The Original

EVALUATION OF INTER-OBSERVER AGREEMENT IN HIF2ALPHA EXPRESSION INTERPRETATION FOR ORAL SQUAMOUS CELL CARCINOMA USING FLEISS' KAPPA ANALYSIS

Shandana Fazal[1], Benish Aleem[2*], Asif Ali [3, 4,] Omeed Ahmad Yousafzai[5, 6,] Alina Mehwish[7], Shagufta Jabeen Mirza[8]

[1] MPhil scholar, Institute of Pathology & Diagnostic Medicine, Khyber Medical University, Peshawar 25100, Pakistan
[2] Assistant Professor, Institute of Pathology & Diagnostic Medicine, Khyber Medical University, Peshawar 25100, Pakistan
[3] Associate Professor, Institute of Pathology & Diagnostic Medicine, Khyber Medical University, Peshawar 25100, Pakistan
[4] Institute of Cancer Sciences, University of Glasgow, Glasgow G12 8QQ, UK
[5] MPhil scholar, Institute of Pathology & Diagnostic Medicine, Khyber Medical University, Peshawar 25100, Pakistan
[6] Chief of Diagnostic Department (NCCP) Kabul
[7] Assistant Professor, Jinnah Medical College, Peshawar 25100, Pakistan
[8] MPhil scholar, Institute of Pathology & Diagnostic Medicine, Khyber Medical University, Peshawar 25100, Pakistan
*Corresponding author
Dr. Benish Aleem
Assistant Professor
Institute of Pathology & Diagnostic Medicine, Khyber Medical University, Peshawar 25100, Pakistan
Benish_zubair@hotmail.com

Abstract
Objective: The aim of this research was to evaluate the degree of agreement among three separate pathologists by employing an H-score method to evaluate the presence of HIF2α (hypoxic-inducible factor-2 alpha) in cases of oral squamous cell carcinoma (OSCC).
Methods: H-scores acquired from immunohistochemistry for HIF2α were converted into nominal data (positive or negative) using a predetermined cutoff. Among the 70 samples of OSCC, the expression of HIF2α was independently analysed by three pathologists. Fleiss' kappa value was used to assess the level of agreement among the raters.
Results: A Fleiss' kappa coefficient for inter-observer agreement of 0.630 ($p < 0.05$) showed that the pathologists' agreement in identifying HIF2α expression was much greater than chance.
Conclusion: According to our findings, pathologists' evaluations of HIF2α expression in OSCC using an H-scoring method demonstrate a significant degree of agreement. These results show that Fleiss' kappa is a useful agreement statistic for investigations involving multiple raters and categorical data.

INTRODUCTION

Oral squamous cell carcinoma (OSCC) arises on the oral cavity's mucosal epithelium. It affects swallowing, appearance, pronunciation, and taste perception and is responsible for around 90% of oral cancers. 377,713 OSCC cases were recorded worldwide in 2020. In addition to an increase in mortality, the Global Cancer Observatory (GCO) projects that by 2040, the incidence of OSCC will increase by almost 40%[1]. Oral squamous cell carcinoma (OSCC) is a major cause of mortality and a significant health issue. The source of the condition can be traced back to the mucosal epithelium lining the oral cavity, throat, and larynx[2]. OSCC is a widespread malignancy affecting the head and neck region, which has exhibited an increasing incidence in several countries, notably among the younger population[3]. The prevalence of OSCC varies with different countries but it has typically been linked to either excessive alcohol use, exposure to carcinogens produced from tobacco, or both[2, 4]. Compared to other sub-sites such as the floor of mouth, gingivae, and retro molar trigone, tongue is the most frequently occurring site and is linked to a greater death rate[5-7].
Oral squamous cell carcinoma (OSCC) accounted for 890,000 new cases and 450,000 deaths in 2018, making it the sixth most frequent cancer globally[8, 9]. According to the Global Cancer Observatory (GLOBOCAN), the incidence of cancer will have risen by 30% by 2030, translating into 1.08 million new cases annually[10, 11]. The risk of having OSCC is often two to four times higher in men than in women[7]. It is still unknown what molecular characteristics and environmental elements contributed to the early formation of OSCC. Research has focused on how alcohol and tobacco use may contribute to the early development of OSCC[12]. In a study chewing areca nut products, such as "betel quid," is linked to oral cavity cancer in populations of Asia-Pacific. The term refers to a

range of modified mixtures that include areca nut (Areca catechu, the carcinogen source), betel leaf (the leaf of Piper betle), slaked lime, tobacco, and/or spices in accordance with local custom[13, 14].

The management of OSCC depends on the stage and histological grade of the disease, either surgical excision, with or without adjuvant therapy (such as radiotherapy or chemo radiotherapy) is typically used for treatment[15]. Features such as differentiation, growth pattern, invasion depth, margins status, vascular invasion or neural invasion, involvement of bone, status of lymph nodes involved, metastasis, extension of the tumor outside, and TNM stage will determine the indication for adjuvant treatment. These features are detailed in the standardized histopathology report of the resection[16].

Under normal oxygen levels, HIF-2α, which is a component of the HIF family responsible for sensing oxygen, is constantly produced and degraded by proline hydroxylases and Factor-inhibiting hypoxia inducible factor (FIH)[17]. Cells are often hypoxic within the tumor microenvironment. Angiogenesis, glycolysis, erythropoiesis, and eventual cancer growth are all mediated by genes whose overexpression is the homeostatic adaptation to this challenge. In a rapidly growing tumor, hypoxic areas are frequently formed[18, 19].

This happens when the tumour mass doesn't receive enough blood due to constant growth, hence leading to activation of HIF family of transcription factors which in turn plays a major role in mediating the cellular response to hypoxia[20, 21]. The presence of HIF-2α in different non vascular locations like kidney, sympathetic nervous system and liver indicates a link between HIF-2α and systems sensitive to hypoxia, such as nerves and blood vessels. Furthermore, increased angiogenesis and hypoxia sensitivity in tumors is often associated with elevated HIF-2α levels[20].

Tumor hypoxia causes resistance to conventional cancer treatments, including radiation and chemotherapy. Therefore the current gold standard for HIF-targeting cancer therapies is immunotherapy combined with targeted treatment[22].

The current study was done to examine HIF-2α expression in OSCC patients through immunohistochemistry. Thus this article aims to evaluate the degree of agreement between different observers/pathologists in classifying the expression of HIF2α in OSCC samples using H-score values[23]. Results were interpreted according to the categories that were proposed by Fleiss for the interpretation of the kappa coefficient of poor agreement (kFleiss<0.40), moderate agreement (kFleiss between 0.40–0.75), and excellent agreement (kFleiss>0.75)[24].

MATERIALS AND METHODS

This study was carried out from 2022 to 2023 at Khyber Medical University's (KMU) Institute of Pathology and Diagnostic Medicine's (IPDM) Histopathology Laboratory. 70 oral squamous cell carcinoma (OSCC) cases recorded between 2017 and 2022 at IPDM, Rehman Medical Institute (RMI), and Khyber College of Dentistry (KCD), Peshawar, were utilized as formalin-fixed paraffin-embedded (FFPE) tissue blocks. Records from the department were used to gather patient information, such as age, gender, tumor location, and histologic variation. A microtome was used to cut each FFPE block into 4 µm slices, which were then floated in a water bath at 45 to 50°C and placed on IHC slides coated with poly-L-lysine and H&E. An Olympus CX21LED light microscope was used for H&E staining for histopathologic grading, and a Canon EOS 550D camera was used to take pictures. ImageJ software (version 1.52a) was used to add scale bars to the photos. The FineTest Rabbit-DAB (Poly-HRP) Detection IHC Kit (Wuhan Fine Biotech Co., Ltd) was used to assess the immunohistochemical (IHC) expression of HIF-2α in OSCC samples. This kit contained peroxidase blocking reagent, HRP, substrate buffer, DAB with chromogenic substrate, antibody dilution buffer, and antigen retrieval solution (10×, pH 9). HIF-2α Rabbit Polyclonal Antibody (Catalog No. Fnab10928, UniProt ID: 2.2.5.2) was the main antibody utilized. A refrigerator, vacuum chamber, oven, microtome, micropipettes, staining troughs, flasks, beakers, slides, cover slips, and microscope were among the standard histopathology tools used in the study. Other standard chemicals included 10% formalin, graded alcohol, xylene, paraffin, distilled water, and H&E stains. HIF-2α expression was found in OSCC tissue samples using immunohistochemistry (IHC). Tissue blocks were sectioned into 4 µm-thick slices using a microtome and then floated on a water bath before being mounted onto poly-L-lysine–coated IHC slides to ensure proper adhesion.

HIF2α scoring

IHC-stained slides were examined under a microscope to determine the expression of HIF-2α in OSCC tissue samples. To avoid bias, the histopathologist was not given access to any clinicopathologic information. Tumor cell cytoplasmic labeling revealed HIF-2α positivity. An H-score between 0 and 300, where 300 denoted strong staining in 100% of tumor cells, was determined by analyzing staining intensity and the proportion of positive cells using a 400× light microscope. Positive and negative cases were categorized using a receiver operating characteristic (ROC) curve analysis cut-off value of 1. With diagnostic accuracy rising as the curve gets closer to the upper left corner of the graph, the ROC curve aids in determining the ideal threshold that best distinguishes between positive and negative cases; on the other hand, a diagonal line indicates a test with no discriminatory power (sensitivity equals specificity)[25, 26].

ANALYSIS OF INTER-OBSERVER AGREEMENT

To determine how consistently HIF2α expression in oral squamous cell carcinoma (OSCC) samples was classified, three pathologists' inter-observer agreement was evaluated. Using H-score, each pathologist independently assessed HIF2α expression in OSCC tissue samples. By evaluating the same histopathological slides or samples and testing interrater agreement for histopathological parameters, the aim is to assess the consistency with which different raters scored using the same method. H-score values were divided into two categories for nominal data: a positive for scores over a predefined cut-off which was calculated as "1" and a negative for scores below or equal to the cut-off.

STATISTICAL ANALYSIS

All the collected data were first structured in Microsoft Excel 2016 and then analyzed with IBM SPSS Statistics Version 25.0. Continuous variables were summarized as mean and standard deviation, whereas categorical variables were reported as frequencies and percentages, with graphical representations where suitable. ANOVA was used to make comparisons between study groups. The agreement between observers was further investigated using kappa statistics and cross-tabulation methods. A p-value less than 0.05 indicated statistical significance.

FLEISS' KAPPA ANALYSIS:

The degree of inter-observer agreement amongst three pathologists was measured using Fleiss' kappa statistic. Fleiss' kappa is appropriate for nominal categorical data and takes into consideration agreement between several raters that goes beyond chance[23, 27-29].

RESULTS:

It was determined that the three raters' inter-observer agreement had a Fleiss' kappa coefficient of **0.630**.

	Kappa	Standard Error	z	Sig.	Lower Bound	Upper Bound
Overall Agreement	**0.630**	0.049	12.905	.000	0.534	0.725

Sample data contains 70 effective subjects and 3 raters.

INTERPRETATION OF RESULTS:

A statistical metric called Fleiss' kappa (κ) is employed to evaluate the degree of agreement between numerous raters or observers when assessing categorical data. Although there are other ways to interpret Fleiss' kappa values, the Landis and Koch (1977) scale is typically used to determine the degree of agreement[30]. According to the Landis and Koch scale, the Fleiss' kappa values are normally interpreted as follows[31]:

K	Interpretation
<0	Poor agreement
0.0-0.20	Slight agreement
0.21-0.40	Fair agreement
0.41-0.60	Moderate agreement
0.61-0.80	Substantial agreement
0.81-1.0	Almost perfect agreement

LANDIS AND KOCH SCALE[31]

A satisfactory degree of inter-observer agreement was seen among the pathologists in this investigation, as indicated by the considerable Fleiss' kappa coefficient (0.630). This suggests that raters were able to interpret HIF2α expression consistently due to the way they used H-score values and the applied cut-off to evaluate the expression.

These explanations offer a qualitative evaluation of the degree of agreement seen amongst several raters or observers. Stronger agreement than would be expected from random chance is shown by an increase in Fleiss' kappa value. The specific factors and environment of a study or research area can influence the interpretation of Fleiss' kappa. Depending on the study's specific requirements and standards, alternative scales or interpretations may potentially be used. Fleiss' kappa values are often interpreted using the Landis and Koch scale.

DISCUSSION:

To understand the tumor's behavior and its potential treatment effects, it is crucial to assess the expression of hypoxia-inducible factor 2 alpha (HIF2α) in oral squamous cell carcinoma (OSCC)[32]. Our results are consistent with other research that has highlighted the role of HIF2α in tumor hypoxia and OSCC aggressiveness. Under hypoxic conditions, HIF2α is recognized to play a critical role in promoting angiogenesis and facilitating tumor growth[33]. The high level of agreement amongst observers indicates that HIF2α expression may be consistently detected and interpreted through the use of immunohistochemical techniques and the approved scoring method. A significant degree of agreement was found among the pathologists when Fleiss' kappa statistic was applied; the computed kappa value was 0.630 which coincides with previous studies[34-36].

A study done by Sohee Park, Hyunho Park and Sang Min Lee in 2022 showed that five readers' agreement on Lung-RADS categorization was substantial (Fleiss kappa, 0.65 [95% CI, 0.63, 0.68]) which corresponds to our study[37]. Guan-Yuh Ho and Matthias Leonhard conducted a study in 2019 to assess the level of agreement among seven examiners. They used the Fleiss method to calculate the kappa value, which was found to be 0.61. The consensus among the seven raters' professional opinions regarding the selected laryngeal electromyography (LEMG) was strong[24].

In 2019, Adam Bonfield and Damian Roland conducted a study in England to assess the Inter-rater reliability (IRR) of the Paediatric Observation Priority Score (POPS), which is a specialized scoring system designed for use in emergency departments for children. The results indicate a statistically significant and reliable agreement of the POPS through Fleiss kappa when evaluating various clinical presentations across healthcare practitioners with varying levels of experience[38]. This Fleiss kappa value is somehow is coinciding with our study.

Fleiss' kappa is a useful tool for analyzing agreement and guaranteeing consistency across different evaluators in studies involving numerous raters or observers, such as those evaluating biomarker expression in tumor samples[39]. This measure emphasizes the consistency in interpreting HIF2α expression amongst many observers by accounting for agreement beyond that which would be predicted by chance. Research findings on HIF2alpha expression in OSCC are better interpreted and have greater validity when Fleiss' kappa analysis is included. It raises the overall standard of the study by offering statistical proof for the validity of data interpretation.

LIMITATIONS

Our study had specific limitations that could have influenced the interpretation of our research findings.

- The fundamental limitation of the study design is the total number of OSCC cases. It influences the implementation of outcomes in the population and the analysis of data.
- Our hospitals had a limited number of cancer registries in terms of statistics.

CONCLUSION:

Our data show good agreement between pathologists' estimates of HIF2α expression in OSCC using a modified scoring approach. These findings demonstrate that Fleiss' kappa is an effective agreement statistic for multi-rater research with categorical data.

REFERENCES

1. Tan Y, Wang Z, Xu M, Li B, Huang Z, Qin S, et al. Oral squamous cell carcinomas: state of the field and emerging directions. International journal of oral science. 2023;15(1):44.
2. Kumar SA, Indu S, Gautami D, VarRuchi SJRJoBV. Oral squamous cell carcinoma (OSCC) in humans: Etiological Factors, diagnostic and therapeutic relevance. 2020;15:10.
3. Joshi P, Dutta S, Chaturvedi P, Nair SJRMmj. Head and neck cancers in developing countries. 2014;5(2).
4. Ford PJ, Rich AMJA. Tobacco use and oral health. 2021;116(12):3531-40.
5. Farhood Z, Simpson M, Ward GM, Walker RJ, Osazuwa-Peters NJTL. Does anatomic subsite influence oral cavity cancer mortality? A SEER database analysis. 2019;129(6):1400-6.
6. Almangush A, Mäkitie AA, Triantafyllou A, de Bree R, Strojan P, Rinaldo A, et al. Staging and grading of oral squamous cell carcinoma: An update. 2020;107:104799.
7. Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JRJNrDp. Head and neck squamous cell carcinoma. 2020;6(1):92.
8. Ghaly AM, Elshenshawy HMA, Abd El Hafez A, Ibrahim MMA, El-Sissi AAIJAB. The prognostic role of hypoxia and the microenvironmental acidity in chemo-radio resistance in oral squamous cell carcinoma patients. 2023;94(3):e20237114.

9. Bolesina N, Gatti G, López de Blanc S, Dhooge S, Rocha D, Fernández E, et al. Oral squamous cell carcinoma (OSCC) tumors from heavy alcohol consumers are associated with higher levels of TLR9 and a particular immunophenotype: Impact on patient survival. 2022;13:941667.
10. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. 2019;144(8):1941-53.
11. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal AJCacjfc. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. 2018;68(6):394-424.
12. Batistella EA, Gondak R, Rivero ERC, Warnakulasuriya S, Guerra E, Porporatti AL, et al. Comparison of tobacco and alcohol consumption in young and older patients with oral squamous cell carcinoma: a systematic review and meta-analysis. 2022;26(12):6855-69.
13. Singh A, Dikshit R, Chaturvedi PJSu, misuse. Betel nut use: the South Asian story. 2020;55(9):1545-51.
14. Waqatakirewa L. Social Cognitive Theory Constructs that Predict Betel Nut Chewing Among Secondary Students: Walden University; 2020.
15. Givony SJJMS. Oral squamous cell carcinoma (OSCC) an overview. 2020;8:67-74.
16. Liu T, David M, Ellis O, Low T-HH, Palme CE, Clark J, et al. Treatment for oral squamous cell carcinoma: impact of surgeon volume on survival. 2019;96:60-5.
17. Gkagkalidis K, Kampantais S, Dimitriadis G, Gourvas V, Kapoukranidou D, Mironidou-Tzouveleki MJMMM. Expression of HIF-2a in clear-cell renal cell carcinoma independently predicts overall survival. 2020;53:229-37.
18. Fang JS, Gillies RD, Gatenby RA, editors. Adaptation to hypoxia and acidosis in carcinogenesis and tumor progression. Seminars in cancer biology; 2008: Elsevier.
19. Liao C, Liu X, Zhang C, Zhang Q, editors. Tumor hypoxia: From basic knowledge to therapeutic implications. Seminars in cancer biology; 2023: Elsevier.
20. Wang W-J, Ouyang C, Yu B, Chen C, Xu X-F, Ye X-QJOR. Role of hypoxia-inducible factor-2α in lung cancer. 2021;45(5):1-10.
21. Jarman EJ, Ward C, Turnbull AK, Martinez-Perez C, Meehan J, Xintaropoulou C, et al. HER2 regulates HIF-2α and drives an increased hypoxic response in breast cancer. 2019;21:1-18.
22. Graham K, Unger EJIjon. Overcoming tumor hypoxia as a barrier to radiotherapy, chemotherapy and immunotherapy in cancer treatment. 2018:6049-58.
23. Tabor MP, Braakhuis BJ, van der Wal JE, van Diest PJ, Leemans CR, Brakenhoff RH, et al. Comparative molecular and histological grading of epithelial dysplasia of the oral cavity and the oropharynx. The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland. 2003;199(3):354-60.
24. Ho G-Y, Leonhard M, Volk GF, Foerster G, Pototschnig C, Klinge K, et al. Inter-rater reliability of seven neurolaryngologists in laryngeal EMG signal interpretation. 2019;276:2849-56.
25. Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. Korean journal of anesthesiology. 2022;75(1):25-36.
26. Polo TCF, Miot HA. Use of ROC curves in clinical and experimental studies. SciELO Brasil; 2020. p. e20200186.
27. Falotico R, Quatto P. Fleiss' kappa statistic without paradoxes. Quality & Quantity. 2015;49(2):463-70.
28. Yan F, Cheng Y-SL, Katabi N, Nguyen SA, Chen H-S, Morgan P, et al. Interobserver variation in evaluating perineural invasion for oral squamous cell carcinoma: phase 2 survey study. Head and Neck Pathology. 2021;15(3):935-44.
29. Farah CS, Janik M, Woo Sb, Grew J, Slim Z, Fox SA. Dynamic real-time optical microscopy of oral mucosal lesions using confocal laser endomicroscopy. Journal of Oral Pathology & Medicine. 2023;52(6):539-47.
30. Foody GMJRsoe. Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification. 2020;239:111630.
31. Marchevsky AM, Walts AE, Lissenberg-Witte BI, Thunnissen EJAoDP. Pathologists should probably forget about kappa. Percent agreement, diagnostic specificity and related metrics provide more clinically applicable measures of interobserver variability. 2020;47:151561.
32. Shi R, Liao C, Zhang QJC. Hypoxia-driven effects in cancer: characterization, mechanisms, and therapeutic implications. 2021;10(3):678.
33. Luo D, Liu H, Lin D, Lian K, Ren HJCE, Biomarkers, Prevention. The clinicopathologic and prognostic value of hypoxia-inducible factor-2α in cancer patients: a systematic review and meta-analysis. 2019;28(5):857-66.
34. Sørensen JS, Vilstrup MH, Holm J, Vogsen M, Bülow JL, Ljungstrøm L, et al. Interrater agreement and reliability of PERCIST and visual assessment when using 18F-FDG-PET/CT for response monitoring of metastatic breast cancer. 2020;10(12):1001.

35. Suh YJ, Lee JW, Shin SY, Goo JM, Kim Y, Yong HSJEr. Coronary artery calcium severity grading on non-ECG-gated low-dose chest computed tomography: a multiple-observer study in a nationwide lung cancer screening registry. 2020;30:3684-91.
36. Radovic N, Ivanac G, Divjak E, Biondic I, Bulum A, Brkljacic BJJomri. Evaluation of breast cancer morphology using diffusion-weighted and dynamic contrast-enhanced MRI: intermethod and interobserver agreement. 2019;49(5):1381-90.
37. Park S, Park H, Lee SM, Ahn Y, Kim W, Jung K, et al. Application of computer-aided diagnosis for Lung-RADS categorization in CT screening for lung cancer: effect on inter-reader agreement. 2022:1-11.
38. Bonfield A, Roland DJEMJ. Inter-Rater reliability in a bespoke scoring tool: the paediatric observation priority score. 2020;37(5):293-9.
39. Shen Y, Yoon MN, Ortiz S, Friesen R, Lai HJDJ. Evaluating Classification consistency of oral lesion images for use in an image classification teaching tool. 2021;9(8):94.