Cytokeratin 19 Immunostain Reduces Variability in Grading Epithelial Dysplasia of the Non-Keratinized Upper Aerodigestive Tract Mucosa

Rima A. Safadi^{1,2,3} · Abdulaziz AlRomaizan^{2,3,4} · Rana S. Alshagroud⁵ · Darshan D. Divakar⁶ · Saeed Alshieban^{2,3,4}

Received: 13 February 2019 / Accepted: 22 April 2019 / Published online: 14 May 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Epithelial dysplasia (ED) grading systems are criticized for low reliability. The effects of diagnostic tests or stains on reducing examiner variability in grading ED of the upper aerodigestive tract have not been investigated. The study aim was to examine the effectiveness of cytokeratin 19 (K19) immunostain on enhancing inter and intraexaminer reliability of ED grading and to reiterate the relation of K19 positivity with epithelial keratinization. The study sample consisted of 122 paraffin blocks that fulfilled the inclusion criteria. Each paraffin block had three sections cut: one immunostained for K19 and two for hematoxylin and eosin stain (H&E). Each examiner graded the study sample in six rounds; three using H&E stain only and three using paired K19–H&E stains. The study examiners were three American-Board certified practicing oral pathologists. The results were analyzed using Krippendorff's alpha, ROC curve, Chi square test and binary logistic regression. Upon the use of paired K19–H&E stains the results showed that the intraexaminer reliability coefficients of grading were improved from 0.70, 0.69, 0.78 to 0.73, 0.88, 0.91 for examiners 1, 2, and 3 respectively. Reliability coefficients for inter-examiners improved from 0.55 to 0.73 (Krippendorff alpha). The accuracy of identifying the diseased cases (high-grade dysplasia) increased from 0.82 to 0.94 (ROC curve). Binary logistic regression revealed that K19 positivity is negatively associated with hyperkeratinization of surface epithelium ($P=0.001$). To conclude, for grading non-keratinized epithelial dysplastic lesions of the upper aerodigestive tract, paired K19–H&E stains proved to reduce inter and intra-examiner variability by highlighting the extension of dysplastic epithelial cells within epithelial thickness, thus identifying the involved epithelial third and assigning a more reliable and better reproducible grade.

Keywords Dysplasia · Grade · K19 · Adjunctive · Reliability · Variability

✉ Rima A. Safadi
safadir@ksau-hs.edu.sa; rsafadi@just.edu.jo;
rimaasafadi@gmail.com

Abdulaziz AlRomaizan
Alromaizanab@ngha.med.sa; romaizana@ksau-hs.edu.sa

Rana S. Alshagroud
ralshagroud@ksu.edu.sa

Darshan D. Divakar
darshandevang@gmail.com

Saeed Alshieban
ShiebanS@ngha.med.sa

² King Saud bin Abdul Aziz University for Health Sciences,
P.O Box: 22490, Riyadh 11426, Saudi Arabia

³ King Abdullah International Medical Research Center,
Riyadh, Saudi Arabia

⁴ Department of Pathology and Laboratory Medicine, King
Abdulaziz Medical City, Ministry of National Guard-Health
Affairs, P.O Box: 22490, Riyadh 11426, Saudi Arabia

⁵ College of Dentistry, King Saud University, Riyadh 11451,
Saudi Arabia

⁶ Dental Health Department, College of Applied Medical
Science, King Saud University, Riyadh 11451, Saudi Arabia

¹ Faculty of Dentistry, Jordan University of Science
and Technology, P.O Box: 3030, Irbid 22110, Jordan

Introduction

Epithelial dysplasia (ED) is a spectrum of abnormal epithelial maturation and cellular aberrations that may or may not precede invasive carcinoma [1]. The severity of ED does not only relate to the likelihood of malignant transformation but also the management protocols; thus great emphasis was laid upon the accuracy and reliability of its grading [2–6]. Unfortunately, it is still universally noticed that oral ED suffers from both intra and inter-examiner variability despite defining the architectural and cytologic criteria of diagnosis in the 2017 World Health Organization (WHO) oral epithelial dysplasia (OED) grading system [7].

Over the years, different grading systems have been proposed in an attempt to achieve better objectivity and reliability of ED grading [5, 8–17]. Some preferred the use of a binary grading system similar to that of laryngeal precursor lesions, a two-tiered system with low-grade and high-grade dysplasia. In order to achieve better grading reliability, the role of the binary and other grading systems have been investigated and in some studies compared with the WHO system [11, 18–22]. Despite the improvement in grading reliability using some of these systems, the prognostic impact is still not assured [19] and the three-tiered WHO OED grading system remains the most widely accepted system until the clinical, prognostic and management validity of the other systems is attained.

In this study, we aimed at investigating a diagnostically helpful and readily available, simple tool that would assist the grading pathologist in assigning a specific ED grade using 2017 WHO OED grading system. The goal of identifying such a tool is not only to achieve better inter and intraexaminer reliability coefficients but also to arrive at assigning the most accurate grade.

Keratins are intermediate cytoplasmic filaments that have a tissue-specific expression [23]. Cytokeratin 19 (K19) is normally expressed in the basal cell layer or stem cells. Its expression in suprabasal locations in oral epithelium has been correlated with the presence and the grade of ED [24–28]. However, no previous studies have investigated the impact of K19 expression on reducing the variability of grading ED. K19 immunorexpression has been shown to be undetectable in keratinized mucosa whether normal or benign hyperplastic. When keratinized epithelium is dysplastic, K19 was expressed only in the basal cell layer of severe ED [29, 30].

This study aimed to explore, for the first time, the role of pairing K19 immunostain with H&E stain in reducing inter and intra-examiner variability taking into consideration the relation of K19 positivity with keratinization status.

Materials and Methods

Case Selection

The archives of the Department of Pathology at King Abdulaziz Medical City, Ministry of the National Guard, Riyadh, KSA were searched for specimens histopathologically diagnosed with ED, for 2013–2018. The upper aerodigestive system was targeted. Biopsy sites included the lips, mouth, tongue, nose, throat, vocal cords, and upper part of the esophagus.

Retrieved reports were reviewed for diagnosis, histopathologic features, comments, age, gender, anatomic location and clinical presentation. The identified cases had their slides and Paraffin blocks retrieved and evaluated. Search keywords and inclusion criteria were primarily as follows:

- Oral, pharyngeal, upper esophageal and laryngeal mucosal biopsies diagnosed clinically with leukoplakia, erythroplakia, or erythroleukoplakia
- Oral, laryngeal or esophageal mucosal biopsy diagnosed histopathologically or has in its report the words: keratosis, hyperkeratosis, acanthosis, epithelial hyperplasia, orthokeratosis, parakeratosis, hyperparakeratosis, mild, moderate or severe dysplasia.
- Enough tissue in paraffin blocks
- Enough surface epithelium for interpretation.

Exclusion criteria included: severe inflammation, adjacent ulceration, basal cell layer degeneration, tangential cut, excessive melanin pigmentation, and poor quality interfering with the diagnosis.

Immunohistochemistry

Eight-micrometer-thick sections were cut from formalin-fixed paraffin embedded tissue blocks and mounted on vectabond-coated and charged glass micro slides, (Leica Bond plus slides REF S21.2113. A. Leica, Germany). Two additional tissue sections were cut for H&E staining following standard protocols. Immunohistochemistry was performed by the standard protocol, using horseradish peroxidase (HRP)–diaminobenzidine (DAB) and counterstained using Mayer's hematoxylin. All immunostaining procedures were done using autostainer (BenchMark XT system, Roche, AZ, USA). The primary antibody was ready to use CK 19 (A53-BIA2.26, Cell Marque, Millipore Sigma, CA, USA). One tissue section in each immunohistochemical run was incubated with non-specific mouse IgG as negative control. Positive control tissue was K19 positive moderately differentiated squamous cell carcinoma of the lateral border of the tongue.

Microscopic Evaluation

All stained sections were evaluated for stain quality after each run. The epithelium was considered positive for K19 if surface squamous epithelial cells showed brown cytoplasmic staining. K19 staining patterns were described as basal and suprabasal.

Study Examiners and Slide Distribution Process

All of the three examiners were Oral and Maxillofacial Pathologists who received their graduate training in the United States and are all Diplomates of the American Board of Oral and Maxillofacial Pathology. The three Oral Pathologists are University Assistant/Associate Professors and are Surgical Oral Pathology practitioners.

The 366 stained sections were distributed by a research assistant in a total of six boxes that were grouped into three sets. In each set: the first box contained paired K19–H&E stains of the same paraffin blocks while the second had H&E stained sections of paraffin blocks other than those of the first box. Before starting the process of grading, the three examiners discussed ED grading criteria, K19 biology and labeling patterns and other details related to the study design. One ED grading round was completed when each examiner finished the three sets of boxes. Each examiner performed three rounds during six months. Completed grading forms were collected together with their respective boxes after each run not to leave any record of the assigned grades. After finishing the three rounds, all records were sent for statistical analysis. The first examiner (RS) had an additional task of identifying the characteristics of the surface epithelium in the examined field, in terms of epithelial thickness and keratinization status. This was done in two separate runs. Slides with two conflicting results were looked at for a third run.

Epithelial Dysplasia Grading

The grading system followed in this study was the 2017 WHO OED grading system which is similar to that for other mucosal sites in the upper aerodigestive tract [7]. The architectural criteria of ED in this system were: irregular epithelial stratification, loss of polarity of basal cells, drop-shaped rete ridges, increased number of mitotic figures, abnormally superficial mitoses, loss of epithelial cell cohesion, keratin pearl formation within rete ridges and premature keratinization in single cells (dyskeratosis). The cytological features were: abnormal variation in nuclear or cellular size and shape, increased nuclear-cytoplasmic

ratio, hyperchromatism, atypical mitotic figures and increased number and size of nucleoli.

Based on the extent of cytological and architectural alterations, ED was classified into three grades: mild dysplasia (basal and suprabasal involvement up to the lower third) moderate dysplasia (up to the midportion of the spinous cell layer), and severe dysplasia (involving the upper third up to full thickness). Severe ED and carcinoma in situ were considered as “severe epithelial dysplasia”. Hyperplasia without cellular or architectural alterations was not considered in the grading scheme. In this study, the latter was assigned with “no dysplasia.” Special types of dysplasia like HPV associated dysplasia were not specifically encountered. For ED grading utilizing paired slides of H&E and K19 immunostain, the examiner assessed H&E stained sections first followed by K19 immunostain. In case of K19 positivity (noted primarily in non-keratinized dysplastic epithelium), the examiner utilized this advantage to better appreciate the presence and superficial extension of dysplastic epithelial cells, thus better assessment of the involved epithelial third. The most superficial extension of K19 positive dysplastic cells was referred to in assigning the grade.

Since grading ED is known for its inter and intraexaminer variability, and because most of the study cases were originally graded by general pathologists, establishing the gold standard of ED grade for this study was decided to be determined by taking the most frequent grade out of the nine readings of the three study raters using only H&E stained sections. This reference grade was used for the frequency distribution of cases included in the study. Likewise, for K19 positive cases, the gold standard was determined by taking the most common grade out of the nine readings of the three examiners using adjunctive K19 immunostain. The K19 gold standard served as the most accurate grade assigned.

Statistical Analysis

Data were entered into the Statistical Package for Social Science (SPSS) (version 20, SPSS Inc., Chicago IL). The assessment of inter-examiner and intraexaminer variability was performed using Krippendorff alpha (Kalpha) [31]. Kalpha is a measure of observed disagreement relative to the disagreement expected by chance [32]. It is the alternative to the known Cohen kappa coefficient when more than two raters are involved. Receiver operating characteristic (ROC) curve was used to assess the accuracy of identifying high grade dysplasia (labeled as diseased cases in ROC curve). ROC curve test was done twice: based on H&E stain alone and paired K19–H&E stains. This statistical test mandate having only binary grades thus for this purpose, the study sample was grouped into low grade and high grade dysplasia following the terminology used in binary grading systems [33] where high grade consisted of moderate and severe

grades while low grade consisted of mild grade in addition to the cases with variable impression of being non dysplastic vs mildly dysplastic. Chi square test and binary logistic regression test were used for K19 expression and staining patterns correlation with the other parameters.

Results

Study Sample

The specimens included in the study were for patients whose age ranged from 21 to 96 years. Specimens of male patients comprised 76% of the study sample. The anatomic location distribution of the biopsied lesions was as follows: oral cavity 49 (40.2%), larynx 47 (38.5%), upper esophagus 14 (11.5%) and oropharynx 12 (9.8%). The most prevalent specific locations were: vocal cords 34 (27.9%) and lateral border of the tongue 22 (18%) followed by esophagus 14 (11.5%). The reported clinical features of the biopsied lesions included: white lesion 32 (26.2%), ulcer 29 (23.8%), mass 28 (22.9%), hoarseness of voice 25 (20.5%) and dysphagia 8 (6.6%).

Gold Standard Diagnoses

Based on H&E staining alone, the gold standard grade distribution of study sample (122) was: “no dysplasia” 5 (4.1%), mild dysplasia 31 (25.41%), moderate dysplasia 48 (39.34%) and severe dysplasia 38 (31.14%). For the K19 positive cases, the gold standard grade distribution based on H&E stain alone was: “no dysplasia,” 3 (6.3%) mild dysplasia 10 (21.3%), moderate dysplasia 17 (36.2) and severe dysplasia 17 (36.2). When the grades were grouped as low grade and high grade for ROC curve analysis, the frequency was: 13 (27.7%) and 34 (72.3%) respectively.

K19 Positivity Associations

There was a statistically significant association between K19 pattern of positivity and ED grade (Chi square, $P=0.02$). K19 positive expression showed significant statistical association with the biopsy location (Chi square $P=0.023$).

The upper esophagus, pharynx, and floor of the mouth showed the highest percentage of K19 positive sections within each location which were 71.4%, 57.1 and 57.1% respectively. All gingival cases (7/7) were K19 negative.

There was a statistically significant association of K19 positivity with non-keratinization and epithelial atrophy status (Chi square, $P<0.001$ for both). Out of the 122 section, sixty-eight showed no hyper keratinization, of which 71.1% were K19 positive while 78% of hyperkeratotic lesions were K19 negative. Forty-two sections showed atrophic

epithelium of which 59.5% were K19 positive, while 75.3% of the hyperplastic epithelium was K19 negative. The desired role of K19 positivity was to highlight the extension of dysplastic epithelial cells within epithelial thickness in order to support the grading process. Figure 1 presents three examples K19 positive, non-keratinized dysplastic epithelium. Figure 1a presents mild ED where K19 positive cells were limited to the lower third of epithelial thickness, supporting mild grade. Another example was of a moderate ED (Fig. 1b) where scattered K19 positive dysplastic cells were noted to reach the upper part of the epithelial middle third. An example of severe ED is presented in Fig. 1c where the positive K19 dysplastic cells extended to the upper epithelial third, being in support of severe ED grade. When the dysplastic epithelium was hyperkeratinized, K19 expression was most of the time negative (78% of the cases) (Fig. 2). In the hyperkeratinized dysplastic epithelium that was K19 positive (22% of the cases), the extension of K19 positive cells was limited primarily to basal and parabasal cells (Fig. 3a). Figure 3b presents dysplastic epithelial tissues with abrupt transition from positive to negative K19 expression in the same tissue, illustrating the effect of hyperkeratinization on K19 expression.

Bivariate logistic regression analysis model was performed to identify the factors that will retain their significant association with K19 positivity. The variables entered in the model were anatomic location, epithelial thickness, epithelial keratinization status, ED grade. The factors that maintained their significant statistical relation with K19 positivity after the regression analysis were only: anatomic location and surface epithelium non-keratinization ($P=0.01$ and 0.001, respectively).

Reliability of Epithelial Dysplasia Grading

For grading ED using H&E stains alone, the examiners #1, 2, 3 had intra-examiner Kappa reliability coefficients of 0.70, 0.69, and 0.78 respectively. After grading the same sections with paired K19–H&E stains, the intra-rater reliability coefficients were: 0.73, 0.88, 0.91 (Table 1). The inter-rater reliability coefficient of ED grading using H&E only and with paired K19–H&E stains was 0.55 and 0.73 respectively (Krippendorff alpha reliability estimate, Table 1).

Study sections with K19 positive results had three more runs of ED grading to evaluate the effect of K19 positivity on the ED grading. The total number of recorded grades after three rounds was 141 (47 K19 positive sections $\times$ 3 rounds). The number of revised grades (upgraded or downgraded) upon the use of paired K19–H&E immunostain was 47 (33.3%) (Table 2). Thirty-four grades were raised, and 16 were lowered. The second examiner had the highest number of upgraded cases among the three examiners. These

Fig. 1 Positive K19 immunostain of three examples of non-keratinized dysplastic epithelium. **a, b** H&E stained section and K19 immunostain of a severe epithelial dysplasia case. Note positive K19 dysplastic cells that extended to the upper epithelial third, being in support of severe epithelial dysplasia grade (laryngeal lesion, original magnification $\times 10$). **c, d** H&E stained section and K19 immunostain of a moderate epithelial dysplasia case where scattered K19 positive

dysplastic cells noted to reach the upper part of the epithelial middle third (arrow heads) (pharyngeal lesion, original magnification $\times 20$). **e** K19 immunostain of a mild epithelial dysplasia case. Note K19 positivity localized to the basal cell layer supporting the diagnosis of a mild epithelial dysplasia grade (floor of the mouth lesion, original magnification $\times 20$)

Fig. 2 K19 negative immunostain in hyperkeratinized dysplastic epithelium (buccal mucosal lesion, original magnification $\times 10$)

Fig. 3 K19 immunostain in dysplastic epithelium illustrating the effect of hyperkeratinization on K19 expression. **a** K19 positive cells limited primarily to basal and parabasal cells in severe epithelial dysplasia of hyperkeratotic epithelium (floor of the mouth lesion, origi-

nal magnification $\times 20$). **b** Dysplastic epithelial tissues with abrupt transition from positive to negative K19 expression in accordance with changes in surface hyperkeratinization (vocal cord lesion, original magnification $\times 10$)

Table 1 Krippendorff's alpha reliability estimate of intra and inter examiner reliability in grading epithelial dysplasia using H&E stain and conjunctive K19 immunostain

Examiner number and stain utilized	Krippendorff's alpha reliability coefficient	95% confidence interval	
		Lower bound	Upper bound
Examiner 1 H&E	0.70	0.59	0.76
Examiner 1 K19	0.83	0.77	0.89
Examiner 2 H&E	0.69	0.59	0.76
Examiner 2 K19	0.88	0.82	0.94
Examiner 3 H&E	0.78	0.72	0.83
Examiner 3 K19	0.91	0.86	0.95
Inter examiner H&E	0.55	0.49	0.61
Inter examiner K19	0.73	0.68	0.78

changes have resulted in improved reliability coefficient as noted in Table 1.

Grading Accuracy

The area under the curve (ROC curve) was used to determine how accurate was the identification of the diseased cases

(high-grade dysplasia) using H&E stain compared to that of paired K19–H&E stains. The average area under the curve for all examiners using H&E stain alone was 0.819. It was raised to 0.941 upon the adjunctive use of K19 immunostain.

Discussion

In this paper, we aimed at testing the impact of pairing K19 immunostain with H&E stain on enhancing the reliability of ED grading using the WHO (2017) system. The WHO ED grading system for oral cavity is similar to that for other mucosal sites in the upper aerodigestive tract. For this study, the sample included ED of the upper aerodigestive epithelial lining. Only the upper part of the esophagus was included since it has a squamous epithelial lining. The originally assigned grade of each case was not considered for the purpose of this study. Assigned ED grades in this study were based on 2017 WHO OED grading system for all study cases regardless of its specific location within the upper aerodigestive lining epithelium.

Although several studies have revealed variability in ED grading [19, 22, 34, 35], only a few studies have investigated supportive grading tests [36] and no studies have tested the actual effect of utilizing such tests or stains on examiners

Table 2 Epithelial dysplasia grades reviewed upon conjunctive use of K19 immunostain

Examiner	No to mild	Mild to no	Mild to moderate	Moderate to mild	Moderate to severe	Severe to moderate	Total number (%)
Examiner 1	2	0	3	6	4	2	17
Examiner 2	6	0	7	0	5	0	18
Examiner 3	2	1	1	2	1	5	12
Total number	10	1	11	8	10	7	47/141 (33.3%)

variability of grading epithelium of the upper aerodigestive tract.

K19 immunoexpression was selected to be the adjunctive stain in grading process because its expression has shown a significant correlation with the extension of dysplastic cells within epithelial thickness [25, 28, 37]. K19 in the suprabasal region is accompanied by a delay of involucrin expression indicating a delay in commitment to terminal differentiation thus supporting that suprabasal K19 expression indicates increased premalignant potential [29]. Based on this, we hypothesized that K19 immunostain would be a valid tool to identify potentially malignant cells at the same time would improve reliability by making the dysplastic cells more visible. Adjunctive K19 immunostain resulted in significant improvement of both intra-examiner and inter-examiner reliability coefficients, which was tested using Krippendorff alpha (Kalpha). Intraexaminer reliability showed 13%, 19% and 13% reliability improvement for the three examiners (0.70, 0.69, 0.78 raised up to 0.83, 0.88, 0.91 respectively). There was 18% improvement in the inter-examiner reliability (0.55 raised to 0.73).

Many methodologists recommend a minimum α coefficient between 0.65 and 0.8 for a reliable test [38–41]. It has been suggested that $\alpha \geq 0.667$ is the lowest conceivable limit [42]. Krippendorff considered Kalpha of > 0.8 as “reasonable to rely on”. He described $0.67 < \alpha < 0.8$ as “allowing for tentative conclusions” while $\alpha < 0.67$ was described as “non- acceptable”. Landis and Koch, described reliability of Kalpha values of 0 to 0.2 as “slight”, 0.21 to 0.40 as “fair”, 0.41 to 0.60 as “moderate”, 0.61 to 0.80 as “substantial”, and 0.81 to 1 as “near perfect”. Landis and Koch described these ratings as “useful benchmarks” [43].

Based on Krippendorff and Landis and Koch, the intraexaminer reliability scores of H&E based grading can be accordingly described as “allowing for tentative conclusion” with moderate reliability, while upon the adjunctive K19 immunostain use, the reliability coefficients can be interpreted as “reasonable to rely on” with “near perfect” reliability. For the inter-examiner reliability, the reading has enhanced from “not- acceptable” to “substantial.”

The examiners related the improvement in reliability to the role of K19 immunostain in highlighting the dysplastic cells and the extent of irregular epithelial stratification, the latter being one of the features with the highest examiners disagreement [23, 44]. Inter and intraexaminer reliability consequently were improved due to more confidence in assigning a grade.

In previous studies, inter and intraexaminer variability of ED grading was related to several factors including the use of different epithelial grading systems, training of the grading pathologist and poor agreement on architectural dysplasia criteria [34, 45, 46]. To control for the effect of training of pathologists that has been linked to inter-examiner

variability the examiners in our study were all diplomats of the American Board of Oral and Maxillofacial Pathology has used the same grading system (WHO 2017) and same grading criteria. Grader education might affect the selection of the grading system to be used; hence it would contribute to increased variability [47]. The second examiner in our study had assigned overall lower grades using H&E only which were revised upon the use of adjunctive K19 immunostain resulting in better inter-examiner reliability. The second rater has been more regularly attending the sign out of general pathology cases where the cervical intraepithelial neoplasia is being graded. The WHO system has been shown to give a higher grade compared to the uterine cervix intraepithelial neoplasia grading system [34]. This variation could be attributed to the difference in keratinization between these two sites [48, 49], or the difference in etiologic factors like a persistent infection of high-risk Human Papilloma Virus in the uterine cervix [50].

Previous studies found that the reliability of grading moderate dysplasia was described as “poor” which was attributed to the lower specific correlation of cytological features with the moderate grade [34]. The correct identification of moderate epithelial grade is clinically significant. In a study by Kujan et al. using 2005 WHO criteria for diagnosing ED, it has been shown that 16 of 30 moderate ED showed malignant transformation at follow up [11]. Our study has shown that K19 immunostain adjunctive use resulted in changing the moderate grade the most among other grades leading to the better achievement of higher reliability coefficients. On the other hand, the “severe” grade showed the lowest variability among graders and the least grade being changed. Thus it is inferred that adjunctive K19 immunostain used will allow to better reliability in grading cases of moderate ED grade.

Although all ED grades carry the risk for malignant transformation [6], until this point of time, higher grades of ED remain the most critical risk factor for SCC development [18, 48, 51–54]. As a result, the accuracy of ED grading is desired. Grading based on adjunctive K19 immunostain rather than H&E stain only was utilized to determine the more accurate/gold standard diagnoses of the study sections as K19 provided better visualization of dysplastic cells in the positive tissue sections. Since a reliable test does not mean a valid/accurate test, the accuracy of adjunctive K19 immunostain impact on the three examiners grades was tested against the gold standard diagnoses. For this purpose, the area under the curve/ROC curve statistics was used. This test mandate having dichotomous variables, so the binary system principle of low and high-grade dysplasia was followed. “No dysplasia” and mild dysplasia were grouped under “low-grade dysplasia” and moderate and severe were under “high-grade dysplasia” category. The average area under the curve was raised from 0.819 to 0.941 upon the adjunctive

use of K19 immunostain, indicating increased accuracy in identifying the high-risk cases.

Our study showed that K19 labeling of surface epithelium was significantly correlated with the tissue site and epithelial keratinization status. Overall, upper esophagus, pharynx, and floor of the mouth showed the highest percentages of K19 positivity while all gingiva cases were K19 negative. This can be explained by our finding of the significant association of K19 positivity with low keratinization status of surface epithelium. Gingiva is relatively more keratinized tissue hence its 100% K19 negativity. It has been shown in previous studies that K19 expression was not detectable in keratinized mucosa whether normal or benign hyperplastic, however, if keratinized epithelium was dysplastic K19 expression was observed only in severe dysplasia grade and was localized to the basal cell layer [29, 30]. The latter finding was also noted in our study (Fig. 3a). A possible suggested explanation is that the early differentiation of basal epithelial cells in hyper keratinized epithelium could be taking place, which is in turn negatively associated with K19 expression [37]. This finding indicates that one of the limitations of the adjunctive K19 immunostain use in grading ED is its negative expression in keratinized mucosa. This biologic characteristic of K19 and its inverse relation to the keratinization status of the dysplastic epithelium should be considered while interpreting K19 sensitivity and specificity in detecting epithelial dysplastic changes. This feature of K19 expression would help understanding the results of some studies that did not consider the keratinizing status of dysplastic epithelial tissue and concluded that K19 immunostain was a marker of low sensitivity in detecting premalignant lesions [55]. With that said, K19 immunostain would not be a helpful tool to be investigated in the lesions of “keratosis of unknown significance,” the concept introduced by Villa and Woo [6].

Conclusions and Future Directions

Pairing K19 immunostain with H&E stain proved to increase intra and inter-examiner reliability and accuracy in grading ED of the non-keratinized upper aerodigestive tract mucosa. K19 immunostain adjunctive use has improved reliability and consistency in grading mild and moderate dysplasia grades. Paired K19–H&E stains is advised for grading ED of non-keratinized epithelium and recommended for the specific locations of upper esophagus, pharynx, and floor of the mouth biopsies. It is recommended to investigate the role of K19 immunostain in improving grading reliability using the binary and other grading systems. It is also suggested to investigate the effect of K19 immunostain on ED grading reliability and accuracy between expert and non-expert oral pathologists, oral and general pathologists.

Acknowledgements The research is funded by King Abdullah International Medical Research Center, King Saud Bin Abdulaziz University for Health Sciences, Ministry of National Guard, Riyadh, Saudi Arabia. Grant Number: RC 17/100. Special thanks to Sarah A. AlJundi, Medical Technologist, Department of Pathology and Laboratory Medicine at King Abdulaziz Medical City, Ministry of National Guard Health Affairs, Riyadh, KSA for her valuable technical support.

Compliance with Ethical Standards

Conflict of interest The authors declare no conflicts of interest of any type.

Ethical Approval The Research was approved by the IRB committee of King Abdullah International Medical Research Center; reference number IRBC/1154/17.

References

1. Harrison LB, Sessions RB, Hong WK. Head and neck cancer: a multidisciplinary approach. Philadelphia: Lippincott Williams & Wilkins; 2009. p. 24.
2. van Hulst AM, Kroon W, van der Linden ES, Nagtzaam L, Ottenhof SR, Wegner I, et al. Grade of dysplasia and malignant transformation in adults with premalignant laryngeal lesions. *Head Neck*. 2016;38(S1):E2284–90.
3. Reibel J. Prognosis of oral pre-malignant lesions: significance of clinical, histopathological, and molecular biological characteristics. *Crit Rev Oral Biol Med*. 2003;14(1):47–62.
4. Cowan C, Gregg T, Napier S, McKenna S, Kee F. Potentially malignant oral lesions in Northern Ireland: a 20-year population-based perspective of malignant transformation. *Oral Dis*. 2001;7(1):18–24.
5. Sadiq H, Gupta P, Singh N, Thakar S, Prabhakar I, Thakral J. Various grading systems of the oral epithelial dysplasia: a review. *Int J Adv Health Sci*. 2015;1:20–6.
6. Villa A, Woo SB. Leukoplakia—a diagnostic and management algorithm. *J Oral Maxillofac Surg*. 2017;75(4):723–34.
7. Reibel JGN, Hille J. Oral potentially malignant disorders and oral epithelial dysplasia. World Health Organization classification of tumors: head and neck tumors. Lyon: IARC Press; 2017. p. 435.
8. Sharma N, Hosmani JV, Tiwari V. Epithelial dysplasia: different grading system and its applications. *J Int Oral Health*. 2010;2(1):1–8.
9. Fleskens S, Slootweg P. Grading systems in head and neck dysplasia: their prognostic value, weaknesses and utility. *Head Neck Oncol*. 2009;1(1):11.
10. Speight PM. Update on oral epithelial dysplasia and progression to cancer. *Head Neck Pathol*. 2007;1(1):61–6.
11. Kujan O, Oliver RJ, Khattab A, Roberts SA, Thakker N, Sloan P. Evaluation of a new binary system of grading oral epithelial dysplasia for prediction of malignant transformation. *Oral Oncol*. 2006;42(10):987–93.
12. Mahajan MC, Hazarey V. An assessment of oral epithelial dysplasia using criteria of Smith & Pindborg Grading System' & Ljubljana Grading System' in oral precancerous lesions. *J Oral Maxillofac Pathol*. 2004;8(2):73.
13. Scully C, Sudbø J, Speight P. Progress in determining the malignant potential of oral lesions. *J Oral Pathol Med*. 2003;32(5):251–6.

14. Gonzalez-Moles MA. Comment on: Küffer and Lombardi “Premalignant lesions of the oral mucosa. A discussion about the place of intraepithelial neoplasia”. *Oral Oncol.* 2002;38(8):809–10.
15. Kuffer J, Lombardi S. Reconsideration oral risk lesions. *Oral Dis.* 2002;38:302–7.
16. Manchanda A, Shetty D-C. Reproducibility of grading systems in oral epithelial dysplasia. *Medicina oral, Patologia oral y Cirugia bucal.* 2012;17(6):e935.
17. Gale NPB. Epithelial precursor lesions. In: Barnes LEJ, Reichart P, Sindransky D, editors. World Health Organization classification of tumours pathology and genetics of head and neck tumours. Lyon: IARC; 2005. p. 177–9.
18. Nankivell P, Williams H, Matthews P, Suortamo S, Snead D, McConkey C, et al. The binary oral dysplasia grading system: validity testing and suggested improvement. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2013;115(1):87–94.
19. Gnepp D. Diagnostic surgical pathology of the head and neck. 2nd ed. Saunders: Elsevier Health Sciences; 2009. p. 647.
20. Gale N, Zidar N, Poljak M, Cardesa A. Current views and perspectives on classification of squamous intraepithelial lesions of the head and neck. *Head Neck Pathol.* 2014;8(1):16–23.
21. Izumo T. Oral premalignant lesions: from the pathological viewpoint. *Int J Clin Oncol.* 2011;16(1):15–26.
22. Kujan O, Khattab A, Oliver RJ, Roberts SA, Thakker N, Sloan P. Why oral histopathology suffers inter-observer variability on grading oral epithelial dysplasia: an attempt to understand the sources of variation. *Oral Oncol.* 2007;43(3):224–31.
23. Trask DK, Band V, Zajchowski DA, Yaswen P, Suh T, Sager R. Keratins as markers that distinguish normal and tumor-derived mammary epithelial cells. *Proc Natl Acad Sci.* 1990;87(6):2319–23.
24. Safadi RA, Musleh AS, Al-Khateeb TH, Hamasha AA-H. Analysis of immunohistochemical expression of k19 in oral epithelial dysplasia and oral squamous cell carcinoma using color deconvolution-image analysis method. *Head Neck Pathol.* 2010;4(4):282–9.
25. Yoshida K, Sato K, Tonogi M, Tanaka Y, Yamane G-Y, Katakura AC. Expression of cytokeratin 14 and 19 in process of oral carcinogenesis. *Bull Tokyo Dent Coll.* 2015;56(2):105–11.
26. Kobayashi T, Maruyama S, Cheng J, Ida-Yonemochi H, Yagi M, Takagi R, et al. Histopathological varieties of oral carcinoma in situ: diagnosis aided by immunohistochemistry dealing with the second basal cell layer as the proliferating center of oral mucosal epithelia. *Pathol Int.* 2010;60(3):156–66.
27. Su L, Morgan PR, Lane EB. Keratin 14 and 19 expression in normal, dysplastic and malignant oral epithelia. A study using in situ hybridization and immunohistochemistry. *J Oral Pathol Med.* 1996;25(6):293–301.
28. Lindberg K, Rheinwald JG. Suprabasal 40 kd keratin (K19) expression as an immunohistologic marker of premalignancy in oral epithelium. *Am J Pathol.* 1989;134(1):89.
29. Jacques CM, Pereira AL, Maia V, Cuzzi T, Ramos-e-Silva M. Expression of cytokeratins 10, 13, 14 and 19 in oral lichen planus. *J Oral Sci.* 2009;51(3):355–65.
30. Antoine J-Y, Villaneau J, Lefeuvre A, editors. Weighted Krippendorff’s alpha is a more reliable metrics for multi-coders ordinal annotations: experimental studies on emotion, opinion and coreference annotation. In: EACL 2014; 2014.
31. Krippendorff K. Reliability in content analysis. *Hum Commun Res.* 2004;30(3):411–33.
32. Gale N, Blagus R, El-Mofty SK, Helliwell T, Prasad ML, Sandison A, et al. Evaluation of a new grading system for laryngeal squamous intraepithelial lesions—a proposed unified classification. *Histopathology.* 2014;65(4):456–64.
33. Tilakaratne W, Sherriff M, Morgan P, Odell E. Grading oral epithelial dysplasia: analysis of individual features. *J Oral Pathol Med.* 2011;40(7):533–40.
34. Fleskens SA, Bergshoeff VE, Voogd AC, van Velthuysen M-LF, Bot FJ, Speel E-JM, et al. Interobserver variability of laryngeal mucosal premalignant lesions: a histopathological evaluation. *Mod Pathol.* 2011;24(7):892.
35. Mallia RJ, Subhash N, Mathews A, Kumar R, Thomas SS, Sebastian P, et al. Clinical grading of oral mucosa by curve-fitting of corrected autofluorescence using diffuse reflectance spectra. *Head Neck.* 2010;32(6):763–79.
36. Mittal KR, Demopoulos RI, Goswami S. Patterns of keratin 19 expression in normal, metaplastic, condylomatous, atrophic, dysplastic, and malignant cervical squamous epithelium. *Am J Clin Pathol.* 1992;98(4):419–23.
37. Carletta J. Assessing agreement on classification tasks: the kappa statistic. *Comput Linguist.* 1996;22(2):249–54.
38. Bland JM, Altman DG. Statistics notes: Cronbach’s alpha. *BMJ.* 1997;314(7080):572.
39. Tavakol M, Mohagheghi MA, Dennick R. Assessing the skills of surgical residents using simulation. *J Surg Edu.* 2008;65(2):77–83.
40. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J Pers Assess.* 2003;80(1):99–103.
41. Dooley K. Questionnaire programming language, version 1 (software package). Washington, DC: US General Accounting Office; 1988.
42. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–74.
43. Bosman FT. Dysplasia classification: pathology in disgrace? *J Pathol.* 2001;194(2):143–4.
44. Speight PM, Abram TJ, Floriano PN, James R, Vick J, Thornhill MH, et al. Interobserver agreement in dysplasia grading: toward an enhanced gold standard for clinical pathology trials. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2015;120(4):474–82.
45. Lasisi F, Mouchli A, Riddell R, Goldblum JR, Cummings OW, Ulbright TM, et al. Agreement in interpreting villous elements and dysplasia in adenomas less than one centimetre in size. *Dig Liver Dis.* 2013;45(12):1049–55.
46. Barnes L. Diseases of the larynx, hypopharynx, and trachea. Surgical pathology of the head and neck. 3rd ed. Boca Raton: CRC Press; 2008. p. 117–208.
47. Gale N, Michaels L, Luzar B, Poljak M, Zidar N, Fischinger J, et al. Current review on squamous intraepithelial lesions of the larynx. *Histopathology.* 2009;54(6):639–56.
48. Wenig BM. Squamous cell carcinoma of the upper aerodigestive tract: precursors and problematic variants. *Mod Pathol.* 2002;15(3):229–54.
49. Zur Hausen H. Papillomavirus infections—a major cause of human cancers. *Biochim Biophys Acta.* 1996;1288(2):F55–78.
50. Mehanna HM, Rattay T, Smith J, McConkey CC. Treatment and follow-up of oral dysplasia—a systematic review and meta-analysis. *Head Neck.* 2009;31(12):1600–9.
51. Boy SC. Leukoplakia and erythroplakia of the oral mucosa—a brief overview. *SADJ.* 2012;67(10):558–60.
52. Amagasa T, Yamashiro M, Uzawa N. Oral premalignant lesions: from a clinical perspective. *Int J Clin Oncol.* 2011;16(1):5–14.
53. McLaren KM, Burnett RA, Goodlad JR, Howatson SR, Lang S, Lee FD, et al. Consistency of histopathological reporting of laryngeal dysplasia. The Scottish Pathology Consistency Group. *Histopathology.* 2000;37(5):460–3.
54. Coltrera MD, Zarbo RJ, Sakr WA, Gown AM. Markers for dysplasia of the upper aerodigestive tract. Suprabasal expression of PCNA, p53, and CK19 in alcohol-fixed, embedded tissue. *Am J Pathol.* 1992;141(4):817–25.