J Oral Pathol Med (2004) 33: 65–70

www.blackwellmunksgaard.com/jopm

Interobserver reliability in the histopathologic diagnosis of oral pre-malignant and malignant lesions

Dena J. Fischer[1,2], Joel B. Epstein[1,3], Thomas H. Morton, Jr[1,4], Stephen M. Schwartz[2,5]

[1]*Department of Oral Medicine, School of Dentistry, Box 356370, University of Washington, Seattle, WA 98195-6370,* [2]*Department of Epidemiology, School of Public Health and Community Medicine, Box 357236, University of Washington, Seattle, WA 98195-7236,* [3]*Department of Oral Medicine and Diagnostic Sciences, College of Dentistry, University of Illinois at Chicago, Chicago, IL 60612,* [4]*Department of Oral Biology, School of Dentistry, Box 357132, University of Washington, Seattle, WA 98195-7132, and* [5]*Program in Epidemiology, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA*

BACKGROUND: The histologic classification of pre-cancerous and cancerous oral lesions has generally shown poor agreement between pathologists, but lesion and patient characteristics that may affect diagnostic reliability have not been explored.

METHODS: Eighty-seven clinically suspicious oral lesions biopsied from 81 patients with previous upper aerodigestive tract cancer were independently classified by their local pathologist and a central pathology committee. Interobserver reliability between the local pathologist and the central pathology committee was measured with weighted kappa (κ_w) statistics and corresponding 95% confidence intervals (CI).

RESULT: The κ_w for pathologic diagnosis was 0.59 (95% CI: 0.45, 0.72), and was higher for lesions without inflammation (0.67 (95% CI: 0.53, 0.80) than inflamed lesions (−0.10 (95% CI: −0.27, 0.07)). Greatest agreement was seen for lesions located in the buccal mucosa/vestibule ($\kappa_w = 0.68$ (95% CI: 0.46, 0.91)) and tongue ($\kappa_w = 0.62$ (95% CI: 0.40, 0.84)). Least agreement was found for lip/labial mucosa lesions ($\kappa_w = -0.04$ (95% CI: −0.34, 0.27)). Punch biopsies ($\kappa_w = 0.67$ (95% CI: 0.54, 0.80)) had greater interobserver reliability than wedge biopsies ($\kappa_w = 0.38$ (95% CI: 0.12, 0.64)).

CONCLUSIONS: These data suggest that the presence of inflammation, lesion site, and biopsy technique modifies the reliability of oral lesion histologic diagnoses.

J Oral Pathol Med (2004) **33**: 65–70

Keywords: mouth diseases/diagnosis; mouth mucosa/pathology; mouth neoplasms/diagnosis; observer variation; oral leukoplakia/diagnosis; oral pathology

Correspondence: Dena J. Fischer, DDS, MSDA, School of Dentistry, Box 356370, University of Washington, Seattle, WA 98195-6370, USA. Tel.: +1 206 221 7014. Fax: +1 206 685 8412. E-mail: denafis@u.washington.edu
Accepted for publication June 25, 2003

Introduction

It is estimated that 27 700 new cases of oral and pharyngeal cancer will be diagnosed in the US during 2003, while 7200 people will die of the disease (1). Because of their location, oral and pharyngeal carcinomas can also be associated with significant morbidity, as these cancers and treatments thereof can be disfiguring and can affect daily functions, negatively impacting quality of life. The 5- and 10-year survival rates are low: 56 and 41%, respectively (1). The survival rates have remained relatively unchanged for the past three decades, probably because of late recognition of the disease (2). Consequently, there is an increasing emphasis on early diagnosis and close monitoring of pre-cancerous lesions.

Because of the variable clinical presentation and the potential for transformation of suspicious lesions, the standard of care is to perform biopsies to obtain 'gold standard' diagnoses. These clinically suspicious lesions are evaluated histologically for the presence and extent of dysplasia. Lesions that exhibit dysplasia are considered pre-neoplastic, with the transformation rate of dysplasia to cancer having been reported as high as 36.4% (3). Epithelial pre-neoplastic and neoplastic oral lesions comprise a continuum from mild dysplasia to carcinoma, with specific criteria of oral epithelial changes defining each stage along the continuum (4, 5). However, because of the subjective nature of histologic diagnoses, individual variations have existed among observers, and further, there has been disagreement as to which criteria contribute more to a lesion's diagnosis (6). This variability is particularly problematic, as incorrect or inconsistent diagnoses may lead to over- or undertreatment and different prognoses for the same lesion.

There have been few formal studies to determine the reliability of pathologic diagnosis for oral pre-cancerous and cancerous lesions. Although criteria exist for histologic classification of these lesions, four previous studies have shown generally poor agreement between pathologists when making diagnoses (7–10). However, these studies were limited by the fact that they did not evaluate characteristics

of lesions that may have affected the reliability. Also, they were limited in the number of pathologists, thereby reducing generalizability of the results. We studied interobserver reliability in the diagnosis of oral epithelial lesions in patients with previous upper aerodigestive tract (UADT) cancers, including whether specific characteristics of lesions (e.g. tumor site, inflammation, and oral cancer risk factor status) influenced the magnitude of interobserver differences.

Materials and methods

The data for this analysis were collected as part of a multicenter study of Tolonium Chloride (OraTest^{TM}, Zila Biomedical Inc., Phoenix, AZ, USA), a rinse designed to assist clinical examination and biopsy site selection of clinically suspicious oral lesions. Patients who had been previously treated for UADT carcinoma were enrolled at 10 centers in the USA and one each in Canada and the UK. All patients provided informed consent under procedures approved by institutional review boards. Inclusion criteria were: age 18 years or older, treated within the past 2 years for UADT carcinoma, capable of providing informed consent, and able and willing to follow instructions. Exclusion criteria were: participation in any experimental study of a drug or device within 30 days of entry, treatment for UADT in the 3 months prior to entry, pregnant/breast feeding females, and any medical condition that would prevent study participation.

After participants provided written consent, a study investigator conducted a thorough examination of each patient's oral mucosa. The exact location of all lesions were noted on a case report form grid, which contained a schematic of dorsal and ventral view of the mouth, including lips, labial mucosa, gingiva, vestibule, buccal mucosa, floor of mouth, hard palate, soft palate, anterior/posterior pillar, and tongue. If a lesion was judged to require immediate histopathologic evaluation, a biopsy was completed at this first visit. Other lesions were biopsied if judged to be clinically suspicious upon re-examination 10–21 days later. Lesions that presented clinically as: (i) homogeneous leukoplakia; (ii) non-homogeneous leukoplakia (erythroleukoplakia); or (iii) erythroplakia were determined to be clinically suspicious (4). Either wedge or 3–4-mm punch biopsies were performed according to the standard procedures at each investigational center. All biopsy specimens were placed in a container with 10% formalin labeled, and sent to the local collaborating pathologist.

The study protocol called for each lesion to be diagnosed by both the local, experienced collaborating pathologist as well as each of two experienced pathologists at the central pathology laboratory. There were 21 local and 3 central pathologists with a mean of 18.6 ± 9.8 (range: 3–35) and 22.0 ± 11.0 years (range: 11–33) of experience, respectively. The breakdown of pathology specialty was as follows: 10 (41.7%) general cancer pathology; 4 (16.7%) general pathology; 4 (16.7%) cancer pathology; 4 (16.7%) head and neck cancer pathology; and 2 (8.3%) oral and maxillofacial pathology. The three central pathologists included two head and neck cancer pathologists and one oral and maxillofacial pathologist. At some field sites, the local pathology laboratory provided the central pathologist with the identical slides upon which the local diagnosis was based. At other field sites, the central pathologists were provided with slides from the sections made immediately adjacent to the biopsy specimen sections that yielded the slides for the local pathology review, i.e. two sets of slides were produced, prepared from alternating sections. Local and central pathologists were blind to the clinical findings, and all pathologists performed their assessments independently of one another. If the two diagnoses among the central pathologists did not agree, the specimen was evaluated by a third pathologist in the central laboratory, and the diagnoses of all three pathologists were reported. In this study, the central pathologists were not in complete agreement for three specimens.

Each biopsy was categorized histologically as: no abnormality, benign, dysplasia, carcinoma *in situ*, invasive carcinoma, and 'other'. Nine of 96 total lesions were categorized as 'other', and they all were diagnosed as radiation fibrosis. Consequently, these lesions were excluded from the analysis. Dysplasia was characterized as cellular (cytologic) abnormalities, which may include: variation in cell size, morphology, and/or staining characteristics, increased and abnormal mitotic figures, or maturation orientation. The pathologists used individual histopathologic experience and expertise in evaluating such cellular abnormalities, rather than collectively agreeing upon the specific morphologic characteristics that should be considered relevant for grading of dysplasia. When dysplasia was seen in the full thickness of the epithelium in any one microscopic field, the diagnosis was carcinoma *in situ*, while squamous cell carcinoma involved disruption of the basement membrane and invasion into the lamina propria by dysplastic cells. Benign lesions were subcategorized as exhibiting keratosis, hyperkeratosis, or hyperplasia, whereas dysplastic lesions were subclassified as either mild, moderate, or severe. Each pathologist also characterized each biopsy according to level (amount) of inflammation (none, mild, moderate, or severe), and the extent (thickness) of dysplasia in the epithelium was recorded for dysplastic lesions ($\leq$1/3, $\leq$2/3, and $>$2/3). Treatment at each center was based upon the diagnosis made by the local pathologist. However, for the purposes of this study, the diagnosis made by the central pathology laboratory was accepted as the definitive diagnosis.

Interobserver reliability was measured with unweighted and weighted kappa (κ) statistics and corresponding confidence intervals (CI) (11). Kappa is a measure of agreement that corrects for the agreement that would be expected by chance. The unweighted kappa considers all disagreements considered to be equally important, while the weighted kappa (κ_w) yields a higher reliability when disagreements between raters are small compared with when they are large. In general, the following scale was used to rate kappa values: poor = 0.00–0.40, good = 0.41–0.70, very good = 0.71–0.80, and excellent = 0.81–1.00. The number of categories evaluated ranged from 3 (i.e. mild, moderate, and severe dysplasia) to 7 for lesion diagnosis. Lesions were then stratified by lesion characteristics, and kappa values were determined within each of these groups. For characteristics that were subject to interobserver variation, such as inflammation, only lesions for which there was agreement for that characteristic were used in determining kappa values. For lesions in which there was disagreement regarding lesion diagnosis,

Table 1 Selected characteristics of oral lesions

Characteristic	*Number of lesions (%)*
Site of biopsy	
Lips/labial mucosa	7 (8.1)
Floor of mouth	9 (10.3)
Tongue	29 (33.3)
Gingiva/hard palate	18 (20.7)
Vestibule/buccal mucosa	21 (24.1)
Other site	3 (3.5)
Biopsy type	
Punch	48 (55.2)
Wedge	39 (44.8)
Inflammation[a]	
Yes	13 (14.9)
No	51 (59.8)

[a]Only includes lesions for which both sets of pathologists agreed on presence or absence of inflammation.

the diagnoses of one central pathologist were chosen at random for the analysis. Kappa values and CI were determined with SAS Statistical Software (Cary, NC, USA).

Results

A total of 87 biopsies of lesions were performed on 81 subjects. The average participant age was 61.6 ± 11.8 years, and 61 (75.3%) were male. 60 (74.1%) of the participants were Caucasian, 16 (19.8%) African–American, 3 (3.7%) Asian, and 1 (1.2%) each Hispanic and Native American. Ten (12.4%) participants reported never having used tobacco, 44 (54.3%) were past users, and 27 (33.3%) of the participants reported current use of tobacco products. Ten (12.4%) participants had never used alcohol, 34 (42.0%) were past users, and 37 (45.6%) of the participants reported current alcohol drinking. Lesion characteristics are presented in Table 1.

The pathologic diagnoses made by the central and the local pathologists are jointly presented in Table 2. For the entire study population, the κ_w for lesion diagnosis was 0.59 (95% CI: 0.45, 0.72). When the pathologic diagnoses were collapsed into three categories of 'no abnormality/hyperkeratosis', 'mild, moderate, or severe dysplasia', and 'carcinoma *in situ*/carcinoma', the κ_w value increased to 0.70 (95% CI: 0.56, 0.84; Table 3). The κ_w value on the extent of inflammation was 0.37 (95% CI: 0.17, 0.56). Lesions demonstrating no inflammation (as agreed upon by all pathologists) had a κ_w for lesion diagnosis that was much higher ($\kappa_w = 0.67$ (95% CI: 0.53, 0.80)) than lesions that were inflamed ($\kappa_w = -0.10$ (95% CI: −0.27, 0.07); Table 3). Lesions exhibiting inflammation were graded no higher than mild dysplasia.

When stratified by intraoral site (Table 4), greatest agreement was seen for lesions located in the buccal mucosa/vestibule ($\kappa_w = 0.68$ (95% CI: 0.46, 0.91)) and tongue ($\kappa_w = 0.62$ (95% CI: 0.40, 0.84)). Least agreement was found for lip and labial mucosa lesions ($\kappa_w = -0.04$ (95% CI: −0.34, 0.27)). Punch biopsies ($\kappa_w = 0.67$ (95% CI: 0.54, 0.80); $n = 48$) had greater interobserver reliability than wedge biopsies ($\kappa_w = 0.38$ (95% CI: 0.12, 0.64); $n = 39$).

When the pathologic diagnoses of carcinoma *in situ* and carcinoma were combined into one category and were compared to lesions diagnosed as a less serious disease (i.e. bivariate comparison), the kappa value was poor (unweighted $\kappa = 0.39$ (95% CI: −0.20, 0.97)). Among dysplastic lesions, agreement as to the dysplasia thickness was poor ($\kappa_w = 0.10$ (95% CI: −0.09, 0.30)).

Kappa values for the presence and degree of inflammation were greatest for current alcohol drinkers ($\kappa_w = 0.50$ (95%

Table 3 Agreement on pathologic diagnoses, overall and by presence or absence of inflammation.

	κ_w *(95% CI)*
Pathologic diagnosis ($N = 87$)	0.59 (0.45, 0.72)
Pathologic diagnosis, three categories[a] ($N = 87$)	0.70 (0.56, 0.84)
Pathologic diagnosis, non-inflammed tissue ($N = 51$)	0.67 (0.53, 0.80)
Pathologic diagnosis, inflammed tissue ($N = 13$)	−0.10 (−0.27, 0.07)

[a]Three pathologic diagnostic categories: 'no abnormality/hyperplasia', 'mild, moderate, or severe dysplasia', 'carcinoma *in situ*/carcinoma'.

Table 4 Agreement on pathologic diagnoses, by lesion site[a]

Lesion site	κ_w *(95% CI)*
Tongue ($N = 29$)	0.62 (0.40, 0.84)
Lip and labial mucosa ($N = 7$)	−0.04 (−0.34, 0.27)
Gingiva and hard palate ($N = 18$)	0.42 (0.07, 0.76)
Buccal mucosa and vestibule ($N = 21$)	0.68 (0.46, 0.91)
Floor of mouth ($N = 9$)	0.49 (0.24, 0.74)

[a]Does not include three lesions, one on the soft palate, and two on the tonsillar pillar, listed as 'other site'.

Table 2 Distribution of pathologic diagnoses according to pathologist (central pathologists vs. local pathologists)

	Central pathologist						
Local pathologist	*No abnormality*	*Hyperplasia/ hyperkeratosis*	*Mild dysplasia*	*Moderate dysplasia*	*Severe dysplasia*	*Carcinoma in situ*	*Carcinoma*
No abnormality	20	16	3	3	1	0	1
Hyperplasia/hyperkeratosis	8	13	2	2	0	0	0
Mild dysplasia	0	0	1	1	0	0	0
Moderate dysplasia	0	0	1	2	1	0	1
Severe dysplasia	0	0	0	0	1	1	0
Carcinoma *in situ*	0	0	0	0	0	1	2
Carcinoma	0	0	0	1	0	0	5

Table 5 Agreement on presence and degree of inflammation, and pathologic diagnoses, by alcohol and smoking habits

	κ_w (95% CI)
Presence and degree of inflammation	
Non-drinkers ($N=10$)	−0.06 (−0.47, 0.35)
Past drinkers ($N=35$)	0.32 (0.08, 0.56)
Current drinkers ($N=42$)	0.50 (0.18, 0.82)
Non-smokers ($N=10$)	0.36 (−0.20, 0.91)
Past smokers ($N=48$)	0.47 (0.23, 0.71)
Current smokers ($N=29$)	0.19 (−0.19, 0.56)
Pathologic diagnoses	
Non-drinkers ($N=10$)	0.51 (0.14, 0.89)
Past drinkers ($N=35$)	0.65 (0.39, 0.91)
Current drinkers ($N=42$)	0.54 (0.37, 0.71)
Non-smokers ($N=10$)	0.71 (0.41, 1.00)
Past smokers ($N=48$)	0.63 (0.48, 0.79)
Current smokers ($N=29$)	0.42 (0.12, 0.71)

CI: 0.18, 0.82)), intermediate for past drinkers ($\kappa_w = 0.32$ (95% CI: 0.08, 0.56)), and very low for non-drinkers ($\kappa_w = -0.06$ (95% CI: −0.47, 0.35)). Agreement regarding the presence of inflammation was lowest for current smokers ($\kappa_w = 0.19$ (95% CI: −0.19, 0.56)) compared to past smokers ($\kappa_w = 0.47$ (95% CI: 0.23, 0.71)) and non-smokers ($\kappa_w = 0.36$ (95% CI: −0.20, 0.91); Table 5).

Agreement in pathologic diagnosis was very similar for current drinkers ($\kappa_w = 0.54$ (95% CI: 0.37, 0.71)), past drinkers ($\kappa_w = 0.65$ (95% CI: 0.39, 0.91)), and non-drinkers ($\kappa_w = 0.51$ (95% CI: 0.14, 0.89)). Kappas for pathologic diagnosis were lowest for current smokers ($\kappa_w = 0.42$ (95% CI: 0.12, 0.71)), intermediate for past smokers ($\kappa_w = 0.63$ (95% CI: 0.48, 0.79)), and highest for non-smokers ($\kappa_w = 0.71$ (95% CI: 0.41, 1.00); Table 5).

Discussion

Relatively few studies have attempted to determine the reliability of pathologic diagnosis for oral pre-cancerous and cancerous lesions. In 1985, Pindborg et al. (7) reported displaying nine photomicrographs of oral lesions at a scientific meeting of oral pathologists. Each of the 72 participants formulated diagnoses of the lesions, ranging from no dysplasia to frank carcinoma. The range of agreement for different histologic diagnoses was 1–78%. For example, a case of carcinoma *in situ* on the buccal mucosa had a 1% agreement, while a case of carcinoma *in situ* on the floor of mouth had a 78% agreement. Those data did not consider the seriousness of diagnostic disagreements, nor were chance agreements taken into account analytically. While not all participants were experienced oral pathologists, there was a great variability in the diagnoses. In another study, two general pathologists and two oral pathologists evaluated 100 consecutive specimens of oral leukoplakia lesions (8). Interobserver agreement between the four pathologists was evaluated. Resulting kappa values ranged from 0.27 to 0.45, regardless of whether the diagnosis was made by a general or oral pathologist. A third reliability study evaluated inter- and intraobserver agreement on 60 oral epithelial lesions among six board-certified oral pathologists (9). Kappa values ranged from 0.15 to 0.41 (poor) for interobserver agreement and ranged from 0.05 to 0.49 (poor) for intraobserver agreement. When expanded to 'within one histological step', the kappa values improved to 0.70–0.88 (excellent) for interobserver comparisons and 0.73–1.00 (excellent) for intraobserver comparisons. A fourth study evaluated the interobserver reliability among four pathologists reviewing 196 leukoplakias (10). Kappa values were determined for two pathologists at a time, and ranged from 0.17 to 0.33 with three categories of mild, moderate, and severe dysplasia. Kappa values were 0.21–0.34 when just two categories of 'favorable' (mild and moderate dysplasia) and 'poor' (severe dysplasia) were used.

The subjectivity of the diagnosis of pre-cancer is not limited to oral lesions. Studies of cervical epithelium, which share many characteristics with oral tissue, have shown variability in the diagnosis and grading of dysplastic lesions (12–14). The highest kappa value in these studies reached 0.69, but only after pathologists agreed on which morphological characteristics should be considered relevant for grading (14). In general, diagnostic tests that require subjective interpretation have yielded poor interobserver reliability.

Our study suggests that oral pathologists can achieve fair to good agreement in the diagnosis of clinically suspicious lesions. Our overall kappa was 0.59 (95% CI: 0.45, 0.72), which is higher than that obtained from previous studies (8–10). When the histologic diagnosis was simplified into three general categories of 'no abnormality/hyperplasia', 'mild/moderate/severe dysplasia', and 'carcinoma *in situ*/carcinoma', the interobserver reliability value improved ($\kappa_w = 0.70$ (95% CI: 0.56, 0.84)), which is contrary to findings from other studies (10). However, the kappa comparing diagnoses of 'carcinoma *in situ*/carcinoma' vs. less serious disease (i.e. bivariate comparison) was 0.39 (−0.20, 0.97). The potentially serious nature of the disease and its treatment mandates a much higher level of diagnostic reliability. Different diagnoses may result in unacceptable levels of over- and underdiagnosis of serious lesions.

Interobserver agreement for pathologic diagnosis was similar in non-drinkers, past drinkers and current drinkers. The kappa was slightly higher among non-smokers compared to past or current smokers, probably because of less tissue trauma in non-smokers.

Nonetheless, depending upon lesion characteristics, the interobserver reliability measures varied greatly. Interobserver agreement varied with lesion site. While it is reassuring to observe good agreement for buccal mucosa/vestibule and tongue lesions, the poor agreement seen in lip and labial mucosa lesions is of concern. It is possible that kappa values for buccal mucosa/vestibule and tongue lesions are high because lesions in these areas are common pre-cancerous/cancerous sites. In particular, within the tongue, there was evidence for greater agreement with lesions on the lateral surface, the common location for tongue lesions, compared to dorsal or ventral lesions (data not shown). Consequently, pathologists may be more familiar with diagnosing lesions from suspicious locations.

Our data also suggest that punch biopsies provide greater interobserver reliability than wedge biopsies. Our study used 3–4-mm punch biopsies, which supplied the pathologist with less tissue than wedges and therefore less of a chance

that the sample would contain heterogenous features that makes a diagnosis difficult. Punch biopsy technique may also result in more consistent sampling and positioning of tissue as punch biopsies were submitted on backing paper, which may allow greater ease of tissue sectioning for histopathologic review.

As assessment of dysplasia thickness has poor reliability in our study, we suggest that this criterion is not reliable in diagnosing grades of dysplasia.

Our study suggests that the presence of inflammation can modify reliability of oral lesion diagnoses. Inflammation may induce reactive atypia, may be associated with dysplastic changes in lesions, and/or may reduce a pathologist's ability to observe dysplastic changes. In this study, epithelial changes in inflamed lesions were graded only as no abnormality, hyperplasia/keratosis, or mild dysplasia. Moreover, interobserver reliability regarding the presence of inflammation was fair at best. Consequently, the severity of lesions exhibiting inflammation may be underreported. Finally, among smokers, the reliability of assessing the presence and extent of inflammation is lower compared to past and non-smokers, possibly because of chronic alterations of tissue that result from smoking.

Our study included 21 local pathologists from a variety of pathology specialties, representing those commonly seen in clinical practice. The three central pathologists included two head and neck cancer pathologists and one oral pathologist. None of the local or central pathologists were aware that their diagnoses were being compared to those made by other pathologists. Two of our 'gold standard' central laboratory pathologists evaluated lesions individually and, when their diagnoses were not in complete agreement, a third pathologist established a diagnosis, and all three diagnoses were reported. We believe that this protocol allowed for highly accurate diagnoses by the central pathology laboratory. One may argue that this spectrum of pathologists might not be representative of those who commonly diagnose suspicious oral lesions. However, previous studies that have evaluated agreement among pathology specialists have found no greater interobserver reliability than among general histopathologists (8, 9, 15).

Other studies have been discounted for having incorporated bias into the selection of pathologic specimens when evaluating the interobserver reliability of histologic specimens in general (16). Examples of sources of bias include non-random selection of slides or inclusion only of more difficult cases. In our study, we evaluated lesions as eligible and consenting patients presented in clinic during the time period of the study, so we did not integrate bias into our lesion selection.

One limitation of our study is the fact that we were only able to obtain 87 lesion specimens, which caused small sample sizes when subgroups were examined. While the large number of pathologists who participated in the study increased the generalizability of the results, it is also possible that the variety of experiences offered by these pathologists was another limitation to the study. Some pathologists may have had excellent interobserver agreement, but this was not seen because the kappa values generated were based on the combined interobserver reliability of all of the pathologists. Furthermore, the study may have been limited by the fact that the two sets of slides evaluated by the local and central pathologists were not always identical, but rather were based in some instances on adjacent microtome sections (steps). While the overall patterns of interobserver reliability are not likely to have been strongly affected by this aspect of our study, we might have observed somewhat higher kappa values on pathologic diagnosis and inflammation status had the paired pathology review for all lesions been based on the same slides.

The results from this study emphasize the need to reduce the variability in the diagnosis of oral pre-cancerous and cancerous lesions. One suggestion to improve diagnostic accuracy has been to combine borderline lesions into one category, such as has been suggested with the 'Bethesda System' for cervical neoplasia (16), as the level of intervention is frequently the same. Nevertheless, the future may rely on the application of molecular markers to aid in the diagnosis of pre-cancerous and cancerous lesions of the oral cavity. Progression from benign to malignant disease is a genetic process, later visualized at the cellular level (phenotypic change) and ultimately at the clinical level. Recent studies have used microsatellite and other genetic and DNA markers to identify molecular genetic profiles of malignant risk in pre-malignant oral lesions (10, 17–25), and such profiles show promise in accompanying histologic diagnoses in the future. Ultimately, the goal is to use molecular tools to assist in earlier identification of high risk lesions and to lead to more accurate histologic diagnoses.

References

1. American Cancer Society. *Cancer Facts & Figures* 2003. Atlanta: American Cancer Society Inc., 2003.
2. Mashburg A. Diagnosis of early oral and oropharyngeal squamous carcinoma: obstacles and their amelioration. *Oral Oncol* 2000; **26**: 253–5.
3. Silverman S, Gorsky M, Kaugars GE. Leukoplakia, dysplasia, and malignant transformation. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 1996; **82**: 117.
4. WHO Collaborating Centre for Oral Precancerous Lesions. Definition of leukoplakia and related lesions: an aid to studies on oral precancer. *J Oral Surg* 1978; **46** (4): 518–39.
5. Pindborg JJ, Reichart PA, Smith CJ, van der Waal J. *Histological Typing of Cancer and Precancer of the Oral Mucosa*, 2nd edn. Springer, New York: World Health Organization, 1997.
6. Warnakulasuriya S. Histological grading of oral epithelial dysplasia: revisited. *J Pathol* 2001; **194**: 294–7.
7. Pindborg JJ, Reibel J, Holmstrup P. Subjectivity in evaluating oral epithelial dysplasia, carcinoma *in situ*, and initial carcinoma. *J Oral Pathol* 1985; **14**: 698–708.
8. Karabulut A, Reibel J, Therkildesn MH, Praetorius F, Nielsen HW, Dabelsteen E. Observer variability in the histologic assessment of oral premalignant lesions. *J Oral Pathol Med* 1995; **24**: 198–200.
9. Abbey LM, Kaugars GE, Gunsolley JC, et al. Intraexaminer and interexaminer reliability in the diagnosis of oral epithelial dysplasia. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 1995; **80**: 188–91.
10. Sudbo J, Bryne M, Johannessen AC, Kildal W, Danielsen HE, Reith A. Comparison of histological grading and large-scale genomic status (DNA ploidy) as prognostic tools in oral dysplasia. *J Pathol* 2001; **194**: 303–10.

11. Armstrong BK, White E, Saracci R. *Principles of Exposure Measurement in Epidemiology.* Oxford: Oxford University Press, 1992, pp. 78–114.
12. Ringsted J, Amtrup F, Asklund C, et al. Reliability of histopathological diagnosis of squamous epithelial changes of the uterine cervix. *Acta Pathol Microbiol Scand Sect A* 1978; **86**: 273–8.
13. De vet HC, Knipschild PG, Schouten HJ, et al. Interobserver variation in histopathological grading of cervical dysplasia. *J Clin Epidemiol* 1990; **43**: 1395–8.
14. De vet HC, Knipschild PG, Schouten HJ, et al. Sources of interobserver variation in histopathological grading of cervical dysplasia. *J Clin Epidemiol* 1992; **45** (7): 785–90.
15. Bosman FT. Dysplasia classification: pathology in disgrace? *J Pathol* 2001; **194**: 143–4.
16. Henson DE. Studies on observer variation should the rules be changed? *Arch Pathol Laboratory Med* 1991; **115** (10): 991–2.
17. Sudbo J, Kildal W, Risberg B, Koppang HS, Danielsen HE, Reith A. DNA content as a prognostic marker in patients with oral leukoplakia. *New Engl J Med* 2001; **344** (17): 1270–8.
18. Rosin MP, Cheng X, Poh CF, et al. Use of allelic loss to predict malignant risk for low-grade oral epithelial dysplasia. *Clin Cancer Res* 2000; **6** (2): 357–62.
19. Mao L, Lee JS, Fan YH, et al. Frequent microsatellite alterations at chromosome 9p21 and 3p14 in oral premalignant lesions and their value in cancer risk assessment. *Nat Med* 1996; **2** (6): 682–5.
20. Partridge M, Pateromichelakis S, Phillips E, Emilion GG, A'Hern RP, Langdon JD. A case control study confirms that microsatellite assay can identify patients at risk of developing oral squamous cell carcinoma within a field of cancerization. *Cancer Res* 2000; **60** (14): 3893–8.
21. Lee JJ, Hong WK, Hittelman WN, et al. Predicting cancer development in oral leukoplakia: ten years of translational research. *Clin Cancer Res* 2000; **6** (5): 1702–10.
22. Mao L. Can molecular assessment improve classification of head and neck premalignancy? *Clin Cancer Res* 2000; **6**: 321–2.
23. Califano J, Westra WH, Meininger G, Corio R, Koch WM, Sidransky D. Genetic progression and clonal relationship of recurrent premalignant head and neck lesions. *Clin Cancer Res* 2000; **6** (2): 347–52.
24. Cairns P, Sidransky D. Molecular methods for the diagnosis of cancer. *Biochim Biophys Acta* 1999; **1423** (2): C11–8.
25. Ahrendt SA, Sidransky D. The potential of molecular screening. *Surg Oncol Clinics N Am* 1999; **8** (4): 641–56.

Acknowledgements

Data were obtained from Zila Pharmaceuticals Inc. from the Zila-ZP-44389–01 study. Dr Dena Fischer was supported by grant K16 DE00161 from the National Institute of Dental and Craniofacial Research.