Inter-examiner Variability in Grading of Oral Epithelial Dysplasia May Cause Suboptimal Management of Oral Potentially Malignant Disorders


ABSTRACT

Oral epithelial dysplasia (OED) grading determines the management guidelines for oral potentially malignant disorders (OPMDs). The subjectivity of OED grading considerably impacts its reliability. **Objective:** This study aimed to assess the reproducibility of and variability in diagnosing and grading OED by oral and medical pathologists, using the conventional WHO 2005 classification. **Material & methods:** Five oral pathologists and one medical pathologist individually examined 200 hematoxylin and eosin-stained histological slides diagnosed as OED from oral pathology archives at the University of Peradeniya. The most experienced examiner’s diagnoses represented the standard for evaluating inter-examiner variability using the unweighted Cohen’s kappa coefficient. **Results:** OED grading among all oral pathologists revealed moderate agreement (kappa value, 0.42–0.50), whereas the medical pathologist showed poor agreement (kappa value, 0.034). The accepted OED diagnoses were mild, 33%; moderate, 24.9%; severe, 32.4%; and no dysplasia, 9.7%. However, 86.5% of the diagnoses by the medical pathologist were mild-no dysplasia. Diagnoses of moderate and severe dysplasia had lesser reproducibility than those of no dysplasia. **Conclusions:** OED grading was only moderately reproducible among oral pathologists and poorly reproducible with regard to the medical pathologist. A more reliable OED grading system is required to improve reproducibility for optimal OPMD management and assessment.

Key words: oral epithelial dysplasia, oral potentially malignant disorders, grading, oral pathologists, reproducibility

INTRODUCTION

Prevention of oral squamous cell carcinoma (OSCC) requires the early detection of changes in the oral environment, which are categorized as oral potentially malignant disorders (OPMDs). The current concept of OPMD management is strictly based on the diagnosis of oral epithelial dysplasia (OED). Dysplasia can be defined as a spectrum of architectural and cytological epithelial changes caused by the accumulation of genetic alterations, associated with an increased risk of progression to OSCC. Several classifications are available for OED diagnosis and grading. However, the conventional World Health Organization (WHO) 2005 classification was the most widely used grading system up to 2017. According to the WHO 2005 classification, OED is categorized into four grades based on the extent and severity of epithelial atypia with regard to its thickness. The architectural and cytological features are newly defined in the WHO 2017 classification, and the possible diagnoses are classified as follows: mild, moderate, and severe dysplasia. Although the WHO 2005 classification included an additional category...
An OED grading system is considered to be acceptable when it has clear value in the clinical environment with regard to management of a lesion, reproducibility, and biological significance. The WHO 2005 system is commonly used in diagnostics; however, it has been criticized for several limitations. Although the diagnostic criteria presents with different grades of increasing severity, they do not directly indicate the malignant transformation potential of the lesion. Similarly, current guidelines for OPMD management do not indicate a specific treatment modality for lesions with moderate dysplasia, which can result in varying outcomes. Management protocols vary from close follow-up to local excision, based on the clinician and institution. The WHO (2005) OED grading system has shown less reproducibility and reliability because of the subjective nature of diagnosis. In the current context, another probable issue is the variability in diagnosing and grading OED among pathologists with different specialties in this field. Nevertheless, OED is a subtle alteration within a thin epithelium, whereas the more commonly observed dysplastic lesions in other regions of the body, such as the cervix, are known to depict a more clear demarcation between the altered cells and normal epithelium. Therefore, the impact of each feature and extent of dysplasia may highly vary based on the diagnosis by an oral pathologist compared with that by a medical pathologist. Eventually, this results in suboptimal diagnosis, management, and poor outcomes in patients with OPMDs. Therefore, this difference in diagnosing OED must be assessed among different pathologists.

Accordingly, this study aimed to evaluate the inter-examiner variability in diagnosing and grading OED using the WHO 2005 classification among six pathologists.

METHODS

The study cohort included 200 hematoxylin- and eosin-stained histological slides diagnosed as different grades of OED that were collected from the archives of the Department of Oral Pathology, Faculty of Dental Sciences, University of Peradeniya. These cases were identified from their codes in the database. All slides were diagnosed using the WHO 2005 classification. These slides were assessed independently by examiners, including five oral pathologists and one medical pathologist. Further, data were collected and analyzed by a separate blinded individual. All examiners were blinded to the initial diagnosis of these patients and the treatment outcome.

Finally, the most experienced examiner’s diagnoses were considered to be the standard to evaluate inter-examiner variability using the unweighted Cohen's kappa coefficient. Unweighted kappa values were interpreted as the following levels of agreement: negative value, disagreement; 0.01–0.20, none or slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–1.00, perfect. The agreement among pathologists in diagnosing and grading OED was assessed.

RESULTS

Inter-examiner agreement in diagnosing and grading OED

Table 1 shows the inter-examiner agreement among the pathologists. Among the oral pathologists, 100% reproducibility of the diagnoses was observed for 48 (25.9%) slides, whereas 80% reproducibility was observed for 50 slides (27%). In these two groups, the agreement was mainly observed with regard to the diagnosis of no dysplasia. In total, 64 (34.6%) slides were observed to have 60% reproducibility for the grades of dysplasia, where the majority represented moderate and severe grades.

The assessment of agreement in diagnosing and grading OED among oral pathologists revealed kappa values in the range of 0.42–0.508. Accordingly, the oral pathologists exhibited moderately good levels of agreement, whereas the medical pathologist revealed only slight agreement (kappa value = 0.034).

Inter-examiner agreement among diagnoses classified according to management

Table 2 shows the inter-examiner agreement among pathologists when the diagnoses was classified according to the management strategy. When the diagnoses were grouped into two categories, no or mild dysplasia and moderate or severe dysplasia, the kappa values were in the range of 0.209–0.945. Furthermore, oral pathologists showed a substantial to an almost perfect level of agreement in the range of 0.819–0.945.

Inter-examiner agreement in diagnosing carcinoma-in-situ and severe dysplasia

Table 3 shows the inter-examiner agreement among oral pathologists in diagnosing carcinoma-in-situ and severe dysplasia. Kappa values in the range of −0.049–0.246 were observed, indicating no agreement to a fair level of agreement.
Table 1. Inter examiner agreement for the different grades of dysplasia classified according to the WHO 2005 classification.

<table>
<thead>
<tr>
<th>Examiner</th>
<th>No dysplasia</th>
<th>Mild dysplasia</th>
<th>Moderate dysplasia</th>
<th>Severe dysplasia</th>
<th>Kappa value</th>
</tr>
</thead>
<tbody>
<tr>
<td>OP 1</td>
<td>9.7% (18)</td>
<td>33% (61)</td>
<td>24.9% (46)</td>
<td>32.4% (60)</td>
<td>-</td>
</tr>
<tr>
<td>OP 2</td>
<td>17.3% (32)</td>
<td>31.4% (58)</td>
<td>16% (30)</td>
<td>35.1% (65)</td>
<td>0.508</td>
</tr>
<tr>
<td>OP 3</td>
<td>21.1% (39)</td>
<td>24.3% (45)</td>
<td>23.3% (43)</td>
<td>31.4% (58)</td>
<td>0.43</td>
</tr>
<tr>
<td>OP 4</td>
<td>13.5% (25)</td>
<td>20.5% (38)</td>
<td>25.9% (48)</td>
<td>40% (74)</td>
<td>0.46</td>
</tr>
<tr>
<td>OP 5</td>
<td>27% (50)</td>
<td>23.2% (43)</td>
<td>22.2% (41)</td>
<td>27.6% (51)</td>
<td>0.42</td>
</tr>
<tr>
<td>MP 1</td>
<td>44.9% (83)</td>
<td>41.6% (77)</td>
<td>10.8% (20)</td>
<td>2.7% (5)</td>
<td>0.034</td>
</tr>
</tbody>
</table>

Results are represented as % (n); OP: oral pathologist; MP: medical pathologist

Table 2. Inter-examiner agreement among diagnoses classified according to management

<table>
<thead>
<tr>
<th>Examiner</th>
<th>No/mild dysplasia</th>
<th>Moderate/severe dysplasia</th>
<th>Kappa value</th>
</tr>
</thead>
<tbody>
<tr>
<td>OP1</td>
<td>42.7% (79)</td>
<td>57.29% (106)</td>
<td>-</td>
</tr>
<tr>
<td>OP2</td>
<td>48.64% (90)</td>
<td>51.35% (95)</td>
<td>0.881</td>
</tr>
<tr>
<td>OP3</td>
<td>45.4% (84)</td>
<td>54.59% (101)</td>
<td>0.945</td>
</tr>
<tr>
<td>OP4</td>
<td>34.05% (63)</td>
<td>65.94% (122)</td>
<td>0.819</td>
</tr>
<tr>
<td>OP5</td>
<td>50.27% (93)</td>
<td>49.73% (92)</td>
<td>0.849</td>
</tr>
<tr>
<td>GP1</td>
<td>86.48% (160)</td>
<td>13.52% (25)</td>
<td>0.209</td>
</tr>
</tbody>
</table>

Results are represented as % (n); OP: oral pathologist; MP: medical pathologist

Table 3: Inter examiner agreement for carcinoma-in-situ and severe dysplasia

<table>
<thead>
<tr>
<th>Examiner</th>
<th>Severe dysplasia</th>
<th>Carcinoma-in-situ</th>
<th>Kappa value</th>
</tr>
</thead>
<tbody>
<tr>
<td>OP1</td>
<td>73.84% (48)</td>
<td>26.15% (17)</td>
<td>-</td>
</tr>
<tr>
<td>OP2</td>
<td>37.03% (20)</td>
<td>62.96% (34)</td>
<td>0.226</td>
</tr>
<tr>
<td>OP3</td>
<td>66.66% (32)</td>
<td>33.33% (16)</td>
<td>-0.049</td>
</tr>
<tr>
<td>OP4</td>
<td>60.34% (35)</td>
<td>39.65% (23)</td>
<td>0.246</td>
</tr>
<tr>
<td>OP5</td>
<td>63.41% (26)</td>
<td>36.58% (15)</td>
<td>0.045</td>
</tr>
</tbody>
</table>

Results are represented as % (n); OP: oral pathologist; MP: medical pathologist

**DISCUSSION**

Diagnosing and grading OED in OPMD is critical because it is the basic factor that determines the management of the disease. Therefore, any influence that can compromise the correct diagnosis should be identified and rectified. The current study was performed to evaluate the inter-examiner reproducibility of OED grading among six pathologists in Sri Lanka. In this study, a moderate level of agreement was observed in diagnosing and grading OED among oral pathologists (Table 1). However, agreement among oral pathologists increased to an almost perfect level when grading was classified according to management requirements (Table 2). With reference to the management requirement, in the Sri Lankan context, OPMD with no or mild dysplasia is managed using the wait-and-watch policy with follow-up at 6-month intervals, with habit intervention. In contrast, OPMD with moderate or severe dysplasia is managed by surgical excision of the lesion, with follow-up at 3–6-month intervals. Therefore, our study revealed
that when an oral pathologist conducts OED grading, >95% of the lesions can receive the appropriate treatment. In contrast, the medical pathologist showed fair-to-poor agreement (Tables 1 and 2). The current study included a medical pathologist to compare our findings with real life observations. In Sri Lanka, medical pathologists often diagnose oral lesions because of the low number and unequal distribution of oral pathologists across the country. According to the WHO 2005 classification, OED is categorized into four grades, namely mild, moderate, and severe epithelial dysplasia and carcinoma-in-situ. However, with regard to statistical analysis in the present study, severe OED and carcinoma-in-situ were combined and considered to be one category (Table 1). When the reproducibility of the diagnoses of severe OED and carcinoma-in-situ was statistically analyzed, even among oral pathologists, only a fair level of agreement was approached (Table 3). Thus, these data additionally confirm that instead of severe OED and carcinoma-in-situ, using only one category is more appropriate, as indicated in the WHO 2017, to decrease inter-examiner variability.

Despite the routine use of the WHO 2005 classification by most reporting pathologists up to 2017, its reproducibility is inadequate. A wide range of kappa values have been reported, including 0.22, 0.37, and 0.58. In the current study, the moderate level of agreement observed among oral pathologists was comparable to an existing study by Kujan et al. Furthermore, no agreement in grading OED has been reported to be because of the subjectivity in the evaluation of established criteria and lack of calibration of criteria used in diagnosing OED. When individual dysplastic features were considered, the highest agreement was observed with regard to the identification of mitotic figures, drop-shaped rete ridges, increased nuclear size, and abnormal variation in cell size. In contrast, the lowest agreement was observed with regard to loss of stratification, loss of polarity of basal cells, abnormal variation in nuclear size, atypical mitotic figures, and nuclear hyperchromatism. Unfortunately, these issues have not been clarified or addressed in the WHO 2017 classification. Accordingly, criteria such as increased mitosis and nuclear hyperchromatism must be suggested in the next revision of the WHO OED classification.

This study has some limitations, of which the main limitation is a single medical pathologist who was able to complete the assessment of all cases. The original study design included three medical pathologists; however, two were unable to participate in the study because of logistical challenges. Another limitation was our inability to correlate OED grading with the outcome of the lesion. This was mainly because information regarding patients lost to follow-up was unavailable. Furthermore, because of logistical challenges, follow-up data could not be obtained because, although the diagnoses were conducted at one center, clinicians who conducted biopsies were distributed throughout the country.

The medical pathologist diagnosed most of the lesions as no dysplasia and mild dysplasia, and no cases of carcinoma-in-situ lesions were diagnosed (Tables 1 and 3). This indicates considerable underdiagnosis that can result in suboptimal management. Although this may not be the sole reason, it may contribute to the high prevalence of OSCC observed in Sri Lanka. In addition, according to Kujan et al. and Nankivell, the most recent WHO classification, binary grading of OED has been introduced as an alternative to improve some of the shortcomings of the conventional OED grading system. Furthermore, Nankivell et al. have shown that the prognostic assessment could be improved by including information regarding risk habits. At present, our center has a report with OED diagnosis with information regarding binary grading. Therefore, it is prudent to introduce this practice to medical pathologists. In addition, if further investigations determine that binary grading is superior to the WHO classification, only using the binary grading in the next revision of the WHO OED classification can be worthwhile. However, it is essential to consider that OED is a continuous pathological process that we attempt to separate into different grades. Therefore, regardless of the extent of careful statistical analysis or categorization, there is some inherent unpredictability in the OED disease course.

CONCLUSION

The present study shows moderate agreement among oral pathologists for OED grading into three separate categories. Diagnoses by the medical pathologist showed poor reproducibility compared with those by the most experienced oral pathologist. Furthermore, we observed that diagnoses of moderate and severe dysplasia had lesser reproducibility than those of no dysplasia. Thus, it is prudent to introduce a more reliable classification for OED grading to improve reproducibility for optimal management, assessment, and prognosis of OPMDs.

CONFLICT OF INTEREST

The authors declared no conflict of interests.

REFERENCES

1. Amarasinghe HK, Usgodaarachchi US, Johnson NW, Laloo RL, Warnakulasuriya S. Betel-quid chewing with or without tobacco is a major risk


(Received August 10, 2018; Accepted March 29, 2019)