Anatomy-Based Algorithms for Detecting Oral Cancer Using Reflectance and Fluorescence Spectroscopy

Sasha McGee, PhD, **Vartan Mardirossian, MD**, **Alphi Elackattu, MD**, **Jelena Mirkovic, PhD**, **Robert Pistey, MD, MA**, **George Gallagher, DMD, DMSc**, **Sadru Kabani, DMD, MA**, **Chung-Chieh Yu, PhD**, **Zimmern Wang, MD**, **Kamran Badizadegan, MD**, **Gregory Grillone, MD**, and **Michael S. Feld, PhD**

G. R. Harrison Spectroscopy Laboratory, Massachusetts Institute of Technology (McGee, Mirkovic, Yu, Badizadegan, Feld), the Departments of Otolaryngology-Head and Neck Surgery (Mardirossian, Elackattu, Wang, Grillone) and Anatomic Pathology (Pistey). Boston Medical Center, the Department of Oral and Maxillofacial Pathology, Boston University Goldman School of Dental Medicine (Gallagher, Kabani), and the Department of Pathology, Harvard Medical School and Massachusetts General Hospital (Badizadegan), Boston, Massachusetts

Abstract

Objectives—We used reflectance and fluorescence spectroscopy to noninvasively and quantitatively distinguish benign from dysplastic/malignant oral lesions. We designed diagnostic algorithms to account for differences in the spectral properties among anatomic sites (gingiva, buccal mucosa, etc).

Methods—In vivo reflectance and fluorescence spectra were collected from 71 patients with oral lesions. The tissue was then biopsied and the specimen evaluated by histopathology. Quantitative parameters related to tissue morphology and biochemistry were extracted from the spectra. Diagnostic algorithms specific for combinations of sites with similar spectral properties were developed.

Results—Discrimination of benign from dysplastic/malignant lesions was most successful when algorithms were designed for individual sites (area under the receiver operator characteristic curve [ROC-AUC], 0.75 for the lateral surface of the tongue) and was least accurate when all sites were combined (ROC-AUC, 0.60). The combination of sites with similar spectral properties (floor of mouth and lateral surface of the tongue) yielded an ROC-AUC of 0.71.

Conclusions—Accurate spectroscopic detection of oral disease must account for spectral variations among anatomic sites. Anatomy-based algorithms for single sites or combinations of sites demonstrated good diagnostic performance in distinguishing benign lesions from dysplastic/ malignant lesions and consistently performed better than algorithms developed for all sites combined.

Keywords

diagnosis; fluorescence; oral cancer; reflectance; spectroscopy

---

Correspondence: Sasha McGee, PhD, Bldg 6-205, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139.

INTRODUCTION

Currently, definitive detection and diagnosis of oral cancer requires biopsy followed by histopathologic assessment of the excised tissue.[1] However, there are several shortcomings to this scheme. First, only a limited number of biopsy specimens can be taken because of the invasiveness of the procedure. On the basis of his or her experience, the physician selects the area of the lesion most likely to show significant disease as the biopsy site. The absence or presence of disease in this specimen is assumed to be representative of the extent of disease in the suspicious lesion as a whole, and this finding often determines whether treatment is indicated.[2] Given the subjective nature of this process, regions of disease can be missed. Significant underdiagnosis has been noted with biopsy, particularly when the lesion is nonhomogeneous or only a single biopsy specimen is obtained.[2,3] Second, the accuracy of pathological classification is limited by significant interobserver and intraobserver variability, largely due to the qualitative nature of the markers used for assessment.[4–6]

Spectroscopy may provide an objective and non-invasive tool for disease diagnosis. Promising findings have been reported for identifying oral lesions by using reflectance and fluorescence spectroscopy [7–13] The majority of these studies have focused on distinguishing healthy mucosa from visible lesions (either a group consisting of dysplastic and malignant lesions ["dysplastic/malignant"] or malignant lesions).[7,8,12–19] However, this separation has no clinical relevance, because these categories are easily distinguished by visual inspection. In some cases, benign lesions are grouped with healthy mucosa samples.[9,20] Most studies combine data from several anatomic sites (eg, gingiva, buccal mucosa) in each category in generating diagnostic algorithms based on differences in the spectral properties (spectral contrast).[7,10,11,13,14,18,21] This approach may not be ideal, because numerous studies have shown significant differences in the spectral properties between various anatomic sites, even for healthy oral mucosa.[7,15,22–24] We refer to the spectral contrast produced by variations in histologic characteristics from site to site as anatomic spectral contrast. The presence of keratin on some sites (particularly the gingiva and hard palate) produces marked anatomic spectral contrast between these sites and nonkeratinized sites.[9,24] A recent study of clinically normal mucosa in healthy volunteers (HVs) by our group demonstrated considerable anatomic spectral contrast even among nonkeratinized sites.[25] The results of our investigation suggested that spectral diagnostic algorithms must be site-specific to ensure accurate disease diagnosis. We use the term anatomy-based algorithms to refer to algorithms that meet this condition, in that they are developed from and applied to a specific site or group of sites. Because most studies combine several sites, the reported diagnostic accuracies may be confounded by anatomic spectral contrast and therefore unreliable.

The distinction of benign lesions from dysplastic/malignant lesions is not possible by visual inspection alone and has considerable clinical significance, yet few spectroscopy studies focus on this separation.[21,26–30] Anatomic spectral contrast can also influence the diagnostic results for this separation if it is not taken into account. In a study by de Veld et al,[28] for most analysis methods, values of less than 0.70 were obtained for the area under the receiver operator characteristic (ROC) curve (ROC-AUC) when data from 11 anatomic sites were combined. Using fluorescence spectroscopy, van Staveren et al[11] combined all sites and achieved a sensitivity of 31 % and a specificity of 67%. Greater success has been achieved in distinguishing lesions when the analysis was limited to a single site or a specific group of sites. Wang et al[30] performed a fluorescence study of patients with lesions located only on the buccal mucosa. They obtained a sensitivity of 81% and a specificity of 96% for distinguishing benign lesions from dysplastic/malignant lesions. Mallia et al[21] successfully developed diagnostic algorithms to discriminate various oral lesions by using fluorescence spectroscopy, but only after excluding lesions of the vermilion border of the lip and of the dorsal and lateral surfaces of the tongue. They suggested that a separate spectral database is

needed for these 3 sites. In another study by Mallia et al,[26] reflectance spectroscopy was used to develop algorithms for all sites combined and for the buccal mucosa (ie, an anatomy-based algorithm). The sensitivities and specificities obtained for algorithms specific to the buccal mucosa were consistently higher for the discrimination of benign lesions from dysplastic lesions and dysplastic lesions from malignant lesions. Müller et al[9] developed a diagnostic algorithm to distinguish benign lesions from dysplastic/malignant lesions specifically for keratinized sites, and they obtained a higher sensitivity and specificity than when keratinized and nonkeratinized sites were combined. Although diagnostic algorithms for single sites or limited groups of sites have been investigated, a systematic approach for developing spectral algorithms for multiple sites has not been developed.

The goal of the present study was to develop a strategy for designing anatomy-based algorithms for multiple anatomic sites. Although the spectral properties of different anatomic sites may vary considerably, certain subsets may be similar enough in their spectral properties such that they can be combined. Furthermore, in many settings, developing a single diagnostic algorithm for spectrally similar sites may be more practical and less time-consuming than developing a diagnostic algorithm for each individual site. Anatomy-based spectral diagnostic algorithms are developed for 2 applications: 1) distinguishing visibly healthy mucosa from visible lesions, and 2) distinguishing benign lesions from dysplastic/malignant lesions. The distinction of clinically normal (healthy) mucosa from clinically abnormal mucosa (lesions) has been the focus of most of the previous studies in the oral cavity. We address the question of whether anatomic spectral contrast affects the diagnostic accuracy of this separation. The second application is the primary objective of this work, since lesions demonstrating disease (dysplasia or malignancy) require intervention.

From the tissue spectra, we extract parameters related to the morphological and biochemical properties. To account for anatomic spectral contrast, we first study healthy mucosa from several anatomic sites to characterize the similarities and differences in their spectroscopy parameters. We then apply these findings by developing common diagnostic algorithms for combinations of sites that share a high degree of spectral similarity. To evaluate the benefit of anatomy-based algorithms, we compared the performance of algorithms developed for these specific combinations of sites to those developed for all anatomic sites combined and to those for a single anatomic site. We also examine the trends in the spectroscopy parameters on the basis of disease category.

PATIENTS AND METHODS

Patients

Patients were recruited from the Department of Otolaryngology–Head and Neck Surgery at Boston Medical Center. The experimental protocol for in vivo data collection was approved by the Institutional Review Board of Boston Medical Center and the Committee on the Use of Humans as Experimental Suhjects at the Massachusetts Institute of Technology. Subjects enrolled in the study included patients undergoing biopsy for clinically suspicious lesions, as well as patients undergoing surgical resection of known dysplastic or malignant lesions. Pregnant women and individuals under the age of 18 years were excluded from the study. Written informed consent was obtained from all subjects to indicate their willingness to participate. Spectral data were collected from tissues with visible abnormalities identified by the clinician. The clinical appearance of the lesions (eg, leukoplakia, erythroplakia) was recorded. After collecting the spectral data, the physician scored the exact area from which the data were acquired using the blade of a small punch biopsy instrument (1.5 or 2 mm in diameter). The tissue sample was subsequently removed using a larger punch biopsy

instrument (3.5 mm diameter) and sent for histopathologic evaluation. In some cases, a scalpel biopsy was performed because of the lesion architecture or location. Spectral data were collected from multiple tissue samples in a patient when the physician selected multiple areas of the lesion for biopsy, or when several biopsies were performed of a lesion to be resected. The histopathologic assessment was performed by 3 pathologists. Specimens were classified as benign, dysplastic, malignant, or indefinite for dysplasia. The consensus diagnosis (agreement between at least 2 of the 3 pathologists) was considered the final diagnosis. An additional set of 710 spectra previously collected from 79 HVs at 9 anatomic sites was also analyzed, as described below.[25] Tissue samples from HVs were assumed to represent healthy mucosa and were not biopsied.

Data Collection

Reflectance and fluorescence spectra were collected from patients using the Fast Excitation Emission Matrix (FastEEM), an investigational instrument developed by our laboratory that has been previously described.[31] A 1.3-mm-diameter optical fiber probe was used to deliver the excitation light and collect the light emitted from the tissue. The optical fiber probe was disinfected with Cidex OPA (Advanced Sterilization Products, Irvine, California) according to the manufacturer's specifications, and placed in contact with the tissue during data collection. No exogenous contrast agents were applied to the tissue. The data collection procedures have been further described elsewhere.[25] Data were collected and analyzed from the following anatomic sites: buccal mucosa (BM), dorsal surface of the tongue (DT), floor of the mouth (FM), gingiva (GI), hard palate (HP), lateral surface of the tongue (LT), retromolar trigone (RT), soft palate (SP), and ventral surface of the tongue (VT).

Model-Based Analysis

We employed physical models to extract morphological and biochemical parameters from the tissue reflectance and fluorescence spectra. Details of the models are presented in other reports.[25,32,33] The inputs to the reflectance model are the reduced scattering coefficient, $\mu s'(\lambda)$, and the absorption coefficient, $\mu a(\lambda)$. By modeling the wavelength dependence of $\mu s'$, we extract 3 parameters: A, B, and C. The A parameter is a scaling parameter proportional to the density of scatterers. The B parameter reflects the size of the scattering particles.[34] The C parameter represents the magnitude of scattering by small scatterers.

The absorption coefficient was modeled as the sum of the contributions from 2 absorbers, hemoglobin (Hb) and β-carotene (βC). Hemoglobin absorption, $\mu a^{Hb}(\lambda)$, was modeled with use of a correction for the effect of vessel packaging, which provides the effective blood vessel radius as an additional fitting parameter.[25] By modeling the absorption, we extracted the following 4 parameters: the concentration of Hb (cHb), the oxygen saturation (α), the effective vessel radius, and the concentration of βC (cβC). Hence, from the reflectance spectra we extracted 7 parameters in all: 3 scattering parameters (A, B, C) and 4 absorption parameters.

We applied a physical model to the measured fluorescence spectra in order to remove distortions introduced by scattering and absorption and obtain what we call the intrinsic fluorescence.[33,35] In order to apply the findings of a study of the spectral properties of healthy mucosa by McGee et al[25] to the present study, data were analyzed at the same 2 excitation wavelengths (308 nm and 340 nm). At 308-nm excitation, the spectra were fit with a linear combination of tryptophan (308 nm), Coll (308 nm), and NADH (308 nm). The number in parentheses represents the excitation wavelength, Coll refers to collagen, and NADH refers to the reduced form of nicotinamide adenine dinucleotide. At 340 nm excitation, the spectra were fit with a linear combination of NADH (340 nm), Coll401 (340 nm), and Coll427 (340 nm), Coll401 (340 nm), and Coll427 (340 nm) are attributed to 2

distinct collagen components.[25] We obtained a total of 6 fluorescence parameters. Overall, a total of 13 spectroscopy parameters were used to characterize the degree of spectral similarity between various sites and to construct diagnostic algorithms.

Identifying Spectrally Similar Anatomic Sites

We compared the spectroscopy parameters extracted from each of the 9 sites in the HV data to identify anatomic sites that shared similar spectral properties and could be combined. The interquartile range exclusion criteria were applied to the data to remove outliers (eg, from probe slippage, tissue movement).[36] For each spectroscopy parameter, a multiple comparison test was performed to compare each pair of sites and identify those that had statistically different distributions. The Tukey-Kramer correction was applied to compensate for multiple comparisons. For each pair of anatomic sites, we calculated a "similarity score," the total number of spectroscopy parameters for which the distributions for the 2 sites were not statistically different. Perfect correspondence between the spectroscopy parameter distributions for a pair of sites would be indicated by a similarity score of 13 (total number of spectroscopy parameters). We considered pairs of sites with a similarity score of at least 10 to be spectrally similar sites.

Spectral Diagnostic Algorithm Development

The diagnostic performance of anatomy-based algorithms developed for spectrally similar sites or individual sites was compared to that of algorithms designed for all sites combined. Logistic regression was used to develop the diagnostic algorithms. The log-likelihood ratio test was used to identify the spectroscopy parameters with the greatest diagnostic potential for the distinction of healthy mucosa from lesions and of benign lesions from dysplastic/malignant lesions.[37] The discriminatory power of each algorithm was evaluated by leave-one-out cross-validation and by calculating the ROC-AUC, sensitivity, and specificity. The sensitivities and specificities quoted in the subsequent analysis were selected based on the Youden index, the point on the ROC curve with the maximum vertical distance from the 45° line.[38]

The fewest parameters necessary to achieve the maximum ROC-AUC for a comparison were retained. To prevent overtraining, we limited the maximum number of spectroscopy parameters used in the model to $n/5$, where n is the number of training samples in the category with the fewest samples. To further ensure the reliable assessment of the diagnostic algorithms, we only analyzed categories in which there were at least 10 samples.

RESULTS

General Description of Data Set

A total of 71 patients were recruited for this study. The average age (±SD) of the patients was 56±14 years. Of the 101 spectra collected from patients, 12 spectra from 7 patients were excluded because of an error noted during data collection, a broken probe, or the poor quality of the biopsy specimen, or because the pathology results were unavailable. Two samples, one of which was a lymphoma and another of which was classified as indefinite for dysplasia, were also excluded. Thus, a total of 87 spectra collected from 64 patients were included in the final analysis (Table 1). The abbreviations for the 9 sites are also shown in Table 1. The average age (±SD) of the 64 patients was 56 ± 15 years. Data collected from HVs on the BM, DT, FM, GI, HP, LT, RT, SP, and VT were used in the subsequent analysis. These groups consisted of 100, 124, 112, 58, 54, 112, 33, 43, and 74 spectra, respectively (not shown in Table 1).

Identification of Spectrally Similar Sites

The set of 710 spectra collected from 9 anatomic sites in HVs was used to calculate similarity scores for each pair of sites and to identify which specific pairs could be combined. There were pronounced differences in the spectral properties among the various anatomic sites (Table 2). For 75% of the comparisons, the similarity score was 7 or less. We identified sites with a high degree of similarity on the basis of a similarity score of at least 10. The following 3 pairs of sites met this criterion: 1) BM and SP, 2) FM and VT, and 3) GI and HP. We refer to pairs of sites with high similarity scores as spectrally similar sites. Just as for the dysplastic/malignant category, we use the slash notation (eg, FM/VT) in the text that follows to refer to a group that combines samples from 2 sites.

For algorithm development, it is important to have a sufficient number of samples in each category. In this study, we used the criterion of at least 10 samples per category. Table 3 lists the numbers of benign and dysplastic/malignant lesions for each of the 3 pairs. For the FM/VT group, we analyzed samples from every category (healthy mucosa, lesions, benign lesions, and dysplastic/malignant lesions). For the BM/SP group, we analyzed samples from the healthy mucosa, lesion, and benign lesion categories.

Distinguishing Healthy Mucosa From Lesions

We first examined the benefit of anatomy-based algorithms when applied to the distinction of healthy mucosa samples from lesions. Table 4 lists the ROC-AUC results for discriminating healthy mucosa (HV data) from visible lesions (patient data). In each case, the results are shown for all sites combined (9 sites total), nonkeratinized sites (all sites except the GI and HP), individual sites, and spectrally similar sites. Our prior study in HVs clearly demonstrated that the GI and HP display significant spectral contrast from most other sites for several spectroscopy parameters.[25] By analyzing nonkeratinized sites separately, we eliminate spectral contrast due to keratin, and can determine whether eliminating it improves the identification of lesions among the remaining sites in the group.

The results in Table 4 indicate that excellent discrimination between healthy mucosa and lesions is possible with spectroscopy (ROC-AUC, 0.81 to 0.97), particularly when the diagnostic algorithms are developed and applied to individual sites or spectrally similar sites. The diagnostic performance of the spectral algorithms was poorest for the group in which all sites were combined (ROC-AUC, 0.81 to 0.85). The spectroscopy parameters frequently identified as diagnostic included NADH (340 nm), cβC, tryptophan (308 nm), Coll427 (340 nm), and α. The effective vessel radius, Coll (308 nm), and A and C parameters were not diagnostic.

When lesions were compared to healthy mucosa, the parameter trends were found to be identical for all 3 separations: healthy mucosa versus all lesions, benign lesions, and dysplastic/malignant lesions, respectively. Lesions were characterized by higher B parameter, cHb, α, tryptophan (308 nm), and NADH (340 nm) and by lower cβC, NADH (308 nm), Coll401 (340 nm), and Coll427 (340 nm). We compared the parameter trends for lesions relative to healthy mucosa to the trends for healthy keratinized sites (GI and HP) relative to healthy nonkeratinized sites. The latter comparison was used as a marker for the influence of keratin; however, mucosal features specific to the GI and HP may also contribute to the trends. The trends for both comparisons were identical for 7 of the 9 spectroscopy parameters employed in the spectral diagnostic models (data not shown).

Distinguishing Benign Lesions From Dysplastic/Malignant Lesions

We next evaluated the benefit of anatomy-based algorithms for a clinically relevant distinction, the discrimination of benign lesions from dysplastic/malignant lesions. Spectral

algorithms were developed for the 4 groups with at least 10 samples in the 2 lesion categories: all sites combined, nonkeratinized sites, the LT alone, and the FM/VT group. The Figure shows ROC curves for all sites combined, the LT alone, and the FM/VT group. The ROC curve for all sites combined demonstrated the poorest discriminatory power.

Table 5 lists the ROC-AUC values, sensitivities, and specificities for the discrimination of benign lesions from dysplastic/malignant lesions. The highest ROC-AUC value (0.75) was obtained for the LT, and once again, the lowest value (0.60) was obtained when all sites were combined. The diagnostic performance of the FM/VT group was considerably better than that of all sites combined.

Table 6 lists the spectroscopy parameters that were combined to yield the performance values shown in Table 5. The diagnostic parameters were almost entirely different from those used to separate healthy mucosa from lesions (data not shown). Although the C parameter was not diagnostic in distinguishing healthy mucosa from lesions, it played a significant role in separating benign lesions from dysplastic/malignant lesions. Similarly, Coll401 (340 nm) had greater significance in separating benign lesions from dysplastic/malignant lesions than in separating healthy mucosa from lesions. The cβC continued to be an important diagnostic parameter. Dysplastic/malignant lesions were characterized by higher C parameter and cHb and by lower cβC, Coll401 (340 nm), and Coll427 (340 nm).

DISCUSSION

The oral cavity is complex, in that it consists of multiple morphologically distinct anatomic sites, each of which is characterized by its own spectral properties. In many clinical studies, only a limited amount of data can be collected in a practical time period; therefore, data from all anatomic sites are combined to maximize the number of samples in each category. Consequently, differences in the histologic characteristics between the anatomic sites may confound the detection of spectral contrast related to the presence of disease. We developed a strategy to evaluate whether certain pairs of sites are comparable in their spectral properties so that one could combine those sites with a high degree of similarity. This approach was designed to address 2 potential pitfalls: 1) combining dissimilar sites may broaden parameter distributions and make it more difficult to detect disease-related spectral contrast and 2) if the sites being combined have different spectral properties and samples are unevenly distributed between the 2 categories being compared, the likelihood is increased that anatomic spectral contrast may compete with or contribute to the apparent disease-related contrast. We identified 3 pairs of spectrally similar sites: 1) BM and SP, 2) FM and VT, and 3) GI and HP.

Distinguishing Healthy Mucosa From Lesions

The results of this comparison demonstrated the significant influence of anatomic spectral contrast on diagnostic accuracy, as well as a clear benefit to combining spectrally similar sites. As in other studies, we obtained excellent results for this comparison; however, anatomy-based algorithms were more accurate (ROC-AUC, 0.90 to 0.97) than were algorithms developed for all sites combined (ROC-AUC, 0.81 to 0.85). We tested an alternative approach for combining multiple sites on the basis of a morphological criterion, only combining nonkeratinized sites together. The performance of the diagnostic algorithm for the nonkeratinized group did not produce ROC-AUC values dramatically higher than those for all sites combined. The number of lesion samples from keratinized sites in our data set was limited (5 samples total); however, this finding is consistent with our previous observation of significant anatomic spectral contrast among nonkeratinized sites, and supports the need for a more effective strategy for determining which sites to combine.[25]

The excellent separations achieved for comparisons between healthy mucosa and lesions may be influenced by hyperkeratosis, ulceration, or other mucosal changes associated with lesions that are not directly linked to malignancy. The spectral contrast between healthy mucosa and visible lesions is largely independent of the presence of disease, as shown by the comparable ROC-AUC values for discriminating healthy mucosa from benign lesions and healthy mucosa from dysplastic/malignant lesions (0.81 to 0.96 and 0.85 to 0.97, respectively). The parallel parameter trends for lesions compared to healthy mucosa and healthy keratinized sites compared to healthy nonkeratinized sites provide added support for the hypothesis that keratin, rather than disease, is a major source of spectral contrast. These results indicate that clinically healthy mucosa samples and samples from clinically abnormal mucosa (lesions) should be treated as distinct entities. The high values for the diagnostic accuracy reported for this comparison in previous studies may in fact be unrealistically high because of the confounding effects of mucosal changes that are unrelated to disease.

The parameter trends we observed when comparing lesions to healthy mucosa are consistent with those of other studies. In distinguishing healthy mucosa from malignant lesions, Amelink et al[12] noted that malignant lesions exhibited a lower α, an increase in blood content, and an increase in scattering slope (B parameter). A decrease in scattering amplitude (A parameter) was also observed, which was not found in our analysis. Müller et al[9] also noted increased NADH and decreased collagen fluorescence for lesions excited at 340 nm. The increase in the B parameter is most likely due to scattering from keratin present in leukoplakias, as previous work has shown that this parameter is greatly affected by this feature.[25] Vascular dilation or inflammation in ulcerated lesions or erythroplakias may give rise to an increased cHb. The decreased collagen fluorescence at 340 nm excitation, which may result from degradation of the extracellular matrix by matrix metalloproteinases, has been noted in numerous spectroscopic studies.[27,39–41] The increased NADH at 340 nm excitation may be due to the increased metabolic activity of abnormal proliferating epithelial cells.

Distinguishing Benign Lesions From Dysplastic/Malignant Lesions

The results for the distinction of healthy mucosa from lesions demonstrated the benefit of anatomy-based algorithms, but this application has no real clinical importance. Therefore, we applied the anatomy-based algorithms to the distinction of benign lesions from dysplastic/malignant lesions. Once again, the best diagnostic performance was obtained for the individual site (LT), and the poorest performance for all sites combined (ROC-AUC, 0.60). The FM/VT group (spectrally similar sites) also produced more accurate results (ROC-AUC, 0.71) than did all sites combined. We did not evaluate the BM/SP group, because it did not meet the criterion of at least 10 samples in each category.

The ROC-AUC values obtained for the distinction of benign lesions from dysplastic/ malignant lesions were significantly lower than those for the distinction of healthy mucosa from lesions. Furthermore, very few parameters were statistically significant in the diagnostic model for the former comparison. Therefore, the diagnostic accuracies reported in studies comparing healthy mucosa to lesions may be misleading, because they do not reflect how spectroscopy performs when applied to clinically important diagnostic distinctions. The heterogeneity of lesions within each category contributes to the challenge of distinguishing lesion categories and can further obscure subtle disease-related changes in the spectroscopy parameters. Some malignant lesions had small foci of invasion in a large area of relatively normal mucosa, whereas others demonstrated more diffuse disease. Similarly, the benign lesions ranged from mildly to highly hyperkeratotic, with various degrees of hyperplasia and inflammation. Interindividual variation is another source of spectral contrast that may affect diagnostic performance.

Another approach for combining sites, employed by some researchers, is to combine healthy mucosa samples with benign lesions. This can dramatically increase the number of samples in the negative-for-dysplasia/malignancy category, because healthy mucosa samples are readily available. However, this approach may produce misleading results for 2 reasons. First, we have shown that there is significant spectral contrast between healthy mucosa and lesions even in the absence of disease. Second, our findings demonstrate a marked difference in spectral contrast between benign lesions and dysplastic/malignant lesions, as compared to healthy mucosa and visible lesions. Therefore, the diagnostic performance may falsely appear to improve as the proportion of healthy mucosa samples in the negative group increases. To test this hypothesis, we combined 10 LT healthy mucosa samples with the benign LT lesions. Whereas an ROC-AUC value of 0.75 was obtained for distinguishing benign lesions from dysplastic/malignant lesions, the ROC-AUC value increased to 0.85 when we used the combined healthy mucosa/benign lesion category (data not shown). When 50 LT healthy mucosa samples were combined with the same benign samples, the ROC-AUC value increased further, to 0.91. These results underscore the importance of considering how data are combined in order to reliably evaluate spectral diagnostic algorithms.

We examined the trends in the spectral parameters used in the diagnostic models with increasing disease severity. The increase in the C parameter is consistent with increased scattering by small particles, as may occur with increased epithelial proliferation. The higher cHb may be the result of vascular dilation and inflammation as seen in erythroplakias or with ulceration (features that are more likely to be associated with dysplasia/malignancy), whereas white lesions are less likely to be dysplastic/malignant upon biopsy.[42,43] Decreased collagen fluorescence was observed once again for this distinction. The reason for the decreased cβC is unknown. β-Carotene is an antioxidant and can be metabolized into vitamin A, which is involved in the differentiation of normal epithelial cells.[44]

Application of Spectral Algorithms in Clinical Practice

The current standard screening method for oral cancer — inspection with palpation — enables inspection of the entire oral cavity. Our probe may not be appropriate for this purpose, because it samples only a small area of tissue. However, when patients with suspicious lesions are referred to an otolaryngologist, spectroscopy may be useful for distinguishing benign lesions from dysplastic/malignant lesions. A study by Waldron and Shafer[42] of 3,256 leukoplakias (the most commonly encountered oral lesions) found that 42.9% of biopsies from FM lesions and 24.2% of biopsies from tongue lesions exhibited dysplasia/malignancy. Of the 3,256 cases of leukoplakia, 6.8% and 8.6%, respectively, were located at these sites. If we consider 1,000 patients with leukoplakia of the tongue or FM, a total of 442 FM lesions and 558 tongue lesions would be expected, given the relative frequency of leukoplakia at these 2 sites noted in that study. Of these patients, 32.5% would be expected to show dysplasia/malignancy (42.9% of FM lesions and 24.2% of tongue lesions). The performance values for the LT and FM/VT sites are similar, so we use the average values of 93% and 64% as the sensitivity and specificity, respectively, of the spectral diagnostic test. Based on these values and the aforementioned prevalence estimates for diseased lesions, the positive and negative predictive values for our test are 55% and 95%, respectively. These results indicate that our spectral algorithms can identify the absence of dysplasia/malignancy, with few false negatives, thereby reducing the number of unnecessary biopsies. A large prospective study is needed to rigorously confirm these findings.

CONCLUSIONS

The results of this study demonstrate the importance of developing anatomy-based algorithms for disease detection in the oral cavity. Combining multiple anatomic sites without accounting for anatomic spectral contrast reduced disease-related contrast and resulted in poor diagnostic performance. Spectral algorithms designed specifically for multiple anatomic sites sharing a high degree of spectral similarity considerably improved diagnostic accuracies. The best diagnostic performance was achieved when anatomy-based algorithms were developed for individual sites. The discrimination of healthy mucosa from lesions yielded excellent results, as in previous studies, but the findings do not extend to the clinically relevant distinction of benign lesions from dysplastic/malignant lesions. However, anatomy-based algorithms significantly improved diagnostic performance for the latter distinction. Diagnostic accuracies for distinguishing healthy mucosa/benign lesions from dysplastic/malignant lesions were dependent on the proportion of healthy mucosa samples and were therefore unreliable, even with the use of anatomy-based algorithms. This study demonstrates a successful approach for developing common algorithms for multiple anatomic sites while accounting for anatomic spectral contrast.

Developing a spectral diagnostic algorithm for every anatomic site may not be necessary to provide a substantial clinical benefit. One potential future direction is to develop a reliable diagnostic tool designed specifically for sites at high risk for oral cancer. To ensure that dysplasia or malignancy is not missed, biopsies are likely to be taken of lesions at these sites; therefore, a tool for identifying benign lesions could reduce the number of unnecessary biopsies. Alternatively, unnecessary biopsies could be reduced by developing a tool for sites at which lesions are frequently encountered but generally prove benign by histopathology (low likelihood of disease). Large, prospective clinical studies are needed to reliably evaluate the efficacy of spectroscopy for these applications. Ultimately, we would like to detect invisible disease, particularly in high-risk individuals and at surgical margins.

Acknowledgments

References

1. Pindborg, J.; Reichart, P.; Smith, C.; van der Waal, I. World Health Organization: histological typing of cancer and precancer of the oral mucosa. Berlin, Germany: Springer-Verlag; 1997.

2. Pentenero M, Carrozzo M, Pagano M, et al. Oral mucosal dysplastic lesions and early squamous cell carcinomas: underdiagnosis from incisional biopsy. Oral Dis 2003;9:68–72. [PubMed: 12657031]

3. Lee JJ, Hung HC, Cheng SJ, et al. Factors associated with underdiagnosis from incisional biopsy of oral leukoplakic lesions. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2007;104:217–25. [PubMed: 17560138]

4. Fischer DJ, Epstein JB, Morton TH Jr, Schwartz SM. Reliability of histologic diagnosis of clinically normal intraoral tissue adjacent to clinically suspicious lesions in former upper aerodigestive tract cancer patients. Oral Oncol 2005;41:489–96. [PubMed: 15878753]

5. Fischer DJ, Epstein JB, Morton TH, Schwartz SM. Interobserver reliability in the histopathologic diagnosis of oral premalignant and malignant lesions. J Oral Pathol Med 2004;33:65–70. [PubMed: 14720191]

6. van der Waal I, Schepman KP, van der Meij EH, Smeele LE. Oral leukoplakia: a clinicopathological review. Oral Oncol 1997;33:291–301. [PubMed: 9415326]

7. Gillenwater A, Jacob R, Ganeshappa R, et al. Noninvasive diagnosis of oral neoplasia based on fluorescence spectroscopy and native tissue autofluorescence. Arch Otolaryngol Head Neck Surg 1998;124:1251–8. [PubMed: 9821929]

8. Majumder SK, Ghosh N, Gupta PK. Support vector machine for optical diagnosis of cancer. J Biomed Opt 2005;10:024034. [PubMed: 15910107]

9. Müller MG, Valdez TA, Georgakoudi I, et al. Spectroscopic detection and evaluation of morphologic and biochemical changes in early human oral carcinoma. Cancer 2003;97:1681–92. [PubMed: 12655525]

10. Onizawa K, Okamura N, Saginoya H, Yusa H, Yanagawa T, Yoshida H. Analysis of fluorescence in oral squamous cell carcinoma. Oral Oncol 2002;38:343–8. [PubMed: 12076697]

11. van Staveren HJ, van Veen RLP, Speelman OC, Witjes MJH, Star WM, Roodenburg JLN. Classification of clinical autofluorescence spectra of oral leukoplakia using an artificial neural network: a pilot study. Oral Oncol 2000;36:286–93. [PubMed: 10793332]

12. Amelink A, Kaspers OP, Sterenborg H, van der Wal JE, Roodenburg JLN, Witjes MJH. Non-invasive measurement of the morphology and physiology of oral mucosa by use of optical spectroscopy. Oral Oncol 2008;44:65–71. [PubMed: 17350324]

13. Sharwani A, Jerjes W, Salih V, et al. Assessment of oral premalignancy using elastic scattering spectroscopy. Oral Oncol 2006;42:343–9. [PubMed: 16321565]

14. Heintzelman DL, Utzinger U, Fuchs H, et al. Optimal excitation wavelengths for in vivo detection of oral neoplasia using fluorescence spectroscopy. Photochem Photobiol 2000;72:103–13. [PubMed: 10911734]

15. Kolli VR, Savage HE, Yao TJ, Schantz SP. Native cellular fluorescence of neoplastic upper aerodigestive mucosa. Arch Otolaryngol Head Neck Surg 1995;121:1287–92. [PubMed: 7576476]

16. Majumder SK, Ghosh N, Kataria S, Gupta PK. Nonlinear pattern recognition for laser-induced fluorescence diagnosis of cancer. Lasers Surg Med 2003;33:48–56. [PubMed: 12866121]

17. Manjunath BK, Kurein J, Rao L, et al. Autofluorescence of oral tissue for optical pathology in oral malignancy. J Photochem Photobiol B 2004;73:49–58. [PubMed: 14732251]

18. Schantz SP, Kolli V, Savage HE, et al. In vivo native cellular fluorescence and histological characteristics of head and neck cancer. Clin Cancer Res 1998;4:1177–82. [PubMed: 9607575]

19. Wang CY, Chiang HK, Chen CT, Chiang CP, Kuo YS, Chow SN. Diagnosis of oral cancer by light-induced autofluorescence spectroscopy using double excitation wavelengths. Oral Oncol 1999;35:144–50. [PubMed: 10435148]

20. Schwarz RA, Gao W, Redden Weber C, et al. Noninvasive evaluation of oral lesions using depth-sensitive optical spectroscopy. Cancer 2009;115:1669–79. [PubMed: 19170229]

21. Mallia RJ, Thomas SS, Mathews A, et al. Laser-induced autofluorescence spectral ratio reference standard for early discrimination of oral cancer. Cancer 2008;112:1503–12. [PubMed: 18260154]

22. Dhingra JK, Perrault DF Jr, McMillan K, et al. Early diagnosis of upper aerodigestive tract cancer by autofluorescence. Arch Otolaryngol Head Neck Surg 1996;122:1181–6. [PubMed: 8906052]

23. de Veld DCG, Skurichina M, Witjes MJH, et al. Autofluorescence characteristics of healthy oral mucosa at different anatomical sites. Lasers Surg Med 2003;32:367–76. [PubMed: 12766959]

24. Pavlova I, Williams M, El-Naggar A, Richards-Kortum R, Gillenwater A. Understanding the biological basis of autofluorescence imaging for oral cancer detection: high-resolution fluorescence microscopy in viable tissue. Clin Cancer Res 2008;14:2396–404. [PubMed: 18413830]

25. McGee S, Mirkovic J, Mardirossian V, et al. Model-based spectroscopic analysis of the oral cavity: impact of anatomy. J Biomed Opt 2008;13:064034. [PubMed: 19123680]

26. Mallia RJ, Thomas SS, Mathews A, et al. Oxygenated hemoglobin diffuse reflectance ratio for in vivo detection of oral pre-cancer. J Biomed Opt 2008;13:041306. [PubMed: 19021314]

27. Mueller MM, Fusenig NE. Tumor-stroma interactions directing phenotype and progression of epithelial skin tumor cells. Differentiation 2002;70:486–97. [PubMed: 12492491]

28. de Veld DCG, Skurichina M, Witjes MJH, Duin RPW, Sterenborg HJCM, Roodenburg JLN. Autofluorescence and diffuse reflectance spectroscopy for oral oncology. Lasers Surg Med 2005;36:356–64. [PubMed: 15856507]

29. de Veld DCG, Skurichina M, Witjes MJH, Duin RPW, Sterenborg HJCM, Roodenburg JLN. Clinical study for classification of benign, dysplastic, and malignant oral lesions using autofluorescence spectroscopy. J Biomed Opt 2004;9:940–50. [PubMed: 15447015]

30. Wang CY, Tsai T, Chen HM, Chen CT, Chiang CP. PLS-ANN based classification model for oral submucous fibrosis and oral carcinogenesis. Lasers Surg Med 2003;32:318–26. [PubMed: 12696101]

31. Tunnell JW, Desjardins AE, Galindo L, et al. Instrumentation for multi-modal spectroscopic diagnosis of epithelial dysplasia. Technol Cancer Res Treat 2003;2:505–14. [PubMed: 14640762]

32. Zonios G, Perelman LT, Backman V, et al. Diffuse reflectance spectroscopy of human adenomatous colon polyps in vivo. Appl Opt 1999;38:6628–37. [PubMed: 18324198]

33. Müller MG, Georgakoudi I, Zhang QG, Wu J, Feld MS. Intrinsic fluorescence spectroscopy in turbid media: disentangling effects of scattering and absorption. Appl Opt 2001;40:4633–46. [PubMed: 18360504]

34. Mourant JR, Freyer JP, Hielscher AH, Eick AA, Shen D, Johnson TM. Mechanisms of light scattering from biological cells relevant to noninvasive optical-tissue diagnostics. Appl Opt 1998;37:3586–93. [PubMed: 18273328]

35. Zhang Q, Müller MG, Wu J, Feld MS. Turbidity-free fluorescence spectroscopy of biological tissue. Opt Lett 2000;25:1451–3. [PubMed: 18066245]

36. Wilcox, RR. Fundamentals of modern statistical methods: substantially improving power and accuracy. New York, NY: Springer; 2001.

37. Kleinbaum, DG.; Klein, M.; Pryor, ER. Logistic regression: a self-learning text. New York, NY: Springer; 2002.

38. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. Am J Epidemiol 2006;163:670–5. [PubMed: 16410346]

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

40. Jordan RCK, Macabeo-Ong M, Shiboski CH, et al. Over-expression of matrix metalloproteinase–1 and –9 mRNA is associated with progression of oral dysplasia to cancer. Clin Cancer Res 2004;10:6460–5. [PubMed: 15475433]

41. Tosios KI, Kapranos N, Papanicolaou SI. Loss of basement membrane components laminin and type IV collagen parallels the progression of oral epithelial neoplasia. Histopathology 1998;33:261–8. [PubMed: 9777393]

42. Waldron CA, Shafer WG. Leukoplakia revisited. A clinicopathologic study 3256 oral leukoplakias. Cancer 1975;36:1386–92. [PubMed: 1175136]

43. Reichart PA, Philipsen HP. Oral erythroplakia — a review. Oral Oncol 2005;41:551–61. [PubMed: 15975518]

44. Steinmetz KA, Potter JD. Vegetables, fruit, and cancer prevention: a review. J Am Diet Assoc 1996;96:1027–39. [PubMed: 8841165]
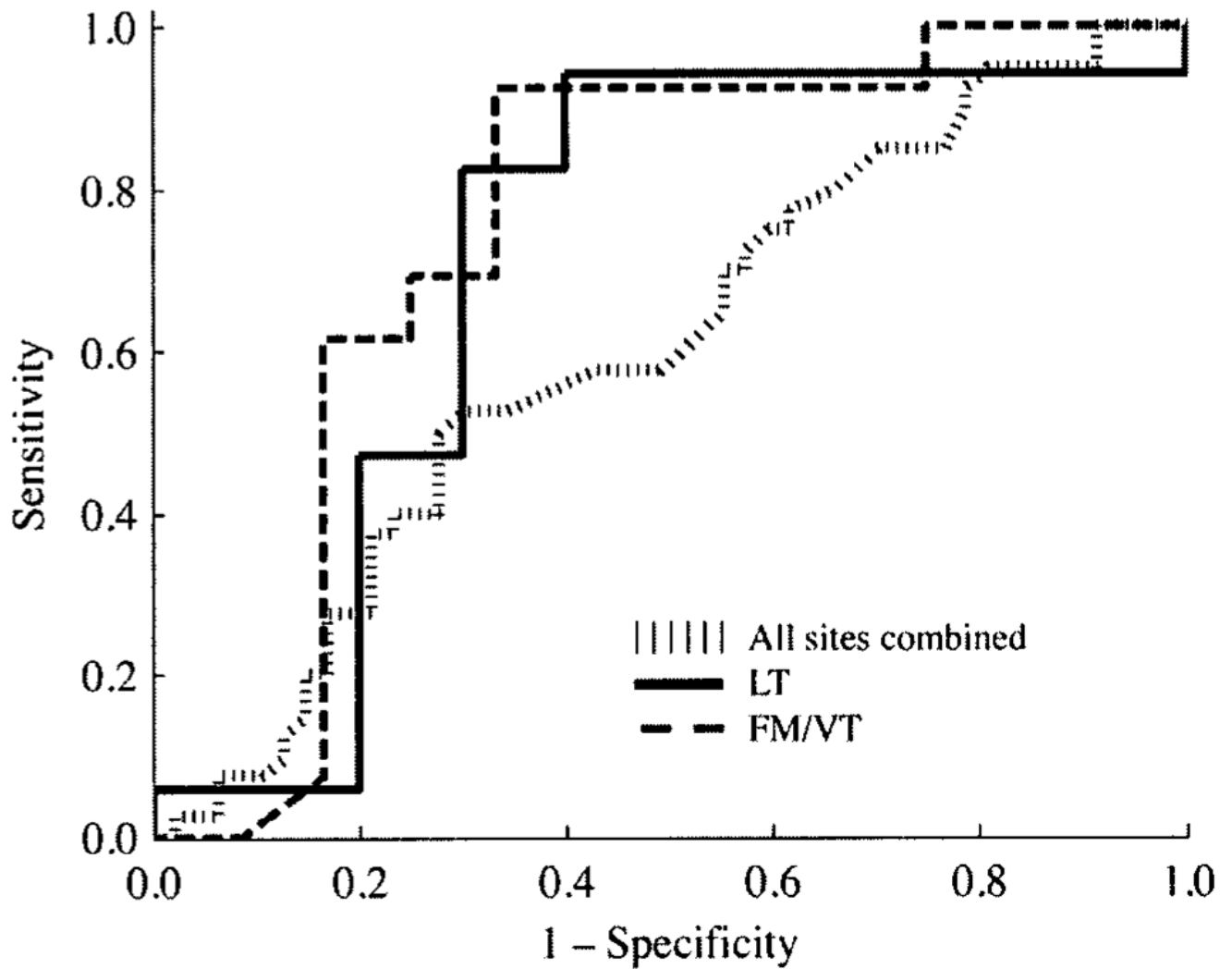
1. .

Receiver operator characteristic curves for discrimination of benign lesions from dysplastic/ malignant lesions for all sites combined (hatched line), lateral surface of tongue (LT; thick dashed line), and floor of mouth/ventral surface of tongue (FM/VT; thick solid line).

TABLE 1

TOTAL NUMBER OF SPECTRA MEASURED AT EACH ANATOMIC SITE IN PATIENTS AND PATHOLOGICAL CLASSIFICATION

Site	Malignant	Dysplastic	Benign	Total
Buccal mucosa (BM)	2	0	11	13
Dorsal surface of tongue (DT)	0	0	4	4
Floor of mouth (FM)	2	11	3	16
Gingiva (GI)	0	1	0	1
Hard palate (HP)	1	1	2	4
Lateral surface of tongue (LT)	4	9	12	25
Retromolar trigone (RT)	0	1	3	4
Soft palate (SP)	4	0	5	9
Ventral surface of tongue (VT)	1	3	7	11
Total	14	26	47	87

TABLE 2

SIMILARITY SCORES FOR EACH PAIR OF ANATOMIC SITES

Site	DT	FM	GI	HP	LT	RT	SP	VT
BM	4	5	2	1	5	8	10	5
DT		6	6	6	3	9	7	3
FM			3	1	5	3	9	12
GI				11	3	7	7	2
HP					3	6	4	0
LT						5	6	8
RT							8	3
SP								8

TABLE 3

DESCRIPTION OF SPECTRALLY SIMILAR ANATOMIC SITES

Similarity Score	Spectrally Similar Sites	Dysplastic/Malignant	Benign
10	BM and SP	6	16
12	FM and VT	17	10
11	GI and HP	3	2

TABLE 4

ROC-AUC VALUES FOR DISCRIMINATION OF HEALTHY MUCOSA FROM LESIONS

Group	Healthy Mucosa vs Lesions	Healthy Mucosa vs Dysplastic/Malignant Lesions	Healthy Mucosa vs Benign Lesions
All sites	0.83	0.85	0.81
Nonkeratinized sites	0.84	0.88	0.84
LT	0.95	0.97	0.93
FM	0.94	0.95	
BM[*]	0.94		0.96
VT	0.95		
FM/VT	0.97	0.97	0.94
BM/SP[*]	0.91		0.90

ROC-AUC – area under receiver operator characteristic curve.

[*] No dysplastic samples.

TABLE 5

ROC-AUC, SENSITIVITY, AND SPECIFICITY VALUES FOR DISCRIMINATION OF BENIGN LESIONS FROM DYSPLASTIC/MALIGNANT LESIONS

Group	ROC-AUC	Sensitivity (Based on Youden Index)	Specificity (Based on Youden Index)
All sites	0.60	53%	70%
Nonkeratinized sites	0.63	49%	76%
LT	0.75	92%	67%
FM/VT	0.71	94%	60%

TABLE 6

SIGNIFICANT DIAGNOSTIC PARAMETERS FOR DISCRIMINATION OF BENIGN LESIONS FROM DYSPLASTIC/MALIGNANT LESIONS

Group	Diagnostic Parameters for Benign Lesions vs Dysplastic/Malignant Lesions
All sites	C parameter, Coll401 (340 nm), cβC
Nonkeratinized sites	C parameter, Coll401 (340 nm), cβC
LT	cβC, Coll401 (340 nm)
FM/VT	C parameter, Coll427 (340 nm), cHb
BM/SP[*]	A parameter

[*] No dysplastic samples.