutsch (1992) Ultraviolet laser-induced fluorescence of colonic tissue: basic biology and diagnostic potential. *Lasers Surg. Med.* **12**(1), 63–78.

29. Welch, A. J., C. Gardner, R. Richards-Kortum, E. Chan, G. Criswell, J. Pfefer and S. Warren (1997) Propagation of fluorescent light. *Lasers Surg. Med.* **21**, 166–178.

30. Zuluaga, A. F., U. Utzinger, A. Durkin, H. Fuchs, A. Gillenwater, R. Jacob, B. Kemp, J. Fan and R. Richards-Kortum (1999) Fluorescence excitation emission matrices of human tissue: a system for in vivo measurement and method of data analysis. *Appl. Spectrosc.* **53**, 302–311.

31. Utzinger, U., M. Brewer, E. Silva, D. Gershensen, R. Bast, M. Follen Mitchell, R. Richards-Kortum (2000) Reflectance spectroscopy for in vivo characterization of ovarian tissue. *Lasers Surg. Med.* (In press)

32. Utzinger, U., V. Trujillo, E. N. Atkinson, M. F. Mitchell, S. B. Cantor and R. Richards-Kortum (1999) Performance estimation of diagnostic tests for cervical pre-cancer based on fluorescence spectroscopy: effects of tissue type, sample size, population and signal-to-noise ratio. *IEEE Trans. BME* **46**, 1293–1303.

33. Cliff, N. (1987) *Analyzing Multivariate Data.* Harcourt Brace Jovanovich, Orlando.

34. Dillon, W. R. and M. Goldstein (1984) *Multivariate Analysis: Methods and Applications.* Wiley, New York.

35. Lachenbruch, P. A. (1975) *Discriminant Analysis.* Hafner Press, New York.

36. Utzinger, U., M. Follen and R. Richards-Kortum (2000) Combined fluorescence and reflectance spectroscopy for precancer detection: how many measurements do we need? In *Biomedical Topical Meetings OSA Technical Digest*, pp. 206–208. Optical Society of America, Washington, DC.

37. Sacks, P. G., H. E. Savage, J. Levine, V. R. Kolli, R. R. Alfano and S. P. Schantz (1996) Native cellular fluorescence identifies terminal squamous differentiation of normal oral epithelial cells in culture: a potential chemoprevention biomarker. *Cancer Lett.* **104**, 171–181.

38. Fuchs, H., U. Utzinger, A. F. Zuluaga, A. Gillenwater, R. Jacob, B. Kemp and R. Richards-Kortum (1998) Combined fluorescence and reflectance spectroscopy: in vivo assessment of oral cavity epithelial neoplasia. In *Technical Digest Summaries of Papers Presented at the Conference on Lasers and Electro-Optics*, Vol. 6, pp. 306–307. Optical Society of America, Washington, DC.

39. Ghadially, F., W. Neish and H. Dawkins (1963) Mechanisms involved in the production of red fluorescence of human and experimental tumours. *J. Pathol. Bact.* **85**, 77–92.

40. Jullien, J. A., M. C. Downer, J. M. Zakrzewska and P. M. Speight (1995) Evaluation of a screening test for the early detection of oral cancer and precancer. *Community Dent. Health* **12**, 3–7.