



Variability in Depth of Invasion Measurements in Carcinomas of the Oral Cavity and the Effect on Pathologic Tumor Staging

Mary K. Allen-Proctor¹ · Mobeen Rahman¹ · Chandana A. Reddy² · Shlomo A. Koyfman² · Deborah J. Chute¹ · Christopher C. Griffith¹

Received: 14 December 2021 / Accepted: 25 February 2022 / Published online: 30 April 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Depth of invasion (DOI) was added to the staging criteria for carcinoma of the lip and oral cavity in the 8th edition of the American Joint Committee on Cancer Staging Manual (AJCC8). However, there are multiple practical challenges to obtaining an accurate DOI measurement with limited data regarding interobserver variability in DOI measurement. The aim of this study was to investigate interobserver variability in DOI measurement and its effect on tumor stage. We performed an electronic medical record search for excisions of squamous cell carcinoma of the oral cavity between January 1, 2010 and December 25, 2017. All slides containing significant tumor were selected for independent blinded DOI measurement by four head and neck pathologists per AJCC8 guidelines. Pathologic stage was assigned in conjunction with reported tumor greatest dimension. Observers recorded the slide used for measurement and potential issues limiting assessment of DOI. Results were compared for reproducibility in DOI and tumor stage using intraclass correlation coefficient (ICC) analysis. A total of 167 cases of oral squamous cell carcinoma with available slides were included. The ICC score for DOI between observers was 0.91339 (>0.9 considered excellent). Only 7.2% of cases had uniform DOI amongst observers. Increasing overall tumor size and average DOI correlated with increasing range in DOI amongst observers. Differences in DOI resulted in differences in pathologic tumor staging (pT) for 15% of tumors. Use of different slides for DOI measurements was significantly associated with different pT staging. In contrast, ulceration and exophytic growth did not correlate with higher DOI or pT variability. Despite the excellent ICC score, differences in DOI measurement resulted in variable pT staging for a considerable number of cases. We therefore recommend consensus for DOI in at least some cases in which potential differences in DOI could alter pT stage assignment.

Keywords Observer variation · Squamous cell carcinoma · Oral cavity · Cancer · Head and Neck Neoplasms · Neoplasm Staging · Prognosis

Introduction

The eighth edition of the American Joint Committee on Cancer Staging Manual (AJCC8) was published in 2017 and contained significant changes for staging head and neck cancers [1, 2]. Considerable changes were made in relation

to staging of human papillomavirus associated carcinoma of the oropharynx and in pathological nodal staging. Pathologic staging of oral cavity tumors was also changed to incorporate depth of invasion (DOI) and remove invasion of extrinsic tongue muscles as a criterion for pT4. Tumor thickness (TT) had long been noted as a potential prognostic indicator for oral cavity carcinoma [3, 4], but more recent evidence suggests DOI may be a better predictor of outcome. Specifically, DOI was shown to be a strong predictor of local recurrence and nodal metastasis [5–7]. As a result of such data, DOI was incorporated into the pathologic tumor (pT) staging of carcinomas of the lip and oral cavity in AJCC8 [2].

Prior to AJCC8, the precise definition of DOI was not clear and frequently confused with TT. In incorporating the

✉ Christopher C. Griffith
griffic8@ccf.org

¹ Department of Anatomic Pathology, Robert J. Tomsich Pathology and Laboratory Medicine Institute, Cleveland Clinic, 9500 Euclid Ave, L2, 44195 Cleveland, Ohio, USA

² Department of Radiation Oncology, Taussig Cancer Institute, Cleveland Clinic, Cleveland, Ohio, USA

use of DOI in oral cavity cancer staging, AJCC8 provided a precise definition for DOI as the deepest extent of tumor measured from the basement membrane of the most adjacent normal surface epithelium. To make this measurement, a horizontal line should be made between adjacent normal epithelial basement membrane and a perpendicular “plumb line” dropped to the tumor’s deepest point of invasion. The differences in TT and DOI measurements are expected to more accurately reflect tumor aggressiveness in that thicker, exophytic tumors have lower DOI while thinner ulcerated tumors have greater DOI. In other words, DOI will be less than TT in exophytic tumors with less extensive invasion and greater than TT in at least some ulcerated tumors. Despite the potential for DOI to better predict disease outcomes, a number of practical challenges arise in accurately and consistently determining DOI. [8, 9]

The aim of this study was to investigate interobserver variability in DOI measurement and the effect on pT staging. Potential sources of discrepancy in staging were also examined. Finally, we offer recommendations with the goal of reducing clinically significant changes in patient care as a result of variability in DOI measurements and pT stage.

Methods

Our institution’s laboratory information system (CoPath) was searched for excisions of squamous cell carcinoma of the oral tongue, gingiva, and floor of mouth performed between January 1, 2010 and December 25, 2017. All glass slides containing significant tumor were selected for independent review by four head and neck pathologists. These four pathologists all practiced in the same institution. One was the current head and neck pathology fellow. The three staff pathologists included two formally trained head and neck pathologists with fellowship training at different institutions. The third staff pathologist did not do a formal head and neck pathology fellowship but had practiced with a focus in this area. Any prior markings on slides were removed. DOI was determined independently by each pathologist following the guidelines provided in AJCC8. Specifically, a line from the basement membrane of the adjacent uninvolved epithelium was used as the surface and DOI was measured with a “plumb line” to the tumor’s greatest depth of invasion [2]. DOI measurements for all observers were obtained by placing a transparent ruler on the glass slide and observed through the microscope. In some cases, the deepest extent of tumor extended to the margin and/or the section edge. In these cases, observers measure the deepest extent of invasion possible and provided this as the DOI for the study. Each observer’s DOI measurement was recorded and used in conjunction with the reported greatest

tumor dimension to assign a pathologic T stage. The specific slide used by each pathologist for measurement and factors potentially impeding DOI assessment were also noted.

Two pathologists (#1 and #4) did not provide DOI measurements in four cases each as they felt DOI could not be measured on the available slides accurately. In these cases, the statistical analysis was performed using only provided DOI measurements. For purposes of statistical analysis, individual cases were grouped based on the “range in DOI” between different pathologists. These “range in DOI” categories were “uniform” (all DOI measurements were identical), “0.5 to < 5 mm” (at least one DOI was different and the overall range in DOI was less than 5 mm), “5 to less than 10 mm” (DOI differences ranged from 5 mm to less than 10 mm) and “≥10mm” (DOI measurements varied more than 10 mm between observers). Various comparisons were made between “range in DOI” groups, average size, average DOI and pT differences. Average DOI was calculated as the average DOI measurements provided by all observers for a particular case.

Statistical Methods

Study data was analyzed using a variety of different statistical tests. Analysis of variance (ANOVA) was done to determine if there were significant differences in DOI among pathologists. The intraclass correlation coefficient (ICC) was used to assess concordance in DOI measurements for each case reviewed among all four pathologists. ANOVA was also used to determine if there were differences in tumor size or DOI by “range in DOI”. The chi-square test for trend was used to determine if there were differences in the likelihood of tumors being classified as ulcerated or with exophytic growth based on “range in DOI” and if “range in DOI” varied based on whether the same slide was reviewed by different pathologists. The unpaired t-test was done to determine if there was a differences in pT stage by tumor size or DOI. Fisher’s exact test was used to determine if pT stage differed based on whether pathologists reviewed the same slide or if tumors were noted to be ulcerated or exophytic. All statistical analyses were performed using SAS v9.4 (SAS Institute, Cary, NC).

Results

A total of 167 cases with available slides of invasive carcinoma were identified. 128 (76.6%) were from oral tongue, 24 from gingiva and 15 from floor of mouth. Overall tumor size was available for 165 cases with an average size of 2.75 cm (median 2.5 cm). Table 1 summarizes DOI measurements by each pathologist. All pathologists had the same median DOI

Table 1 Overview of depth of invasion by observer

	N	Median	Mean	Max	p-value
Tumor size (cm)	165	2.5	2.75	7.5	
DOI					0.6503*
DOI Path 1 (mm)	163	8	8.5	27	
DOI Path 2 (mm)	167	8	9.0	26	
DOI Path 3 (mm)	167	8	8.2	22	
DOI Path 4 (mm)	163	8	8.4	25	
DOI Average (mm)	167	7.88	8.514	23	

Abbreviations: DOI – Depth of invasion, N – Number

*p-value from ANOVA, there were no significant p-values when comparing DOI between pathologists

of 8 mm. Although pathologist 3 trended toward a lower mean and max DOI, ANOVA analysis showed no significant difference (p-value = 0.6503) in overall DOI measurements by individual pathologists. Differences in oral subsites (e.g. oral tongue, gingiva, buccal) could not be determined in this study due to insufficient numbers of non-oral tongue primary lesions.

Table 2 summarizes cases by “range in DOI” group based on overall differences in DOI measurements by pathologists. 12 of 167 (7.2%) had uniform DOI measurements and these represented tumors with an average overall size of 1.68 cm. The majority of cases, 124 of 167 (74.3%) had range of DOI between 0.5 and less than 5 mm and these tumors averaged 2.43 cm in overall tumor size. A range of DOI of 5 mm or more was observed in a total of 31 of 167 (18.6%) of cases, and these tumors had average overall size greater than 4 cm. When correlating tumor size with ranges of DOI, larger tumors and those with a higher average DOI were significantly more likely to have a wider range in DOI measurements. Ulceration or exophytic growth appeared to

influence the level of agreement in DOI between observers but, these differences were not statistically significant. Cases with a narrower range in DOI measurements were also significantly more likely to have had observers all measure from the same slide.

In total, 25 of 167 (15%) cases had difference in pT stage based on the individual observer DOI measurements. Not surprisingly, the rate of cases with differing pT stages was greater in those cases that had wider ranges in individual observer DOI. Differences in pT did not show significant association with overall tumor size, average DOI or growth pattern (ulcerated or exophytic). Similarly to the ranges in DOI, cases with uniform pT between all observers were associated with a higher rate of all observers choosing the same slide to measure DOI. The figure demonstrates different DOI measurements obtained from several slides for the same case (Figure 1). The variable DOI measurements in this case resulted in differences in pT stage between pT1 and pT2. Table 3 provides additional details on those cases with different pT staging between observers. In most cases (21 of 25), three of the four observers provided the same pT despite differences in DOI with only a single observer concluding a different pT stage. Only 4 of the 25 cases had an even split of 2 vs. 2 for disagreements in pT. Most (21 of 25) of the differences in pT were a result of the DOI differences spanning the staging cut-point of 10 mm. These differences in pT were only a single step different and most commonly resulted in differences between pT2 and pT3.

Despite the differences in DOI and pT detailed above, the ICC analysis showed overall excellent agreement in DOI measurements between observers by case with an ICC of 0.91339 (Table 4). Given the significance of using the same

Table 2 Summary of tumor characteristics by depth of invasion range and pT stage

		N (%)	Average tumor size (cm)	Average DOI (mm)	Ulcerated, N (%)	Exophytic, N (%)	Same slide measured, N (%)
DOI range	Uniform	12 (7.2)	1.68	4.75	8/12 (66.7)	3/12 (25)	4/12 (33.3)
	0.5 to <5 mm	124 (74.3)	2.43	7.38	64/124 (51.6)	13/124 (10.5)	61/124 (49.2)
	5 to <10 mm	25 (15.0)	4.30	13.81	15/25 (60)	0/25 (0)	5/25 (20.0)
	≥10 mm	6 (3.6)	5.02	17.46	4/6 (66.7)	1/6 (16.7)	0/6 (0)
p-value			<0.0001****	<0.0001****	0.7147*	0.1213*	0.0111*
pT	Same pT	142 (85.0)	2.66	8.31	74/142 (52.1)	15/142 (10.6)	66/142 (46.5)
	Different pT	25 (15.0)	3.30	9.66	17/25 (68)	2/25 (8)	4/25 (16.0)
p-value			0.1004***	0.2659***	0.1914**	>0.9999**	0.0012**
	All cases	167	2.8	8.53	91/167(54.5)	17/167 (10.2)	69/167 (41.3)

*p-value from Chi-square test for trend

**p-value from Fisher’s exact test

***p-value from unpaired t-test

****p-value from ANOVA, comparisons between pair ‘Uniform and 0.5 to <5 mm’ and pair ‘5 to <10 mm and ≥10 mm’ were not significant, comparisons between other pairs were significant

Fig. 1 Case with variable depth of invasion measurements using different slides and resulting in different pathologic T staging. Two observers used slide A4 (top panel) and provided DOI measurements of 4 mm (blue) and 5 mm (yellow). One observer used slide A5 (bottom panel) to obtain a measurement of 6 mm (green). A fourth observer also provided a measurement of 4 mm but from slide A3 (not shown). This tumor had an overall dimension of 1.2 cm and therefore the 6 mm depth measurement resulted in pT2 staging while the other measurements resulted in pT1.

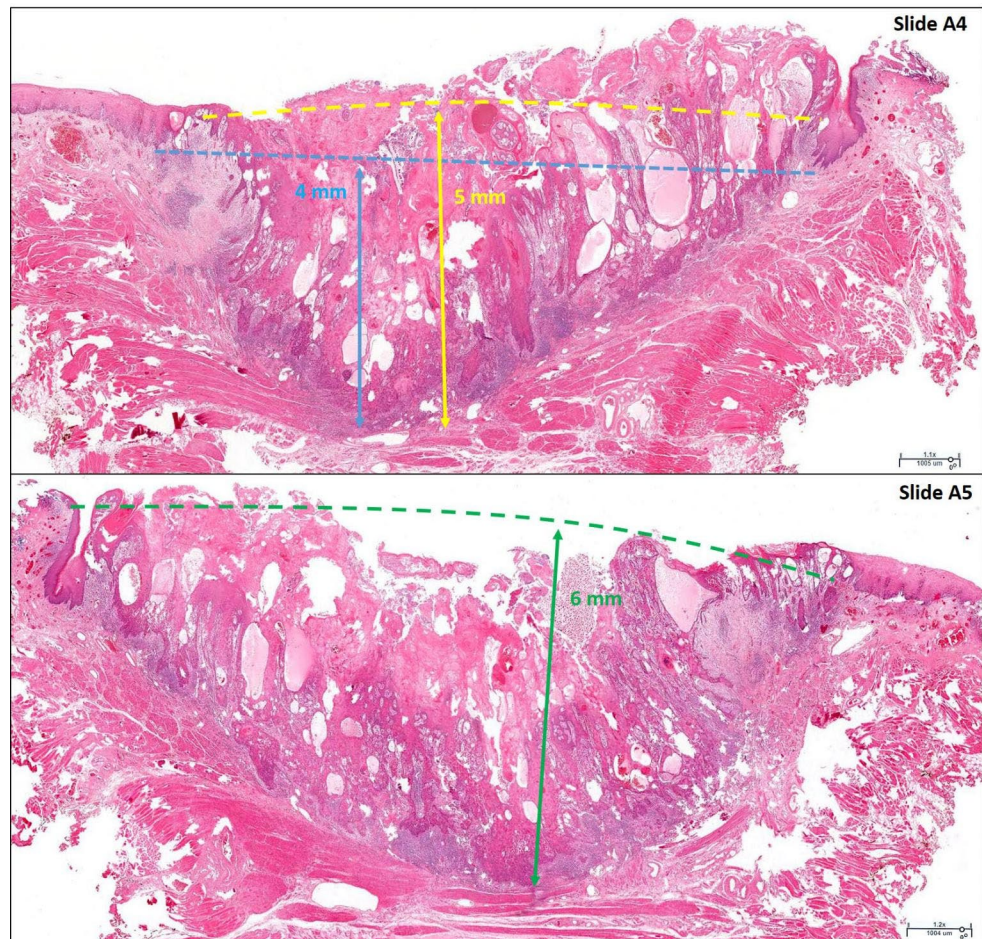


Table 3 Summary of cases with differences in pT

Observers	2 vs. 2	4
	1 vs. 3	21
Cut-point of disagreement	5 mm	4
	10 mm	21
pT stage difference	pT1–pT2	5
	pT2–pT3	14
	pT3–pT4	6

Explanatory notes: Observers refers to the number of observers with disagreement in pT stage (i.e. 2 vs. 2 indicates an even split with 2 providing one pT stage and the other 2 providing another pT stage)

Cut-point of disagreement refers to the nearest 5 or 10 mm measurement that resulted in different pT stages being provided by various observers

pT stage difference indicates the stages that various observers determined on cases. No case had more than a single pT stage discrepancy

slide in measuring DOI, ICC was compared between cases in which the same slide was used and those in which different slides were used. These calculations demonstrated higher ICC (0.98277) when the same slide was used compared to those with different slides (0.87415).

Discussion

The 2017 release of AJCC8 was notable for, among other things, recognizing the importance of DOI as a prognostic indicator for carcinomas of the oral cavity by incorporating this as a parameter for pT staging. Multiple studies have provided evidence to support the correlation between tumor depth (i.e. DOI) and risk of nodal metastasis, local recurrence, and survival but the methods for how to best obtain the DOI measurement have been varied [6, 7, 11]. In addition to the importance of DOI in tumor staging, DOI is also touted by some authors as a useful feature in deciding to perform elective neck dissection and/or the extent of nodal dissection [12, 13]. These prognostic features of DOI highlight the importance in accurate and reliable DOI measurements amongst individual pathologists.

In an attempt to standardize DOI measurement, and thus improve interobserver reproducibility in staging, AJCC8 included a detailed definition of DOI. AJCC8 defined DOI as the deepest depth of tumor measured from the nearest adjacent uninvolved epithelial basement membrane. From this surface reference point, a perpendicular plumb line is dropped to the tumor's deepest point of invasion. This

Table 4 Intraclass correlation coefficient scores for comparison of outcomes among 4 pathologists

	ICC score	ICC group [10]
DOI for all subsites (n = 660 observations)	0.91339	Excellent
DOI for all subsites when same slide used for DOI (n = 274 observations)	0.98277	Excellent
DOI for all subsites when different slide used for DOI (n = 386 observations)	0.87415	Good

measurement is used in conjunction with the tumor's greatest overall dimension, usually measured grossly, to assign a pathologic tumor stage.

Although the process for DOI measurement seems straightforward, several authors have noted challenges in determining accurate DOI. Berdugo et al. detail several difficulties in determining DOI including a lack of residual carcinoma on the resection, carcinoma extending to the deep margin and extratumoral perineural invasion/angiolymphatic invasion [8]. Another study on archival samples by Kukreja et al. also noted challenges in determining the proper starting point for measuring DOI (the so called "horizon" in their article) [9]. Difficulty in determining the horizon was seen in the majority of cases (78.9%) and was related to an absence of adjacent normal mucosa, the rounded or angulated nature of the mucosal surface and/or convoluted nature of endophytic tumors. We informally encountered many of the same issues in the present study, with curved, ulcerated, and irregular surfaces. Of note, extratumor foci of perineural invasion and angiolymphatic invasion were not noted in our series. Despite these challenges and the potential ambiguity of DOI measurements, Kukreja et al. confirmed that DOI greater than 5 mm predicted worse disease free survival while TT did not.

The current study did not aim to identify all the challenges in measuring DOI but rather to determine the reliability and correlation of these measurements between observers. Our data show that despite a high level of overall correlation in DOI measurements (i.e. excellent ICC), individual observer DOI differences have the potential to influence pT staging, particularly when the measurement is near the 5 or 10 mm thresholds. AJCC8 guidelines utilize depths of 5 and 10 mm as breakpoints in staging and disagreements in DOI around these values has the potential to lead to different pT staging. As larger tumors had higher variability in DOI, it is not surprising that most cases (21 of 25) with pT disagreement were due to differences in DOI measurements surrounding the 10 mm cut-off important for staging of tumors greater than 2 cm in overall dimension.

Our data indicate that DOI and pT differences were more common with larger tumors and those with higher average DOI. In contrast, ulceration and exophytic tumor growth were not significantly correlated with greater differences in

DOI or pT staging. Interestingly, the observer's choice of slide for DOI measurement was important in DOI variability. We found significantly higher variability in DOI and pT when each observer used different slides compared to cases in which the same slide was used by all observers. This study provided all tumor slides to the observers and allowed each to select the slide they felt provided the most accurate DOI as would be the situation when reporting a case in practice. This finding suggests that the choice of slide used for DOI measurement is an important factor in variable DOI and is an important point to resolve when attempting consensus in a group.

The challenges in determining DOI have been highlighted by various authors and some have suggested that measurements of TT are more straightforward. A recent study by Salama et al. found close correlation in DOI and TT overall but also suggests that TT may be a better predictor of tumor behavior [14]. Future studies could examine variability in TT as well and if there is less variability, TT may in fact make for a better measure in tumor staging and for predicting behavior.

In summary, our findings show that despite the overall high correlation in DOI, differences resulted in potential changes to pT in 15% of cases. Given the importance of pathologic T staging, we therefore recommend that cases be shared and discussed within a group to reduce this variability and to develop consensus in these measurements. This is particularly true for cases in which the DOI is close to a cut-point that might change the pT stage (i.e. 5 or 10 mm). Consensus on the slide used for the most accurate DOI measurement likely represents an important step in this process.

Authors' contributions: All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Mary Allen-Proctor, Mobeen Rahman, Chandana Reddy, Deborah Chute and Christopher Griffith. The first draft of the manuscript was written by Kathy Allen-Proctor and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding No funds, grants, or other support was received.

Availability of Data and Material Excel files.

Code Availability Not applicable.

Declarations

Conflict of Interest The authors have no relevant financial or non-financial interests to disclose.

Ethics Approval Cleveland Clinic IRB approval CC 00192.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

References

1. Lydiatt WM, Patel SG, O'Sullivan B, Brandwein MS, Ridge JA, Migliacci JC, et al. Head and neck cancers-major changes in the American Joint Committee on cancer eighth edition cancer staging manual. *CA Cancer J Clin American Cancer Society*. 2017;67:122–37.
2. Amin MD, Edge S, Greene F, editors. *AJCC Cancer Staging Manual*. 8th ed. New York: Springer; 2017.
3. Spiro RH, Huvos AG, Wong GY, Spiro JD, Gnecco CA, Strong EW. Predictive value of tumor thickness in squamous carcinoma confined to the tongue and floor of the mouth. *Am J Surg Am J Surg*. 1986;152:345–50.
4. Shim SJ, Cha J, Koom WS, Kim GE, Lee CG, Choi EC, et al. Clinical outcomes for T1-2N0-1oral tongue cancer patients underwent surgery with and without postoperative radiotherapy. *Radiat Oncol. Radiat Oncol*; 2010;5 PMID 20504371.
5. Ebrahimi A, Gil Z, Amit M, Yen TC, Liao CT, Chaturvedi P, et al. Primary tumor staging for oral cancer and a proposed modification incorporating depth of invasion: An international multicenter retrospective study. *JAMA Otolaryngol - Head Neck Surg. American Medical Association*; 2014. pp. 1138–48.
6. Fukano H, Matsuura H, Hasegawa Y, Nakamura S. Depth of invasion as a predictive factor for cervical lymph node metastasis in tongue carcinoma. 19: *Head Neck*. John Wiley and Sons Inc.; 1997. pp. 205–10.
7. Al-Rajhi N, Khafaga Y, El-Husseiny J, Saleem M, Mourad W, Al-Otieschan A, et al. Early stage carcinoma of oral tongue: Prognostic factors for local control and survival. *Oral Oncol Oral Oncol*. 2000;36:508–14.
8. Berdugo J, Thompson LDR, Purgina B, Sturgis CD, Tuluc M, Seethala R, et al. Measuring Depth of Invasion in Early Squamous Cell Carcinoma of the Oral Tongue: Positive Deep Margin, Extratumoral Perineural Invasion, and Other Challenges. *Head Neck Pathol. Humana Press Inc.*; 2019;13:pp. 154–61.
9. Kukreja P, Parekh D, Roy P. Practical Challenges in Measurement of Depth of Invasion in Oral Squamous Cell Carcinoma: Pictographical Documentation to Improve Consistency of Reporting per the AJCC 8th Edition Recommendations. *Head Neck Pathol Springer*. 2020;14:419–27.
10. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016 Jun;15(2):155–63.
11. Pentenero M, Gandolfo S, Carrozzo M. Importance of tumor thickness and depth of invasion in nodal involvement and prognosis of oral squamous cell carcinoma: A review of the literature. *Head Neck. Head Neck*; 2005;25:1080–91.
12. CGF van L, MAJ de R YPK. H M, I TH, JA H, et al. Depth of invasion in early stage oral cavity squamous cell carcinoma: The optimal cut-off value for elective neck dissection. *Oral Oncol. Oral Oncol*; 2020. p. 111.
13. R K, K D. Do ipsilateral lateralized squamous cell carcinomas of oral tongue with a depth of invasion (DOI) greater than 10 mm warrant a bilateral neck dissection to prevent future regional recurrences on contralateral side? A continuing conundrum. *Oral Oncol. Oral Oncol*; 2021;120.
14. AM S, C V, A NKAK, DK Y. Z, et al. Depth of invasion versus tumour thickness in early oral tongue squamous cell carcinoma: which measurement is the most practical and predictive of outcome? *Histopathology. Histopathology*. 2021;79:325–37.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.