



Inter- and Intrarater Reliability and Agreement Among Danish Head and Neck Pathologists Assessing Extranodal Extension in Lymph Node Metastases from Oropharyngeal Squamous Cell Carcinomas

Chadi Nimeh Abdel-Halim^{1,2} · Max Rohde^{1,2} · Stine Rosenkilde Larsen³ · Tina Marie Green^{2,3} · Benedicte Parm Uihøi⁴ · Nina Claire Woller⁵ · Oke Gerke^{2,6} · Poul Flemming Høilund-Carlsen⁶ · Jens Ahm Sørensen^{2,7,8} · Christian Godballe^{1,2}

Received: 2 May 2022 / Accepted: 15 June 2022 / Published online: 13 July 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Background Extranodal extension (ENE) in lymph node metastases is one of the most important prognostic factors in head and neck squamous cell carcinomas. Studies have shown inconsistency among pathologists in the assessment of ENE. The aims of this study were: (1) to determine the interrater and intrarater reliability and agreement in the assessment of ENE among Danish pathologists and (2) to test if a standardized assessment method may increase interrater agreement.

Methods Four Danish head and neck pathologists assessed ENE presence or absence in 120 histological slides from lymph nodes with oropharyngeal squamous cell carcinoma metastases (first round). Subsequently, guidelines were introduced to the pathologists and a new assessment was performed (second round). Finally, two of the pathologists assessed the slides to determine intrarater reliability and agreement (third round).

Results Interrater kappa coefficients varied between 0.57 and 0.67 in the first round and between 0.59 and 0.72 in the second round. The intrarater agreement between round 2 and 3 was 0.88 for pathologist 1 and 0.92 for pathologist 2 with resulting kappa coefficients of 0.76 (95% CI 0.64–0.88) and 0.84 (95% CI 0.74–0.94), respectively.

Conclusion We found a moderate level of reliability and agreement among pathologists for ENE in lymph node metastases from oropharyngeal squamous cell carcinomas. The intrarater reliability and agreement was generally higher than interrater measures. Interrater agreement was slightly improved by standardized assessment.

Keywords Head and neck cancer · Oropharynx · Squamous cell carcinoma · Extranodal extension · Lymph node metastases · Reliability · Histopathology · Interobserver

Introduction

Head and neck cancer is the 7th most frequent cancer in the world [1]. The vast majority is squamous cell carcinomas (SCC) in the pharynx, larynx or oral cavity, which are

commonly associated with risk factors such as smoking, alcohol, and age [2]. Cervical lymph node metastases are frequent and have adverse prognostic significance for the patients [3, 4].

✉ Chadi Nimeh Abdel-Halim
chadi.abdelhalim@rsyd.dk

¹ Department of ORL – Head & Neck Surgery and Audiology, Odense University Hospital, J. B. Winsløvs Vej 4, 5000 Odense, Denmark

² Department of Clinical Research, University of Southern Denmark, J.B. Winsløvs Vej 19.3, 5000 Odense, Denmark

³ Department of Pathology, Odense University Hospital, J. B. Winsløvs Vej 15, 5000 Odense, Denmark

⁴ Department of Pathology, Aarhus University Hospital, Palle Juul Jensens Boulevard 99, 8200 Aarhus N, Denmark

⁵ Department of Pathology, Copenhagen University Hospital, Inge Lehmanns Vej 14, Copenhagen, Denmark

⁶ Department of Nuclear Medicine, Odense University Hospital, J. B. Winsløvs Vej 4, 5000 Odense, Denmark

⁷ Research Unit for Plastic Surgery, Odense University Hospital, Odense, Denmark

⁸ University of Southern Denmark, Odense, Denmark

Extranodal extension (ENE) is the growth of lymph node metastasis beyond the lymph node capsule and is of critical importance for correct management of patients with head and neck SCC (HNSCC) [5–10]. Several studies have shown significantly worsened survival and increased relapse rates in HNSCC patients with ENE leading to the implementation of ENE in the latest TNM staging system in 2017 [11–13]. Subsequently, ENE is currently considered an indication for adjuvant chemoradiotherapy and contraindicative for unimodal surgical treatment [14–16].

Despite of being of high clinical importance, pathological findings qualifying a diagnosis of ENE seem to be uncertain [17, 18]. Thus, the reported incidence of ENE in HNSCC lymph node metastases varies between 20 and 85% [19–21]. Reasons may be insufficient imaging technologies and variability in histopathological conclusions among pathologists [22]. A recent systematic review of the literature identified 44 unique definitions of ENE, elucidating the lack of a succinct and generally accepted definition [23]. Obviously, this raises concerns about the reliability of the histopathological diagnosis of ENE and the successive treatment planning.

The aims of this study were: (1) to determine the interrater and intrarater reliability and agreement in the histopathological assessment of ENE among Danish pathologists and (2) to test if introduction of a standardized assessment method may increase interrater agreement.

Materials and Methods

This study was performed as a prospective reliability and agreement study in accordance with the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) [24]. Legal approval from the Research Board of Odense University Hospital (OUH) was obtained before the start of the study.

Sample Size

The sample size estimation was based on assumptions made for the proportion of agreement between every pair of two pathologists. An 80% minimum proportion of agreement (the null hypothesis) with a precision of 0.1 on each side was assumed. With a prevalence of ENE around 63%, the number of histological slides required for the study was calculated to 123 at a significance level of 5% and a power of 80%. Calculations were made according to the methods described by Hong et al. [25].

Study Sample

One-hundred-twenty-three histological slides were identified through a combined search in the electronic patient

journal system at the Department of ORL—Head & Neck Surgery and Audiology, OUH, Denmark and the local database at the Department of Pathology, OUH. We searched for patients from the time period 1st of November 2014 to 31st of October 2019, classified as having oropharyngeal squamous cell carcinoma (OSCC) with the code DC102, DC103, DC108, DC109 or DC090-DC099 according to the International Classification of Diseases version 10 (ICD-10). Subsequently, we performed a search in the pathology database from the same period to identify lymph nodes from the neck with squamous cell carcinoma coded T081* or T082* and M807* or M8083* excluding P31060 (fine needle aspirations) according to the systematized nomenclature of medicine (SNOMED). Finally, we were able to match the results from the two searches to identify patients with OSCC, who had surgically removed metastatic lymph nodes. From these patients, we identified all slides with lymph node metastases and extracted a consecutive list with the 123 newest slides. Slides with no evidence of lymph node metastasis were excluded.

Demographic and health data on the included patients were collected from electronic medical reports. Variables collected were birthdate, gender, diagnosis, p16 status and TNM classification.

All lymph nodes for evaluation had been mounted as histological slides, stained with standard hematoxylin and eosin. The slides were then digitally scanned at $\times 40$ magnification using Hamamatsu Nano Zoomer S360. For this study, the samples were retrieved as digital images from the local database. All images were anonymized before they were saved on an encrypted external hard drive, which was physically sent by mail to the pathologists in the assessment team.

Human Papilloma Virus (HPV) association was determined by immunohistochemical staining for p16 as part of the diagnostic work-up.

Pathologic Evaluation

Four specialized head and neck pathologists from the three largest head and neck cancer centers in Denmark were invited to participate in the study. One (NCW) from the University Hospital of Copenhagen, one from Aarhus University Hospital (BPU) and two from our own center at OUH (SRL & TMG).

The assessment of the slides was done using NanoZoomer Digital Pathology View v. 2.7.52 (© Hamamatsu Photonics K.K.). Each digitalized slide was evaluated twice by all pathologists through round 1 and round 2. Additionally, two of the pathologists evaluated the slides in a third round for intrarater reliability and agreement. To reduce recall bias, repeated evaluations of the histological slides were separated by a minimum interval of three months. Also, the sequence of the slides was randomized for every round. After each

round the pathologists handed in the external hard drive and did not receive it again before the beginning of the next round. The pathologists were blinded to clinical information, previous evaluations, and the ratings of the other participating pathologists. However, they were informed about the title and the purpose of the study, thus knowing they were rating cervical lymph node metastases from patients with OSCC in order to estimate reliability and agreement parameters. Ratings were conducted independently, and the pathologists were specifically instructed not to communicate with each other about assessment results during the study.

In the first round, the pathologists reviewed each slide to determine the presence or absence of ENE based on immediate impression and clinical routine. Additionally, they were asked to grade each slide from 0 to 4 according to the classification system of ENE presented by Lewis et al. [26].

In the second round, the pathologists reviewed each slide again for assessment of ENE and grading. However, this time they were instructed to follow our proposed uniform definition of ENE when determining the presence and grade of ENE: “*Squamous cell carcinoma within the confinement of a metastatic lymph node that grows through the lymph node capsule or beyond the lymph node contour into the adjacent tissue regardless of the size of the extension. In case of doubt, the presence of a desmoplastic response confirms the suspicion of ENE*” [23]. In addition, they were introduced to the latest guidelines from the International Collaboration on Cancer Reporting (ICCR) and asked to refer to these in case of doubt [27].

In the third round, two pathologists repeated the exact same procedure with the same instructions as in the second round.

Statistics

To examine the reliability and absolute agreement within and between raters, kappa coefficients and proportions of agreement were calculated [24]. Cohen’s kappa coefficient was calculated for interrater reliability between every two raters in each round and for intrarater reliability between round 2 and 3 for each pathologist. For the overall reliability among all pathologists, Fleiss’ multi-rater kappa coefficient was calculated for each round. The proportion of agreement was reported as the count of observed agreements between pathologists divided with the total number of evaluated slides. In addition, weighted kappa coefficients were calculated in the analyses of ENE grading reliability between two raters. We defined our own weighting matrix, giving full credit to absolute agreement, 0.75 credit to disagreement between grade 0 and 1 and between grade 2 and 3, half credit to disagreement between grade 2–3 and 4, and finally no credit to disagreement between grade 0–1 and grade 2–4. Exploratory hypothesis testing

of bivariate associations of categorical variables was done with Fisher’s exact test, and the significance level was 5% (two-sided). All analyses were done using Stata/BE v17.0 (StataCorp LLC, College Station, Texas, USA). Graphics were done with Microsoft Excel 2016.

Results

Study Sample

Of the original 123 slides, one was excluded due to no evidence of carcinoma in the lymph node, whereas two others were excluded because they appeared to originate from mucosal biopsies rather than lymph nodes. In total, 120 histological slides of separate cervical lymph nodes with metastatic squamous cell carcinoma from 54 OSCC patients were included. The median number of slides per patient was two (range 1–8). The average age was 62.2 and more than 2/3 of the patients were men. Two of the patients were originally diagnosed with cancer of unknown primary tumor (CUP) but relapsed at their oropharyngeal T-site half a year and three years later, respectively. There was a majority of p16+ cancers and most of the patients had low T-, N-, and M-stages. See Table 1 for detailed patient characteristics.

Table 1 Patient characteristics

Total, n	Patients	54
Age, mean \pm SD	Years	62.2 \pm 10.1
Sex, n (%)	Male	39 (72%)
	Female	15 (28%)
Cancer diagnosis, n (%)	Oropharynx	52 (96%)
	CUP	2 (4%)
p16 Status, n (%)	Positive	38 (70%)
	Negative	16 (30%)
T-stage, n (%)	T0	15 (28%)
	T1	28 (52%)
	T2	9 (17%)
	T3	1 (2%)
	T4	1 (2%)
N-stage, n (%)	N1	31 (57%)
	N2	18 (33%)
	N3	5 (9%)
M-stage, n (%)	M0	54 (100%)
	M1	0 (0%)
Recurrent cancer, n (%)	No	39 (72%)
	Yes	15 (28%)

n count, CUP cancer of unknown primary tumor

Raters

Four pathologists from three different hospitals across Denmark rated the same 120 digitalized histological slides. Two of the pathologists from the same hospital (Odense University Hospital) had three replicate observations each (round 1 to 3), whereas the other two pathologists evaluated the slides only through the first two rounds. One of the pathologists did not report data for one of the slides in the first round and therefore only had 119 observations in that round. All pathologists were consultants with median of 8 years of head and neck pathology experience ranging from 2 to 15 years.

Interrater Reliability and Agreement

In the first round, the overall proportion of observed agreement for the histopathological diagnosis of ENE between the four pathologists was 0.66 corresponding to a kappa coefficient of 0.61 (95% CI 0.51–0.71). In the second round, the proportion of agreement increased to 0.72 and a kappa

coefficient of 0.68 (95% CI 0.58–0.77). The number of observed lymph nodes with ENE ranged between 45 (38%) and 55 (46%) in the first round and 36 (30%) and 51 (43%) in the second round. In Table 2, the total count of observed ENE is stated for each pathologist and round. The kappa coefficients between two pathologists ranged from 0.57 (95% CI 0.42–0.72 between pathologist 1 and 4) to 0.66 (95% CI 0.53–0.80 between pathologist 1 and 3) in the first round and 0.59 (95% CI 0.45–0.74 between pathologist 3 and 4) to 0.78 (95% CI 0.66–0.89 between pathologist 2 and 4) in the second round. The overall difference between round 1 and 2 in kappa coefficients and proportions of agreement was 0.066 and 5.30% points, respectively. See Fig. 1 for the boxplot of kappa coefficients in the two rounds and Supplementary Table 1 for a detailed overview of reliability estimates and agreement measures for the histopathological diagnosis of ENE.

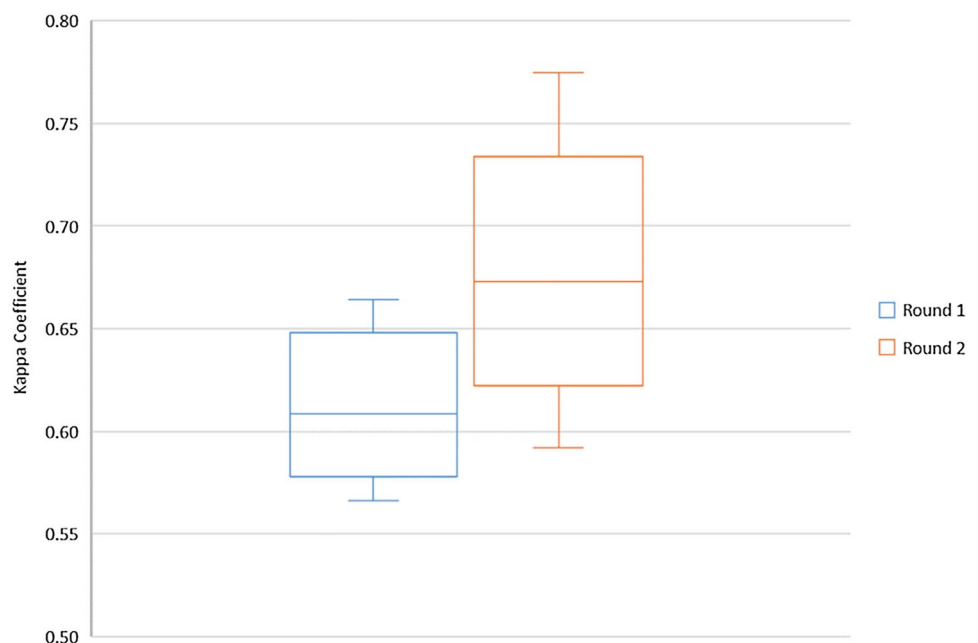
The overall proportion of observed agreement for ENE grading between the four pathologists was 0.24 in the first round corresponding to a kappa coefficient of 0.35 (95% CI 0.32–0.39). In the second round, after the introduction of our proposed definition of ENE, the proportion of agreement increased to 0.30 with a kappa coefficient of 0.41 (95% CI 0.34–0.44). The distribution of grades for each pathologist is illustrated in Fig. 2. The kappa coefficients between two pathologists ranged from 0.28 (95% CI 0.21–0.29 between pathologist 2 and 3) to 0.42 (95% CI 0.38–0.47 between pathologist 1 and 2) in the first round and 0.31 (95% CI 0.28–0.36 between pathologist 1 and 3) to 0.51 (95% CI 0.45–0.60 between pathologist 3 and 4) in the second round. The overall difference between round 1 and 2 in kappa coefficients and proportions of agreement was 0.060 and 5.6%

Table 2 Total count of ENE

Pathologist	Round 1	Round 2	Round 3
1	46/120 (38%)	48/120 (40%)	48/120 (40%)
2	55/120 (46%)	46/120 (38%)	47/120 (39%)
3	45/120 (38%)	36/120 (30%)	–
4	51/119 (43%)	51/120 (43%)	–

The total count of observed ENE through round 1–3 for each pathologist. Data is reported as number of histological slides with ENE divided with total number of evaluated histological slides

Fig. 1 Boxplot of interrater reliability measures for presence/absence of ENE between every two pathologists in round one and two. The plotted kappa coefficients are between pathologist 1 and 2, 1 and 3, 1 and 4, 2 and 3, 2 and 4, and 3 and 4. The rectangles represent the second and third quartiles, the horizontal line inside represents the median, and the horizontal lines outside represent the maximum and minimum



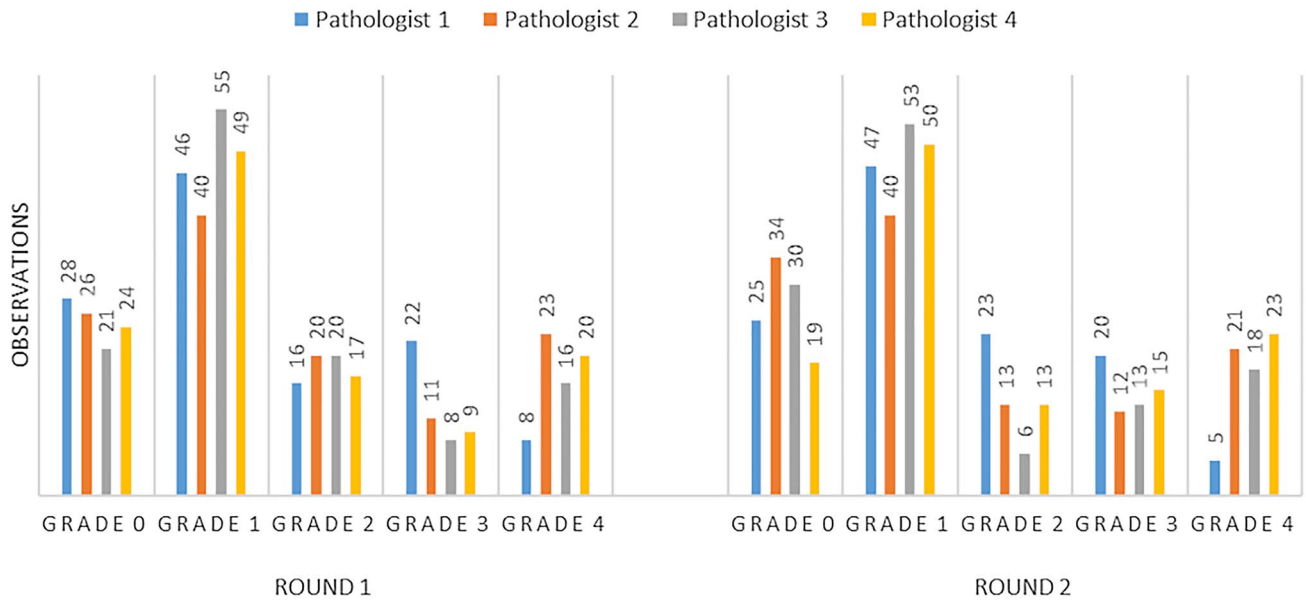


Fig. 2 Bar chart with the count of observed histological slides graded from 0 to 4 according to Lewis' ENE classification system

points, respectively. The weighted kappa coefficients ranged from 0.46 (95% CI 0.38–0.56 between pathologist 1 and 4) to 0.57 (95% CI 0.50–0.59 between pathologist 2 and 4) in the first round and 0.53 (95% CI 0.46–0.59 between pathologist 1 and 3) to 0.66 (95% CI 0.58–0.75 between pathologist 2 and 4) in the second round. See Supplementary Table 2 and 3 for all reliability estimates and agreement measures for ENE grading.

Interrater Agreement and p16 Status

Eighty-nine histological slides from 38 patients with p16+ cancers were included in the study. A median of 30 (range 25–37) (33%) were diagnosed with ENE in the first round and 28 (range 21–31) (31%) in the second round. The proportion of ENE in histological slides with p16- cancer was 61% in both rounds. There was a higher proportion of agreement for p16- (81%) than p16+ (61%) cancers in the first round with a p-value of 0.076. However, this changed in the second round to almost no correlation with p16 status with a p-value of 0.49. See Table 3 for agreement and p16 status.

Interrater Agreement in Primary Versus Recurrent Cancers

No difference was detected in the proportion of agreement between histological slides from primary and recurrent cancers. P-values were calculated to 0.82 and 0.99 for respectively round 1 and 2.

Table 3 Agreement according to p16 status

Agreement	p16-	p16+	Total	<i>p</i> -value
Round 1				
Yes	25 (81%)	54 (61%)	79 (66%)	
No	6 (19%)	34 (39%)	40 (34%)	
Total	31 (100%)	88 (100%)	119 (100%)	<i>0.076</i>
Round 2				
Yes	24 (77%)	62 (70%)	86 (72%)	
No	7 (23%)	27 (30%)	34 (28%)	
Total	31 (100%)	89 (100%)	120 (100%)	<i>0.49</i>

Statistically significant values are given in italics ($p > 0.05$)

Contingency table with the frequency distribution of agreement and p16 status in round 1 and 2

P-values are calculated with Fisher's exact test (two-sided)

Intrater Reliability and Agreement

Two pathologists evaluated the same histological slides in a third round. The proportion of observed agreement for the histopathological diagnosis of ENE between round 2 and 3 was 0.88 for pathologist 1 and 0.93 for pathologist 2. This resulted in the kappa coefficients 0.76 (95% CI 0.64–0.88) and 0.84 (95% CI 0.74–0.94).

The proportion of observed agreement for grading of ENE between round 2 and 3 was 0.72 for pathologist 1 and 0.81 for pathologist 2 with resulting kappa coefficients of 0.61 (95% CI 0.56–0.68) and 0.75 (95% CI 0.65–0.86), respectively.

Calculating the weighted kappa coefficients, agreement proportions changed to 0.84 for pathologist 1 and 0.89 for pathologist 2 with the weighted kappa coefficients of 0.71 (95% CI 0.62–0.81) and 0.80 (95% CI 0.75–0.81).

Discussion

To our knowledge, this study is the most extensive inter- and intrarater study on ENE in HNSCC. It showed a moderate level of reliability and agreement among Danish pathologists in the assessment of histopathological ENE in lymph node metastases from OSCC and only a minor trend towards improvement after introduction of standardized methods for evaluation. These findings indicate significant inconsistencies in the histopathological diagnosis of ENE and call for methods and definitions, which may improve diagnostic certainty.

In advanced and extensive cases, ENE may be detectable by clinical examination or imaging. The diagnostic accuracy of different imaging technologies has been evaluated in several studies and found to be rather low [22]. Thus, neck dissection with histopathological assessment of extirpated lymph nodes including evaluation of ENE is still considered as the gold standard in cancer staging.

Few studies regarding diagnostic certainty of ENE have been performed. In 2012, van den Brekel et al. examined the observer variation in ENE diagnosis among 10 pathologists [17]. Forty-one metastatic lymph nodes from 18 HNSCC patients were evaluated in two rounds. They found poor interrater agreement with overall kappa values of 0.42 and 0.49. Kappa coefficients between two pathologists ranged from 0.14 to 0.73 in the first round and from 0.20 to 0.75 in the second round. The study did not include information regarding p16/HPV status. In 2015, Lewis et al. showed an interobserver agreement of 48% with a kappa-value of 0.51 (95% CI 0.36–0.64) among five pathologists rating ENE in node metastases [18]. The study included 50 histological slides from 50 patients with p16+ OSCC. Interestingly, in a second assessment round, they introduced a self-developed grading system from zero to two based on the degree of tumor invasion into the perinodal tissue. Dichotomizing the grades to \pm ENE, they achieved a better agreement ratio at 68% corresponding to a kappa value of 0.64 (95% CI 0.47–0.78). The kappa values between two pathologists ranged between 0.36 and 0.72 in the first round and 0.51 and 0.75 in the second round.

Our study included fewer raters but a larger study sample with both p16+ and p16- OSCC. Moreover, the raters were asked both to assess the presence of ENE and to apply the non-validated grading system, developed by Lewis et al. in 2011 [26]. The generally higher concordance in our study may reflect the beneficial value of a grading system as a

supportive tool. Nevertheless, we found a lower agreement using the classification system by Lewis et al. compared to a simple dichotomous assessment of \pm ENE. This result is not surprising, as the classification system allows five different outcomes and thereby a higher risk of inconsistency. Instead of dichotomizing data, we found it more relevant to do an analysis of weighted kappa coefficients categorizing grade 0–1 (ENE not present) as similar, but not equal, and the same for grade 2–4 (ENE present). The analysis showed considerably higher levels of concordance, supporting that much of the discrepancy was caused by disagreements within similar ENE categories.

Interestingly, our results showed a tendency towards a difference in disagreement between p16+ and p16- tumors. The difference was not statistically significant ($p=0.076$). However, looking at the first round, we found an overall disagreement of 39% in assessment of metastases from p16+ carcinomas vs. 19% for p16-, indicating higher difficulties in evaluation of p16+ lymph node metastasis. An explanation may be that lymph nodes with metastasis from p16+ OSCC often grow in a ballooning/cystic manner, where the capsule is stretched and thickened (Fig. 3). Some pathologists may consider this being ENE caused by a breach in the capsule and fibrotic reaction to tumor growth in the perinodal tissue with the formation of a so-called pseudo-capsule [28]. However, other pathologists would say that the metastasis is within the confinement of the lymph node and therefore not indicative of ENE. Although current studies have shown lack of correlation between ENE in p16+ lymph nodes and prognosis, pathologists should still report ENE in these cases for educational and research purposes [29–31]. Our study supports this, since disagreement on the histopathological diagnosis potentially could be an explanation for the lack

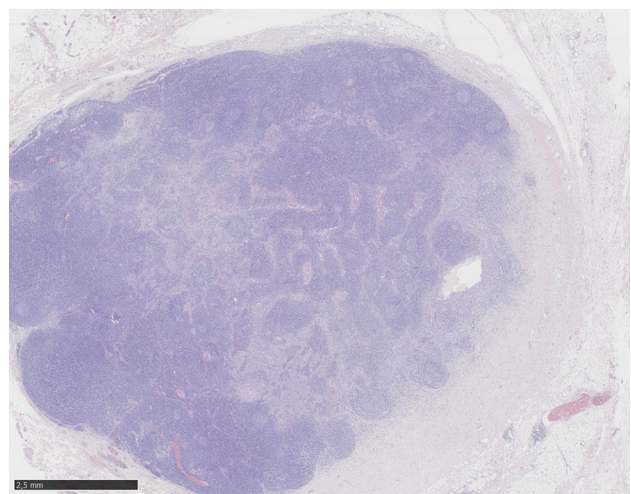


Fig. 3 Lymph node with metastasis from p16+ oropharyngeal squamous cell cancer and pseudocapsule. According to our definition, this finding does not qualify as ENE. (hematoxylin–eosin)

of association with worsened prognosis in this cohort of patients.

A major reason for the variation in the diagnosis of ENE among pathologists may be the lack of consensus on a histopathological definition. We saw a trend towards an increase in agreement and a decrease in the fraction of lymph nodes with ENE after introducing the pathologists to a uniform definition of ENE. The higher reliability was also evident in the intrarater kappa coefficients for the two pathologists that completed a third evaluation round. Based on our results and the current literature, we recommend pathologists to discuss and agree on a consensus definition of ENE, preferably at an international level.

However, we also acknowledge the fact that a written definition cannot stand alone, since pathologists could interpret it differently. This was reflected in the higher spread of kappa coefficients between every two pathologists in the second round compared to the first round (Fig. 1). The more cautious ENE determination in the second round may be due to the provided guideline, which in some cases recommended a more conservative approach.

Limitations

A limitation to the study was the analysis of reliability and agreement based on one single histological slide for each outcome. In the daily clinical setting, pathologists would in cases of doubt, examine additional sections and levels and sometimes even do further analyses to make their final diagnosis. In suboptimal tissue sections where part of the lymph node capsule is not present, the assessment is hampered, and the pathology report would include a qualitative description stating the diagnostic uncertainty. This was not possible in our study since we demanded a definitive binary *yes* or *no* answer.

Another limitation was the use of scanned histology images. At the time of the study, digitalized slides were not used in the daily routine work in all institutions across the country. For that reason, the assessment of the slides did not reflect the daily routine for all the pathologists. However, digitalized slides were easier to distribute among the geographically separated pathologists. In return, we were able to assemble a team of pathologists with many years of head and neck experience. Moreover, the risk of the pathologists aligning their results was lower; we were able to randomize the order of histological slides for each round; and we were in better control of the minimum interval of time between rounds, minimizing recall bias even more.

Intrarater reliability was calculated based on only two of the pathologists, since the other two pathologists did not participate in the third round. This was a compromise between logistics and importance of calculating intrarater

kappa coefficient for every pathologist. We do not believe, the inclusion of the other pathologists would have changed our results.

In conclusion, ENE diagnosis in lymph node metastases from OSCC is a significant challenge even for highly specialized pathologists with many years of experience. It is very important for researchers, clinicians, and patients to be aware of the limitations in the pathologic assessment of ENE as shown in this study, since it may have huge impact on the successive management of the patients. In the era of de-escalation therapy for p16+ OSCC with ENE, these concerns must be taken into consideration. [32–34].

Conclusion

The current study showed moderate level of reliability and agreement among Danish pathologists in the assessment of histopathological ENE in lymph node metastases from OSCC. The intrarater reliability and agreement was generally higher than between different pathologists. Interrater reliability increased slightly after the introduction of defined guidelines and a new standardized ENE definition. We conclude that an agreed-upon ENE definition is helpful but cannot stand alone in increasing the interrater agreement.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12105-022-01468-z>.

Acknowledgements This work was supported by the University of Southern Denmark; the Region of Southern Denmark; Odense University Hospital; and the Harboe fund. The sponsors had no involvement in study design, data collection, analysis, interpretation of data, writing of the report, or decision to submit the article for publication.

Author Contributions Study concepts and design: All authors. Data acquisition: CAH, SRL, TMG, BPU, NCW. Quality control of data and algorithms: CAH, MR, SRL, TMG, BPU, NCW, CG. Data Analysis and interpretation: CAH, MR, SRL, TMG, BPU, NCW, OG, CG. Statistical analysis: CAH, MR, OG, CG. Manuscript preparation, editing, and review: All authors.

Funding This work was supported by the University of Southern Denmark; the Region of Southern Denmark; Odense University Hospital; and the Harboe fund. The sponsors had no involvement in study design, data collection, analysis, interpretation of data, writing of the report, or decision to submit the article for publication.

Data Availability Data has not been published elsewhere whole or part. All data and materials comply with fields standards.

Code Availability Additional data can be found in the attached supplementary material.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval Legal approval from the Research Board of Odense University Hospital was obtained before the start of the study. The study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

References

- Chow LQM. Head and neck cancer. *N Engl J Med.* 2020;382:60–72.
- Argiris A, Karamouzis MV, Raben D, Ferris RL. Head and neck cancer. *Lancet.* 2008;371:1695–709.
- Woolgar JA. Histopathological prognosticators in oral and oropharyngeal squamous cell carcinoma. *Oral Oncol.* 2006;42:229–39.
- Tsai TY, Iandelli A, Marchi F, Huang Y, Tai SF, Hung SY, et al. The prognostic value of lymph node burden in oral cavity cancer: systematic review and meta-analysis. *Laryngoscope.* 2021;132:88–95.
- Pilborough AE, Lambert DW, Khurram SA. Extranodal extension in oral cancer: a role for the nodal microenvironment? *J Oral Pathol Med.* 2019;48:863–70.
- Majercakova K, Valero C, Lopez M, Garcia J, Farre N, Quer M, et al. Postoperative staging of the neck dissection using extracapsular spread and lymph node ratio as prognostic factors in HPV-negative head and neck squamous cell carcinoma patients. *Oral Oncol.* 2018;77:37–42.
- Hararah MK, Stokes WA, Jones BL, Oweida A, Ding D, McDermott J, et al. Nomogram for preoperative prediction of nodal extracapsular extension or positive surgical margins in oropharyngeal squamous cell carcinoma. *Oral Oncol.* 2018;83:73–80.
- Shevach J, Bossert A, Bakst RL, Liu J, Misiukiewicz K, Beyda J, et al. Extracapsular extension is associated with worse distant control and progression-free survival in patients with lymph node-positive human papillomavirus-related oropharyngeal carcinoma. *Oral Oncol.* 2017;74:56–61.
- Mermod M, Tolstong G, Simon C, Monnier Y. Extracapsular spread in head and neck squamous cell carcinoma: a systematic review and meta-analysis. *Oral Oncol.* 2016;62:60–71.
- Bhattacharya P, Mukherjee R. Lymph node extracapsular extension as a marker of aggressive phenotype: classification, prognosis and associated molecular biomarkers. *Eur J Surg Oncol.* 2021;47:721–31.
- Bennett SH, Futrell JW, Roth JA, Hoye RC, Ketcham AS. Prognostic significance of histologic host response in cancer of the larynx or hypopharynx. *Cancer.* 1971;28:1255–65.
- Dunne AA, Muller HH, Eisele DW, Kessel K, Moll R, Werner JA. Meta-analysis of the prognostic significance of perinodal spread in head and neck squamous cell carcinomas (HNSCC) patients. *Eur J Cancer.* 2006;42:1863–8.
- Huang SH, O'Sullivan B. Overview of the 8th edition TNM classification for head and neck cancer. *Curr Treat Opt Oncol.* 2017;18:40.
- Bernier J, Cooper JS, Pajak TF, van Glabbeke M, Bourhis J, Forastiere A, et al. Defining risk levels in locally advanced head and neck cancers: a comparative analysis of concurrent postoperative radiation plus chemotherapy trials of the EORTC (#22931) and RTOG (# 9501). *Head Neck.* 2005;27:843–50.
- Bernier J, Domette C, Ozsahin M, Matuszewska K, Lefebvre JL, Greiner RH, et al. Postoperative irradiation with or without concomitant chemotherapy for locally advanced head and neck cancer. *N Engl J Med.* 2004;350:1945–52.
- Cooper JS, Pajak TF, Forastiere AA, Jacobs J, Campbell BH, Saxman SB, et al. Postoperative concurrent radiotherapy and chemotherapy for high-risk squamous-cell carcinoma of the head and neck. *N Engl J Med.* 2004;350:1937–44.
- van den Brekel MW, Lodder WL, Stel HV, Bloemena E, Lee-mans CR, van der Waal I. Observer variation in the histopathologic assessment of extranodal tumor spread in lymph node metastases in the neck. *Head Neck.* 2012;34:840–5.
- Lewis JS Jr, Tarabishy Y, Luo J, Mani H, Bishop JA, Leon ME, et al. Inter- and intra-observer variability in the classification of extracapsular extension in p16 positive oropharyngeal squamous cell carcinoma nodal metastases. *Oral Oncol.* 2015;51:985–90.
- Coatesworth AP, MacLennan K. Squamous cell carcinoma of the upper aerodigestive tract: the prevalence of microscopic extracapsular spread and soft tissue deposits in the clinically N0 neck. *Head Neck.* 2002;24:258–61.
- Carter RL, Barr LC, O'Brien CJ, Soo KC, Shaw HJ. Transcapsular spread of metastatic squamous cell carcinoma from cervical lymph nodes. *Am J Surg.* 1985;150:495–9.
- Brannan AG, Johnstone PA, Cooper J. Extracapsular tumor extension in cervical lymph nodes: reconciling the literature and seer data. *Head Neck.* 2011;33:525–8.
- Abdel-Halim CN, Rosenberg T, Dyrvig AK, Hoiland-Carlsen PF, Sorensen JA, Rohde M, et al. Diagnostic accuracy of imaging modalities in detection of histopathological extranodal extension: a systematic review and meta-analysis. *Oral Oncol.* 2021;114: 105169.
- Abdel-Halim CN, Rosenberg T, Larsen SR, Hoiland-Carlsen PF, Sorensen JA, Rohde M, et al. Histopathological definitions of extranodal extension: a systematic review. *Head Neck Pathol.* 2021;15:599–607.
- Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol.* 2011;64:96–106.
- Hong H, Choi Y, Hahn S, Park SK, Park BJ. Nomogram for sample size calculation on a straightforward basis for the kappa statistic. *Ann Epidemiol.* 2014;24:673–80.
- Lewis JS Jr, Carpenter DH, Thorstad WL, Zhang Q, Haughey BH. Extracapsular extension is a poor predictor of disease recurrence in surgically treated oropharyngeal squamous cell carcinoma. *Mod Pathol.* 2011;24:1413–20.
- Bullock MJ, Beitler JJ, Carlson DL, Fonseca I, Hunt JL, Katabi N, et al. Data set for the reporting of nodal excisions and neck dissection specimens for head and neck tumors: explanations and recommendations of the guidelines from the international collaboration on cancer reporting. *Arch Pathol Lab Med.* 2019;143:452–62.
- Sinha P, Lewis JS Jr, Piccirillo JF, Kallogjeri D, Haughey BH. Extracapsular spread and adjuvant therapy in human papillomavirus-related, p16-positive oropharyngeal carcinoma. *Cancer.* 2012;118:3519–30.
- Maxwell JH, Ferris RL, Gooding W, Cunningham D, Mehta V, Kim S, et al. Extracapsular spread in head and neck carcinoma: impact of site and human papillomavirus status. *Cancer.* 2013;119:3302–8.

30. Klozar J, Koslabova E, Kratochvil V, Salakova M, Tachezy R. Nodal status is not a prognostic factor in patients with HPV-positive oral/oropharyngeal tumors. *J Surg Oncol*. 2013;107:625–33.
31. Geiger JL, Lazim AF, Walsh FJ, Foote RL, Moore EJ, Okuno SH, et al. Adjuvant chemoradiation therapy with high-dose versus weekly cisplatin for resected, locally-advanced HPV/p16-positive and negative head and neck squamous cell carcinoma. *Oral Oncol*. 2014;50:311–8.
32. Cmelak AJ, Ferris RL, Chen AM, Seiwert T, Burtness B. Treatment de-intensification for HPV-positive oropharynx cancer: what is currently acceptable? *J Clin Oncol*. 2021;39:2732–3.
33. Ma DJ, Van Abel KM. Treatment de-intensification for HPV-associated oropharyngeal cancer: a definitive surgery paradigm. *Semin Radiat Oncol*. 2021;31:332–8.
34. Moore EJ, Van Abel KM, Routman DM, Lohse CM, Price KAR, Neben-Wittich M, et al. Human papillomavirus oropharynx carcinoma: aggressive de-escalation of adjuvant therapy. *Head Neck*. 2021;43:229–37.

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