

Inter-observer Agreement in Laryngeal Pre-neoplastic Lesions

Sulen Sarioglu · Fulya Cakalagaoglu · Sahande Elagoz · Unsal Ersoy ·
Demet Etit · Sema Hucumenoglu · Ilgin Karaman · Fulya Koybasioglu ·
Sezer Kulacoglu · Irem Onur Paker · Gulay Ozbilim · Yasemin Ozluk ·
Ugur Pabuccuoglu · Arzu Ruacan · Selda Seckin · Aysun Uguz · Ozlem Saraydaroglu ·
Ali Veral · Dilek Yilmazbayhan · Hulya Ellidokuz

Received: 4 May 2010 / Accepted: 30 August 2010 / Published online: 21 September 2010
© Humana 2010

Abstract In this series, laryngeal preneoplastic lesions were evaluated by the classifications of the World Health Organization (WHOC), Ljubljana (LC) and squamous intraepithelial neoplasia (SINC) by multiple observers. The inter-observer agreement (IA) by WHOC for laryngeal lesions had been previously evaluated, but to the best of our knowledge, there are no data for LC and SINC. H&E stained slides from 42 laryngeal biopsies were evaluated by fourteen participants according to WHOC and LC, and SINC was additionally applied by 6. The results were analyzed statistically. The diagnoses which were favored by most participants for each case, according to WHOC, were as follows: squamous cell hyperplasia ($n = 5$; 12%),

mild dysplasia ($n = 11$; 26.2%), moderate dysplasia ($n = 12$; 28.6%), severe dysplasia ($n = 7$; 16.7%), carcinoma in situ ($n = 5$; 12%), and invasive squamous cell carcinoma ($n = 2$; 4.8%). There was a significant difference between the participants for all three classifications; some participants gave lower or higher scores than the others. The mean correlation coefficients (MCC) of the participants were higher for WHOC compared to LC (0.55 ± 0.15 and 0.48 ± 0.14 , respectively). The mean linear-weighted kappa (wKappa) values of participants were not significantly different (0.42 ± 0.10 , 0.41 ± 0.12 and 0.37 ± 0.07 for WHOC, LC and SINC, respectively).

S. Sarioglu (✉) · I. Karaman · U. Pabuccuoglu
Department of Pathology, Dokuz Eylul University Faculty
of Medicine, 35340 Inciralti, Izmir, Turkey
e-mail: sulen.sarioglu@deu.edu.tr

F. Cakalagaoglu · D. Etit
Department of Pathology, Yesilyurt Atatürk Research
and Education Hospital, Izmir, Turkey

S. Elagoz
Department of Pathology, Cumhuriyet University Faculty
of Medicine, Sivas, Turkey

U. Ersoy · S. Hucumenoglu · F. Koybasioglu · I. O. Paker
Department of Pathology, Ankara Diskapi Research
and Education Hospital, Ankara, Turkey

S. Kulacoglu · S. Seckin
Department of Pathology, Ankara Numune Research
and Education Hospital, Ankara, Turkey

G. Ozbilim
Department of Pathology, Akdeniz University Faculty
of Medicine, Ankara, Turkey

Y. Ozluk · D. Yilmazbayhan
Department of Pathology, Istanbul University Faculty
of Medicine, Istanbul, Turkey

A. Ruacan
Department of Pathology, Hacettepe University Faculty
of Medicine, Ankara, Turkey

A. Uguz
Department of Pathology, Cukurova University Faculty
of Medicine, Adana, Turkey

O. Saraydaroglu
Department of Pathology, Uludag University Faculty
of Medicine, Bursa, Turkey

A. Veral
Department of Pathology, Ege University Faculty
of Medicine, Izmir, Turkey

H. Ellidokuz
Preventive Oncology, Dokuz Eylul University Institute
of Oncology, Izmir, Turkey

The kappa values in this series are in agreement with those in previous literature for WHOC, and the similar results obtained for LC and SINC are novel findings. Although the MCC of WHOC was higher, as the wkappa was not significantly different, the findings in this series are not in favor of any of the classifications for better IA for pre-neoplastic laryngeal lesions.

Keywords Inter-observer agreement · Larynx · Pre-neoplastic lesions · World Health Organization classification · Ljubljana classification · Squamous intraepithelial neoplasia classification

Introduction

There have been many proposals for the classification of pre-neoplastic lesions of the laryngeal mucosa, but there is yet no final mutual agreement. Among these, the classifications, World Health Organization (WHOC), Ljubljana (LC) and squamous intraepithelial neoplasia (SINC) are the most widely favored [1–6]. These three are included in the latest classification of the World Health Organization [7]. The categories of these classifications are intended for determining the prognosis. The inter-observer agreement (IA) or variability is very important for any classification. There are a few studies about the IA of WHOC for head and neck lesions [8–18], but there are no data for LC and SINC [14]. In the present study, IA of laryngeal biopsies has been evaluated by WHOC, LC and SINC by multiple observers, in order to determine which of the classifications is better than the others.

Materials and Methods

Hematoxylin- and eosin (H&E)-stained sections from 42 laryngeal biopsies were gathered from 14 pathologists interested in head and neck pathology, participating from 8 cities and 11 centers. The slide set was posted among centers and each participant was requested to review microscopic slides and evaluate the sections preferentially according to WHOC, LC and SINC, based on the criteria described in the recent article by Gale et al. [6] (Table 1). In addition to these diagnostic categories, epithelium without any hyperplastic and neoplastic lesions and invasive squamous cell carcinoma were also accepted. It was announced to the participants that if they give a diagnosis with two categories like “moderate-severe dysplasia”, the higher category would be accepted for statistical analysis. The pathologists knew only the localization of the lesion during the evaluation of the slides.

Table 1 World Health Organization 2005 (WHO), Ljubljana (L) and squamous intraepithelial neoplasia (SIN) classifications [7]

WHO	L	SIN
Squamous cell hyperplasia	Squamous cell hyperplasia	
Mild dysplasia	Basal/parabasal hyperplasia	SINI
Moderate dysplasia	Atypical hyperplasia	SINII
Severe dysplasia	Atypical hyperplasia	SINIII
Carcinoma in situ	Carcinoma in situ	SINIII

All the participants were requested to send their diagnosis electronically to one center and all the data was gathered for statistical analysis.

The participants were asked to complete a questionnaire including the years spent as a pathology specialist, years of special interest or specific working in head and neck pathology, the number of laryngeal biopsies evaluated in September 2009, the classification they apply during daily practice for laryngeal pre-neoplastic lesions, and comments about the working set and organization of the study.

Statistical Analysis

The Friedman and Wilcoxon Signed Rank Test and the Spearman’s Correlation Analysis (SPSS, 11.00), as well as the weighted linear kappa analysis were performed (MedCalc).

Results

Distribution of Cases

The diagnosis which was favored by most of the participants was listed for each case. According to these, the distribution of laryngeal biopsies by WHOC was: squamous cell hyperplasia 5 (12%), mild dysplasia 11 (26.2%), moderate dysplasia 12 (28.6%), severe dysplasia 7 (16.7%), carcinoma in situ 5 [12], and invasive squamous cell carcinoma in 2 (4.8%) cases.

The Profile of the Participants:

The participants had been pathology specialists for 5–24 years and they had been dealing with head and neck pathology for 5–18 years. The number of laryngeal biopsies diagnosed per month ranged from 11 to 31. WHOC was the predominantly applied classification, SINC was applied by one, and WHOC and LC were applied together by one, while two participants reported each case according to the three classifications.

Evaluation Results of Laryngeal Biopsies

All the participants ($n = 14$) applied both the WHOC and LC to laryngeal biopsies, but only 6 applied SINC.

There was a significant difference between the participants when all the cases were considered by WHOC (Friedman's Test 0,000). Two of the participants diagnosed the cases with lower (mean: 2.24 ± 1.32 and 2.60 ± 1.29) and one diagnosed the cases with higher grades (mean: 3.62 ± 1.15) than the others (Wilcoxon Signed Ranks Test with Bonferroni Correction $P = 0.000$).

There was also a significant difference between the participants when all the cases were considered by LC (Friedman's Test 0,000). Again, two of the participants diagnosed the cases with lower (mean: 2.45 ± 0.74 and 2.50 ± 0.63) and one diagnosed the cases with higher grades than the others (mean: 3.05 ± 0.73) (Wilcoxon Signed Ranks Test ≤ 0.05 with Bonferroni Correction $P = 0.000$), but the participant who gave significantly higher grades was somebody else this time.

There was a significant difference between the participants when all the cases were considered by SINC (Friedman's Test 0,000). One of the participants, who was one of the above-mentioned participants, diagnosed the cases as lower grade than the others (2.14 ± 1.20) (Wilcoxon Signed Ranks Test $P < 0.003$).

Ninety-one pairs of Spearman Correlation Analyses were performed for WHOC results, each comparing a pair of participants. There was no correlation for 2 (2%) comparisons. Correlations were mild, moderate and strong for 33 (36.3%), 45 (49.5%) and 11 (12.1%) pairs, respectively.

For LC, the results were no correlation, mild, moderate and strong correlations for 13 (14.3%), 34 (37.4%), 40 (43.9%) and 4 (4.4%) comparisons, respectively.

The SINS results were for 15 comparisons, as only 6 participants applied this method. There were mild, moderate and strong correlations for 4 (26.7%), 10 (66.6%) and 1 (6.6%) pairs (Table 2).

The mean correlation coefficients (MCC) of WHOC was significantly higher than that for LC (mean 0.55 ± 0.15 and 0.48 ± 0.14 , $P = 0.000$, Paired Samples t test).

The MCC of 6 participants who evaluated the cases by three methods, WHOC, LC and SINC, were compared, but

there was no significant difference ($P = 0.24$, Friedmans Test).

The Linear kappa analysis was performed for WHOC, LC and SINC for 65, 54 and 7 pairs as there were missing values for other comparisons, preventing the kappa analysis. The mean kappa values for WHOC, LC and SINC were 0.40 ± 0.11 (range: 0.19–0.65), 0.42 ± 0.12 (range: 0.5–0.71) and 0.42 ± 0.11 (range: 0.29–0.62), respectively (Table 3). The mean values of kappa statistics for matching pairs with WHOC, LC and SINC were 0.42 ± 0.10 (0.19–0.56) (33 pairs), 0.41 ± 0.12 (0.15–0.63) (33 pairs) and 0.37 ± 0.07 (0.29–0.47) (5 pairs). The results of the matching pairs with available kappa values were compared and there was no significant difference between WHOC and LC ($P = 0.44$ Paired samples t test, 33 comparisons), WHOC and SINC, as well as LC and SINC ($P = 0.078$ and $P = 0.144$, respectively, Wilcoxon Signed Ranks Test).

Discussion

The classification of head and neck squamous mucosal pre-neoplastic lesions has been a controversial issue. The criteria of the three most favored classifications, WHOC, LC and SINC, can be compared, but none of the categories match perfectly. Mild dysplasia of WHOC and SINI share features with basal/parabasal hyperplasia of LC. Moderate and severe dysplasia or SINII and III have many features of atypical hyperplasia of LC, but SINIII also includes the lesions that are described in the context of carcinoma in situ by WHOC and LC [1–6, 18].

The criteria for uterine cervical dysplasia have been adapted to laryngeal and oral lesions by WHOC, but the prognostic value and IA of this method in the head and neck are not as good as they are for the cervix [19]. The oral cavity is the other head and neck region where these classifications are applied. The mean IA with WHOC for oral cavity is also lower than that of the cervix; the kappa values were 0.15–0.41 and 0.23–0.45 for two series [8, 9] and the mean kappa value was 0.58 in another [12]. For laryngeal cases, the kappa value was 0.32 in the series by McLaren et al. [11] with WHOC. There are other series with categories censored, but we do not believe that these

Table 2 The distribution of correlation coefficients (r) comparing pairs of observers' results for the laryngeal biopsies by the World Health Organization classification (WHOC), Ljubljana classification (LC) and squamous intraepithelial neoplasia classification (SINC) (Spearman's Correlation Analysis)

	Non $P < 0.05$	Mild $0.25 < r \leq 0.5$	Moderate $0.5 < r \leq 0.75$	Strong $r > 0.75$	Total
WHOC	2 (2)	33 (36.3)	45 (49.5)	11 (12.1)	91 (100)
LC	13 (14.3)	34 (36.3)	40 (43.9)	4 (4.4)	91 (100)
SINC	–	4 (26.7)	10 (66.6)	1 (6.6)	15 (100)

Table 3 The distribution of wKappa values for pairs of observers, with the World Health Organization classification (WHOC), Ljubljana classification (LC) and squamous intraepithelial neoplasia classification (MedCalc)

	0.01–0.20 (%)	0.21–0.40 (%)	0.41–0.60 (%)	0.61–0.80 (%)	Total
WHOC	1 (1.5)	32 (49)	31 (47.7)	2 (3)	65 (100)
LC	3 (7.7)	13 (33.3)	20 (51.3)	3 (7.7)	39 (100)
SINC		4 (57)	2 (28.6)	1 (14.3)	7 (100)

series reflect the IA of WHOC [14, 15]. The lack of data about the IA of the head and neck lesions by LC and SINC is the basis of this study, in addition to providing more data about the IA of WHOC.

Studies for determining IA have some shortcomings. During the examination, the participants are asked to perform things differently from their daily practices. They have to evaluate the same group of patients following each other. They cannot demand new sections or apply additional methods or make a comment about the findings, and they have to select a category. They cannot ask for a repeat biopsy. The sections being evaluated may be from different centers with variable qualities of tissue processing and staining. They are asked to apply different classifications from what they used to do. For example, in this series, only two participants stated that he/she applied the three classifications in daily practice, and one stated that he/she used two (WHOC and LC). The others applied only one, with a high rate of WHOC.

During the reporting of the biopsies, the responsibility is high, and probably more time is spent compared to a research project like the present one. There may be bias towards seeing a pre-neoplastic lesion, as the working set is intended for this. Last but not least, an important factor may be related to the errors saving the results, during the evaluation, writing onto the computer, and homogenization of the data for statistical analyses. We tried to overcome the above-mentioned factors with as much care as possible.

The number of cases studied in this series are lower than the previous ones which evaluated the IA of WHOC for laryngeal biopsies, but in this series, there were more participants and the participants were requested to perform three classifications allowing the comparison of classifications in one series [8, 9, 11, 12]. The participants in this series were pathologists interested in head and neck pathology; hence, the results may not reflect all the pathologists who do not have a special interest or experience in this field. On the other hand, although all the participants were interested in head and neck pathology, some gave statistically significant lower or higher results than the others.

In this series, the weighted kappa statistics were preferred as unweighted kappa considers all disagreements to be equally important, but the weighted kappa provides

better results if there is a low degree of disagreement and worse results if the degree of disagreement is high. For best results with this method, the categories should preferably be nearly equally different from each other [20]. As we intended to prefer the weighted kappa statistics, we had to censor the simple and basal-parabasal hyperplasia categories for LC, as in the recent article by Gale et al. [6]; these two were grouped together as benign lesions with low malignant potential.

In this series, the correlation coefficients of the WHOC were higher than LC for laryngeal pre-neoplastic lesions. The wkappa values of LC, WHOC and SINC were not different. These kappa values were in concordance with the previous findings, yielding fair, nearly moderate agreement (wkappa = 0.40) by WHOC. Both LC and SINC results showed moderate agreement (wkappa = 0.42 and 0.42, respectively), but there was no difference between these three classifications.

Previous studies have suggested that the molecular basis for LC classification for laryngeal neoplastic progression is stronger than for the others [21–24]. This series presents data about IA of WHOC, and novel information about LC and SINC of laryngeal pre-neoplastic lesions. We could not achieve good IA wkappa values by any of the three most favored classifications for preneoplastic laryngeal lesions, in agreement with the previous findings. These results probably point to requirements for the improvement of classifications or pathologists, or even better, both. Although the MCC of WHOC was higher as the wkappa were not significantly different, the findings in this series were not in favor of any of the classifications for better IA for laryngeal lesions.

References

1. Kambic V, Gale N. Significance of keratosis and dyskeratosis for classifying hyperplastic aberrations of laryngeal mucosa. *Am J Otolaryngol.* 1986;7(5):321–3.
2. Resta L, Colucci GA, Troia M, Russo S, Vacca E, Delfino VP. Laryngeal intraepithelial neoplasia (LIN). An analytical morphometric approach. *Pathol Res Pract.* 1992;188:517–23.
3. Michaels L. The Kambic-Gale method of assessment of epithelial hyperplastic lesions of the larynx in comparison with the

- dysplasia grade method. *Acta Otolaryngol* (Stockh). 1997;Suppl 527:17–20.
4. Hellquist H, Cardesa A, Gale N, Kambic V, Michaels L. Criteria for grading in the Ljubljana classification of epithelial hyperplastic laryngeal lesions. A study by members of the working group on epithelial hyperplastic laryngeal lesions of the European Society of Pathology. *Histopathology*. 1999;34:226–33.
 5. Zerdoner D. The Ljubljana classification—its application to oral epithelial hyperplasia. *J Craniomaxillofac Surg*. 2003;31:75–9.
 6. Gale N, Michaels L, Luzar B, Poljak N, Zidar N, Fischinger J, Cardesa A. Current review on squamous intraepithelial lesions of the larynx. *Histopathology*. 2009;54:639–56.
 7. Gale N, Pilch BZ, Sidransky D, Westra WH, Califano J. Epithelial precursor lesions. In: Barnes L, Eveson JW, Reichart P, Sidransky D, editors. *World Health Organization Classification of Tumours. Pathology & Genetics of Head and Neck Tumours*. Lyon: IARC Press; 2005. p. 140–3.
 8. Abbey LM, Kaugars GE, Gunsolly 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.
 9. Karabulut A, Reibel J, Therkildsen 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.
 10. Abbey LM, Kaugars GE, Gunsolley JC, Burns JC, Page DG, Svirsky JA, Eisenberg E, Krutchkoff DJ: the effect of clinical information on the histopathologic diagnosis of oral epithelial dysplasia. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. 1998;85:74–7.
 11. McLaren KM, Burnett RA, Goodlad JR, Howatson SR, Lang S, Lee FD, et al. Consistency of histopathological reporting of laryngeal dysplasia. *Histopathology*. 2000;17:460–3.
 12. Tabor MP, Braakhuis BJ, van der Wal JE, van Diest PJ, Leemans CR, Brakenhoff RH, Kummer JA. Comparative molecular and histological grading of epithelial dysplasia of the oral cavity and the oropharynx. *J Pathol*. 2003;199:354–60.
 13. Brothwell DJ, Lewis DW, Bradley G, Leong I, Jordan RC, Mock D, Leake JL. Observer agreement in the grading of oral epithelial dysplasia. *Community Dent Oral Epidemiol*. 2003;31:300–5.
 14. Fischer DJ, Epstein JB, Morton TH, Schwartz SM. Interobserver variability in the histopathologic diagnosis of oral pre-malignant and malignant lesions. *J Oral Pathol Med*. 2004;33:65–70.
 15. 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;42:987–93.
 16. Warnakulasuriya S, Reibel J, Bouquot J, Dabelsteen E. Oral epithelial dysplasia classification systems: predictive value, utility, weaknesses and scope for improvement. *J Oral Pathol Med*. 2008;37(3):127–33.
 17. Eversole LR. Dysplasia of the upper aerodigestive tract squamous epithelium. *Head and Neck Pathol*. 2009;3:63–8.
 18. Fleskens S, Slootweg P. Grading systems in head and neck dysplasia their prognostic value, weaknesses and utility. *Head and Neck Oncol*. 2009;1:11.
 19. Roteli-Martins CM, Derchain SF, Martinez EZ, Siqueira SA, Alves VA, Syrjänen KJ. Morphological diagnosis of HPV lesions and cervical intraepithelial neoplasia (CIN) is highly reproducible. *Clin Exp Obstet Gynecol*. 2001;28(2):78–80.
 20. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med*. 2005;37(5):360–3.
 21. Sarioglu S, Ozer E, Kirimca F, Sis B, Pabuccuoglu U. Matrix metalloproteinase-2 expression in laryngeal preneoplastic and neoplastic lesions. *Pathol Res Pract*. 2001;197:483–6.
 22. Sengiz S, Pabuccuoglu U, Sarioglu S. Immunohistological comparison of the World Health Organization (WHO) and Ljubljana classifications on the grading of preneoplastic lesions of the larynx. *Pathol Res Pract*. 2004;200:181–8.
 23. Zidar N, Gale N, Cör A, Kambic V. Expression of Ki-67 antigen and proliferative cell nuclear antigen in benign and malignant epithelial lesions of the larynx. *J Laryngol Otol*. 1996;110:440–5.
 24. Poljak M, Gale N, Kambic V, Ferluga D, Fischinger J. Overexpression of p53 protein in benign and malignant laryngeal epithelial lesions. *Anticancer Res*. 1996;16:1947–52.