Interrater variation of vascular classifications used in enhanced laryngeal contact endoscopy

Mehlum, Camilla Slot; Døssing, Helle; Davaris, Nikolaos; Giers, Anja; Grøntved, Ågot Møller; Kjaergaard, Thomas; Möller, Sören; Godballe, Christian; Arens, Christoph

Published in:
European Archives of Oto-Rhino-Laryngology

DOI:
10.1007/s00405-020-06000-z

Publication date:
2020

Document version:
Accepted manuscript

Citation for pulished version (APA):
Mehlum, C. S., Døssing, H., Davaris, N., Giers, A., Grøntved, Å. M., Kjaergaard, T., Möller, S., Godballe, C., & Arens, C. (2020). Interrater variation of vascular classifications used in enhanced laryngeal contact endoscopy. *European Archives of Oto-Rhino-Laryngology*, *277*(9), 2485-2492. https://doi.org/10.1007/s00405-020-06000-z

Reliability in Contact Endoscopy

Interrater Variation of Vascular Classifications Used in Enhanced Laryngeal Contact Endoscopy.

Camilla Slot Mehlum, MD (1); Helle Døssing, MD, PhD(1); Nikolaos Davaris, MD (2); Anja Giers, MD(2); Ågot Møller Grøntved, MD (1); Thomas Kjaergaard, MD, PhD (3); Sören Möller, MSc, PhD (4); Christian Godballe, MD, PhD (1); Christoph Arens, MD, PhD(2).

(1) Department of ORL Head & Neck Surgery and Audiology, Odense University Hospital, J. B. Winsløwsvej 4, 5000 Odense, Denmark

(2) Department of Otorhinolaryngology, Head and Neck Surgery, Magdeburg University Hospital, Otto-von-Guericke University, Germany.

(3) Department of Otorhinolaryngology‐ Head and Neck Surgery, Aarhus University Hospital, Palle Juul‐ Jensens Boulevard 165, 8200 Aarhus N, Denmark

(4) OPEN ‐ Open Patient Data Explorative Network and Department of Clinical Research, Odense University Hospital and University of Southern Denmark, J. B. Winsløwsvej 9, 5000 Odense, Denmark

Corresponding author: Camilla Slot Mehlum, Department of ORL Head & Neck Surgery and Audiology, Odense University Hospital, J. B. Winsløwsvej 4, 5000 Odense, Denmark. E-mail Camilla.mehlum@rsyd.dk. ORCHID ID 0000-0002-5740-612X

Disclosures: none

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

Funding information: This study was supported by funds from the Toyota Foundation, which enabled the purchase of contact endoscopes for research use.

Ethical approval: All procedures performed involving human participants were in accordance with the ethical standards of the University Hospital Magdeburg, Germany, and Odense University Hospital, Denmark at which the study were conducted. Informed consent was obtained from all participants included in the study from University Hospital Magdeburg, Germany. According to Danish legislation patient consent was not required for the participants included in the study from Odense University Hospital, Denmark as no identifying information about participants was investigated in the study. The study was approved by the Danish Data Protection Agency, the local hospital management and the legal department at Odense University Hospital (permit number 20/15307).

Interrater Variation of Vascular Classifications Used in Enhanced Laryngeal Contact Endoscopy.

ABSTRACT

Purpose

Combined use of contact endoscopy (CE) and Narrow Band Imaging (NBI) is suggested for visualization of specific vascular changes indicative of glottic neoplasia. We investigated the interrater reliability and agreement in 3 recognized classification systems of vascular changes applied to images from CE + NBI in patients suspected for glottic neoplasia.

Methods

Six experienced head and neck surgeons familiar with NBI rated 120 images obtained by CE + NBI by 3 classification systems of vascular changes as suggested by Ni et al. (N-C), Puxeddu et al. (P-C), and the European Laryngological Society (ELS-C). Three raters were experienced in CE, and three raters had only limited experience with CE. Crude agreement and Fleiss´ kappa with 95% confidence interval were estimated for all 6 raters, and for the 2 levels of expertise for each original classification system and for dichotomized versions of the N-C and the P-C based on suggested neoplastic potential.

Results

The interrater crude agreement and the corresponding kappa values for the ELS-C were good and significantly higher than those for the N-C and P-C for all raters, irrespective of the level of experience with CE ($p<0.0001$). There were no significant differences between the N-C and the P-C ($p=0.16$). Kappa was considerably improved for both the N-C and the P-C to a level not different from the ELS-C ($p=0.21$-0.71) when their 5 original categories were pooled into dichotomized classifications.

Conclusion

Difficulties in reliably classifying vascular changes in CE + NBI are evident. Two-tier classification systems are the most reliable.

Keywords

Contact endoscopy, Narrow Band Imaging, enhanced endoscopy, reliability, larynx, glottic.

INTRODUCTION

The superficial glottic vasculature is clearly visualized by enhanced endoscopy (EE), which is a common term for endoscopic imaging techniques designed to visualize the vasculature by enhancing hemoglobin. Assessment of the vascular patterns by use of EE is now considered an integral part of the evaluation of glottic lesions[1, 2]. The most frequently used equipment for EE is probably the Narrow Band Imaging (NBI, Olympus®).

Specific vascular patterns have proved to be useful indicators of cancer[3, 4] and in particular vessels arising from the deep layers of the mucosa as intraepithelial papillary capillary loops (IPCLs) are stimulated by neo-angiogenetic factors in tumor growth[1]. The vasculature has been characterized in classification systems with varying degrees of complexity and number of categories. In 2016 the European Laryngological Society (ELS) proposed a simplified 2-tier classification (hereafter "ELS-guideline")[1] as an alternative to other, more complex systems such as those proposed by Ni et al. in 2011 (hereafter "Ni-classification")[4] and Puxeddu et al. in 2015 (hereafter "Puxeddu-classification")[5]. The vascular characteristics applied in these 3 classification systems are summarized in Table 1. The Ni-classification and the Puxeddu-classification were originally established to predict laryngeal cancer[4, 5], whereas the ELS-guideline was established to distinguish benign, non-neoplastic lesions from premalignant or malignant (neoplastic) glottic lesions[1]. The term glottic "neoplasia" is used as a common term for cancer and precursor lesions throughout this paper as it may be more relevant to differentiate neoplastic lesions from non-neoplastic lesions, than to differentiate early cancer from precursor lesions, in contemporary planning of treatment[6-10].

The glottic mucosa can be viewed in detail by the use of contact endoscopy (CE) which provides images of cells and vascular structures in the superficial layers of the mucosa with 60x or 150x magnification. When the technique was first adapted from gynecology for laryngological use by Andrea et al[11, 12], the mucosa was stained with 1% methylene blue before endoscopy. The procedure was cumbersome and time-consuming and it required extensive experience in histological assessment to be able to interpret the pictures taken[13, 14] as the focus was largely on cell changes in the mucosa. Vascular changes were occasionally described, but their significance was uncertain[15-17]. Several authors suggested that there was high diagnostic accuracy of CE in detecting premalignant and malignant laryngeal lesions[13, 14, 17-21], but the technique was not widely used. In 2003, Arens et al. described the use of CE enhanced by autofluorescence as Compact Endoscopy[20], and in 2006 the same authors published spectrometric measurements in Compact endoscopy performed without staining the mucosa[22]. The authors described that vessels were better contrasted to the surrounding tissue and concluded that Compact endoscopy was a useful complementary tool in the detection and delineation of laryngeal cancer and premalignant lesions. The combined use of CE and EE without tissue staining was presented in 2014 by Arens et al. during the 10th congress of the ELS in Antalya, Türkey. The technique was subsequently reported almost simultaneously in the beginning of 2015 by Puxeddu et al.[5] and by Arens et al.[23, 24] who described specific microvascular irregularities and visible capillary loops in areas with dysplasia and increasing vascular alterations proportional to the severity of dysplasia. The CE technique applied with EE was soon elaborated upon in the ELS-guideline[1] and other reports with promising results for assessment of larynx[25, 26], oral cavity and oropharynx[27]. However, an element of subjectivity in the interpretation of vascular patterns in CE has been described**[28]** and it is largely unknown how reliable available vascular classification systems are when applied to images obtained by CE. We thus aimed to investigate the interrater reliability and agreement in 3 classification systems of vascular changes applied to images from CE + NBI in patients suspected for glottic neoplasia. Based on ratings from 6 experienced head and neck surgeons, with varying levels of expertise in CE, we compared the calculated reliability and agreement for each of the classification systems and for subgroups of raters. The rated images originated from patients with a spectrum of non-neoplasia, precursor lesions and cancer, and thus resembled a clinical situation. The diagnostic accuracy of CE + NBI was outside the scope of this present paper and the ratings were not assessed for diagnostic accuracy as the images had not necessarily originated from the most severely affected part of the glottic lesion.

METHODS

This article was prepared according to the Guidelines for Reporting Reliability and Agreement Studies[29]. We chose a sample size of 120 images, well above the number suggested by Hong et al. for sample size in kappa statistics[30].

The images used for this study originated from endoscopic procedures performed in 2 university hospital departments of head and neck surgery (departments of Otorhinolaryngology at the University Hospital Magdeburg, Germany, and Odense University Hospital, Denmark) between

June 15th and November 15th, 2019. A total of 120 images were obtained by CE (KARL STORZ, Tuttlingen, Germany) with EE (NBI®, Olympus, Japan) from 15 patients, with 5-15 images from each patient and half of the images from each institution. The patients were treated by endoscopic cordectomy for suspected glottic neoplasia and the final histopathology reported 3 cases of non-neoplasia, 7 precursor lesions and 5 carcinomas. The images were pseudonymized (by C.S.M.) before the rating. The 6 raters were all experienced head and neck surgeons at university hospitals. Every rater evaluated all of the 120 images in terms of vascular changes and scored each image by the 3 classification systems described in Table 1. Three of the raters were very familiar with the use of CE and NBI (C.A., A.G., N.D.; hereafter "experts"), whereas the 3 remaining raters (C.G., H.D., C.S.M.) were very familiar with the use of NBI but had only limited experience in the interpretation of CE (hereafter "beginners").

The ratings were conducted independently at 2 different hospitals. The raters were allowed to view the images as many times and for as long as they needed, but were told not to discuss their ratings before the results had been sent for further processing. Individual assessments, as opposed to consensus ratings, were chosen in order most to resemble the daily clinical routine in hospital departments. The raters were blinded to the final histological diagnoses and their distribution, and the assessment of the other raters, but were aware that their assessments were intended for an interrater reliability study. The raters were provided with pre-designed tables for use during the rating with four columns for each image, one column for each classification system. The categories in each classification were numbered above the columns (Ni-classification type I-V; Puxeddu classification type 0-IV; ELS-guideline perpendicular vessels yes/no). In the last column comments could be added. For uniformity, all of the raters were provided with the original papers in which the classifications were first described[1, 4, 5] along with an instructional PowerPoint® presentation of the classifications to accompany the electronic images. In case of doubt, the raters were instructed to choose "worst-case scenario" for each image.

The degree of interrater reliability and agreement, kappa statistics were estimated using Stata 16 (StataCorp LP, College Station, TX). The crude agreement was calculated as the number of times that absolute agreement occurred. The corresponding calculated unweighted Fleiss´ kappa value reflected the chance-corrected agreement. Crude agreement and Fleiss´ kappa with 95% confidence interval (CI) was estimated for each classification system for all 6 raters, and for the 2 subgroups of "experts" and "beginners". Furthermore, we estimated crude agreement and Fleiss´ kappa of the ratings when categories were pooled into vascular changes indicative of neoplasia or not according to the original papers (Table 1). Hence, we introduced binary categories for the Puxeddu-classification (type 0-II non-neoplastic, type III-IV neoplastic) and for the Ni-classification (type I-III non-neoplastic, IV-V: neoplastic). In addition, we tested for differences in kappa between the classification systems and between beginners and experts using the asymptotic normality of kappa statistics.

RESULTS

The results of the individual ratings according to the Ni-classification, Puxeddu-classification and ELS-guideline are provided in Table 2. The estimates of interrater reliability and agreement, with 95% CI for each classification system, are presented in Table 3. The interrater crude agreement and the corresponding kappa values for the ELS-guideline were significantly higher than those for the original Ni- and Puxeddu-classifications for all raters and for subgroups of experts and beginners ($p<0.0001$). There were no significant differences between the original Ni- and Puxeddu-classifications ($p=0.16$).

In the calculations based on binary categories for the Puxeddu-classification (type 0-II non-neoplastic, type III-IV neoplastic) and for the Ni-classification (type I-III non-neoplastic, IV-V: neoplastic) the crude agreement for all 3 classification systems were almost identical, as were the kappa values for the Ni-classification and ELS-guideline, whereas kappa was slightly lower for the Puxeddu-classification in the subgroup of beginners, which reduced the overall kappa. Kappa was considerably improved for both the Ni- and the Puxeddu classifications to a level not different from the ELS-guideline ($p=0.21$ - 0.71) when their 5 original categories were pooled into a dichotomized classification.

No significant differences in reliability were detected between experts and beginners, except for the expert ratings by the Ni-classification for which kappa was significantly higher than for the beginners' ratings ($p<0.05$).

DISCUSSION

The kappa values found in our study indicate an inability of the original Ni-classification and the original Puxeddu-classification to make clear distinctions between adjacent categories of the vasculature in CE + NBI, resulting in difficulties for raters to provide reliable and reproducible categorizations. There were no significant differences between the 2 systems (p=0.16). The ELS-guideline provided considerably more reliable ratings (p<0.0001). This is relevant for clinicians if they are to rely on CE + NBI to determine patient treatment. The higher crude interrater agreement and kappa value for the ELS-guideline reflects a higher degree of agreement among raters. However, it is well known that the reliability of ratings is improved by use of fewer categories, since kappa values depend on the prevalence of ratings in each category, and a greater number of categories will tend to lower the kappa value[31]. Thus, in the modified 2-tier ratings by the Ni-classification, and the Puxeddu-classification, the agreement and corresponding kappa values increased significantly, and to the same level as the ELS-guideline (p=0.21 - 0.71). The expert ratings by the Ni-classification resulted in significantly higher kappa than for the beginners' ratings (p<0.05). For the remaining ratings, no significant differences in reliability were detected between experts and beginners.

A commonly used reliability grading proposed by Landis[32] and modified by Altman[33] defined kappa values <0.20 as poor agreement, 0.21 to 0.40 as fair agreement, 0.41 to 0.60 as moderate agreement, 0.61 to 0.80 as good agreement, and 0.81 to 1 as very good agreement. On this scale, the kappa estimates for the ELS-guideline and the modified 2-tier versions of the Ni- and Puxeddu-classifications in our study indicated good interrater agreement in all subgroups analyzed except for beginners' ratings by Puxeddu (2-tier) that only indicated moderate agreement. The kappa estimates for the original Ni- and Puxeddu classifications indicated fair agreement, except for the expert ratings by Puxeddu-classification that indicated poor agreement. However, these labels are arbitrary and do not necessarily indicate the clinical relevance of the results as even a "very good reliability" result can be associated with clinically unacceptable discordance between raters, depending on the consequences of the result [29](e.g. if one of the categories requires specific handling).

Estimates of reliability of CE has only rarely been reported. Warnecke et al.[13] reported the assessment of images from CE with white light and methylene blue staining of 42 lesions in larynx, oropharynx and hypopharynx (32 benign, 10 malignant). The authors described a high interrater reliability (kappa 0.81) when 3 assessors rated the cytological features in the images into a dichotomized classification of benign or malignant. The interrater reliability of CE with white light and methylene blue staining of nasopharyngeal mucosa has also been investigated. In images from 34 previously irradiated and 20 non-irradiated patients, Pak et al.[34] reported an overall kappa value of 0.78 resp. 0.86, irrespective of the variable level of expertise among the 3 raters and suggested CE to be used for localizing suspicious sites and thus identify where to biopsy in the case of suspected nasopharyngeal carcinoma.

Despite the fact that EE with NBI and similar methods has been used for the assessment of head and neck for a number of years, there are only a few descriptions of the reliability. Schossee et al.[35] reported kappa values of 0.23-0.59 when 84 NBI images of vocal folds were presented to 3 consultants for classification of 6 types of longitudinal vascular changes, with only slight differences between white light and NBI. In a recent study by Ni et al[36], white light images and NBI images of 30 cases of glottic leukoplakia were presented to 8 experienced laryngologists and 8 less experienced laryngologists. The images were rated by all the participants before and after an hour- long NBI training course that comprised basic principle of NBI technology, the Ni-classification, and a suggested, new diagnostic classification for glottic leukoplakia[37]. The kappa values for experienced/less experienced/overall raters were for white light 0.38/0.21/0.29, for NBI before training 0.51/0.50/0.50, and for NBI after training 0.65/0.58/0.62. Hence, NBI resulted in significantly higher kappa than white light alone, and a very short, specific training course seemed to further enhance the reliability of NBI in glottic leukoplakia.

As far as we know, only one very recent paper by Davaris et al. has reported the reliability of CE + NBI[38]. In their study, 3 experienced otolaryngology specialists and 3 less-experienced otolaryngology residents assessed potential perpendicular vasculature according to the ELS-guideline in CE + NBI images from 68 patients with various laryngeal lesions. The overall agreement among all 6 raters was substantial (Fleiss´ kappa 0.70 (95% CI 0.64-0.76)) which is very similar to our findings. However, they found significant difference between the calculated Fleiss´ kappa values of the two groups of raters as the level of agreement was moderate for the residents (Fleiss´ kappa 0.51 (95% CI 0.37-0.65)) and almost perfect for the specialists (Fleiss´ kappa 0.92 (95% CI 0.78 – 1.01)). In this present study, no significant differences in reliability were detected between experts and beginners. These slightly contradictory results may be related to the fact that the “beginners” in the present study were actually experienced otolaryngologists, familiar with NBI and the classification of vasculature (and only “beginners” in terms of CE) as opposed to residents who are expectedly less experienced in several areas.

Although none of the above papers are directly comparable to the results in the present paper, they do, overall, point to a non-negligible interrater variation in the assessment of glottic lesions with CE, NBI and CE + NBI. As expected, when using 2-tier categorization, the highest degree of agreement and credibility is achieved. Hence, in our study the highest reliability of CE + NBI was achieved for the 2-tier ELS-guideline and comparable reliability was achieved for the modified Ni- and Puxeddu classification when the ratings were pooled into a dichotomized classification indicative of neoplasia or non-neoplasia. However, clinicians dealing with patients with glottic lesions should be critical in interpreting CE with or without the aid of EE, as even experienced laryngologists can experience difficulties in providing reliable diagnosis.

Strengths and Limitations

To our knowledge, this is the first published study to provide comparable estimates of interrater assessment for the Ni-classification, Puxeddu-classification and ELS-guideline of vascular changes visualized in CE + NBI. The study was designed and conducted with the specific intention of assessing reliability and agreement in vascular classification of CE. The fully crossed design improved the validity of the reliability estimates. Individual assessment of images was chosen to most resemble clinical practice. On the other hand, the use of still images and not video sequences, and the option for the assessors to repeatedly assess the images was not completely comparable to clinical reality.

The raters were experienced head and neck surgeons with extensive laryngological expertise and experience with NBI and comparable imaging techniques. Hence the results may not be representative of laryngologists in general. All of the raters were less familiar with the Puxeddu-classification, than with the other 2 classification systems, which may have influenced the comparisons.

CONCLUSION

Assessment with CE + NBI provides visualization of vascular changes in glottic lesions that may indicate neoplasia and thus assists the clinician in making a correct diagnosis before treatment. However, the results of the current study indicate difficulties in reliably classifying vascular changes in CE + NBI by the Ni-classification[4] and the Puxeddu-classification[5] when used in their original 5-tier version, even for very experienced head and neck surgeons. This is important knowledge from both clinical and scientific perspectives, for instance in future automated evaluation of NBI-CE using machine learning algorithms[28]. When the ratings of the 2 systems are pooled into 2-tier categories intended to distinguish neoplastic glottic lesions from non-neoplastic lesions, the reliability of both systems increase significantly to a level comparable to the reliability achieved with the ELS-guideline[1].

REFERENCES

1. Arens, C., et al., *Proposal for a descriptive guideline of vascular changes in lesions of the vocal folds by the committee on endoscopic laryngeal imaging of the European Laryngological Society.* Eur Arch Otorhinolaryngol, 2016. **273**(5): p. 1207-14.
2. Simo, R., et al., *European Laryngological Society: ELS recommendations for the follow-up of patients treated for laryngeal cancer.* Eur Arch Otorhinolaryngol, 2014. **271**(9): p. 2469-79.
3. Piazza, C., et al., *Narrow band imaging and high definition television in the assessment of laryngeal cancer: a prospective study on 279 patients.* Eur Arch Otorhinolaryngol, 2010. **267**(3): p. 409-14.
4. Ni, X.G., et al., *Endoscopic diagnosis of laryngeal cancer and precancerous lesions by narrow band imaging.* J Laryngol Otol, 2011. **125**(3): p. 288-96.
5. Puxeddu, R., et al., *Enhanced contact endoscopy for the detection of neoangiogenesis in tumors of the larynx and hypopharynx.* Laryngoscope, 2015. **125**(7): p. 1600-6.
6. Peretti, G., et al., *Pre- and intraoperative assessment of mid-cord erythroleukoplakias: A prospective study on 52 patients.* European Archives of Oto-Rhino-Laryngology, 2003. **260**(10): p. 525-528.
7. Cosway, B. and V. Paleri, *Laryngeal dysplasia: an evidence-based flowchart to guide management and follow up.* J Laryngol Otol, 2015. **129**(6): p. 598-9.
8. Cosway, B., M. Drinnan, and V. Paleri, *Narrow band imaging for the diagnosis of head and neck squamous cell carcinoma: A systematic review.* Head Neck, 2016. **38 Suppl 1**: p. E2358-67.
9. Rohde, M., et al., *Aggressive elimination of precancerous lesions of the vocal cords to avoid risk of cancer.* Dan Med J, 2012. **59**(5): p. A4399.
10. Trolle, W.T., JF; Charabi, B; Schytte, S; Kjaergaard, T; Ulhøj, BP; Lyhne, NM; Lambertsen, K; Groentved, AM; Mehlum, CS; Schultz, JH; Overgaard, J; Godballe, C; https://dahanca.dk/assets/files/Pro_VejledningDahanca27.pdf. *[Guidelines for management of laryngeal intraepithelial neoplasia (LIN) and T1A-glottic cancer]*. 2019 [cited 2019 June 11th].
11. Andrea, M., O. Dias, and A. Santos, *Contact endoscopy during microlaryngeal surgery: a new technique for endoscopic examination of the larynx.* Ann Otol Rhinol Laryngol, 1995. **104**(5): p. 333-9.
12. Andrea, M., O. Dias, and A. Santos, *Contact endoscopy of the vocal cord: normal and pathological patterns.* Acta Otolaryngol, 1995. **115**(2): p. 314-6.
13. Warnecke, A., et al., *Contact endoscopy for the evaluation of the pharyngeal and laryngeal mucosa.* Laryngoscope, 2010. **120**(2): p. 253-8.
14. Tarnawski, W., et al., *The role of computer-assisted analysis in the evaluation of nuclear characteristics for the diagnosis of precancerous and cancerous lesions by contact laryngoscopy.* Adv Med Sci, 2008. **53**(2): p. 221-7.
15. Sone, M., et al., *Vascular evaluation in laryngeal diseases: comparison between contact endoscopy and laser Doppler flowmetry.* Arch Otolaryngol Head Neck Surg, 2006. **132**(12): p. 1371-4.
16. Cikojevic, D., I. Gluncic, and V. Pesutic-Pisac, *Comparison of contact endoscopy and frozen section histopathology in the intra-operative diagnosis of laryngeal pathology.* J Laryngol Otol, 2008. **122**(8): p. 836-9.
17. Carriero, E., et al., *Preliminary experiences with contact endoscopy of the larynx.* Eur Arch Otorhinolaryngol, 2000. **257**(2): p. 68-71.
18. Wardrop, P.J., S. Sim, and K. McLaren, *Contact endoscopy of the larynx: a quantitative study.* J Laryngol Otol, 2000. **114**(6): p. 437-40.
19. Arens, C., et al., *[Endoscopic imaging techniques in the diagnosis of laryngeal carcinoma and its precursor lesions].* Laryngorhinootologie, 1999. **78**(12): p. 685-91.
20. Arens, C., et al., *Compact endoscopy of the larynx.* Ann Otol Rhinol Laryngol, 2003. **112**(2): p. 113-9.

21. Klancnik, M., I. Gluncic, and D. Cikojevic, *The role of contact endoscopy in screening for premalignant laryngeal lesions: a study of 141 patients.* Ear Nose Throat J, 2014. **93**(4-5): p. 177-80.
22. Arens, C., et al., *Spectrometric measurement in laryngeal cancer.* Eur Arch Otorhinolaryngol, 2006. **263**(11): p. 1001-7.
23. Arens, C. and S. Voigt-Zimmermann, *[Contact endoscopy of the vocal folds in combination with narrow band imaging (compact endoscopy].* Laryngorhinootologie, 2015. **94**(3): p. 150-2.
24. Arens, C., H. Glanz, and S. Voigt-Zimmermann, *Vascular lesions of vocal folds - Part 2: Perpendicular vascular lesions.* Laryngo- Rhino- Otologie, 2015. **94**(11): p. 738-744.
25. Arens, C., et al., *Narrow band imaging for early diagnosis of epithelial dysplasia and microinvasive tumors in the upper aerodigestive tract.* Hno, 2017. **65**(Suppl 1): p. 5-12.
26. Mannelli, G., L. Cecconi, and O. Gallo, *Laryngeal preneoplastic lesions and cancer: challenging diagnosis. Qualitative literature review and meta-analysis.* Crit Rev Oncol Hematol, 2016. **106**: p. 64-90.
27. Carta, F., et al., *Enhanced contact endoscopy for the assessment of the neoangiogenetic changes in precancerous and cancerous lesions of the oral cavity and oropharynx.* Eur Arch Otorhinolaryngol, 2016. **273**(7): p. 1895-903.
28. Esmaeili, N., et al., *Novel automated vessel pattern characterization of larynx contact endoscopic video images.* Int J Comput Assist Radiol Surg, 2019. **14**(10): p. 1751-1761.
29. Kottner, J., et al., *Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed.* J Clin Epidemiol, 2011. **64**(1): p. 96-106.
30. Hong, H., et al., *Nomogram for sample size calculation on a straightforward basis for the kappa statistic.* Ann Epidemiol, 2014. **24**(9): p. 673-80.
31. Jakobsson, U. and A. Westergren, *Statistical methods for assessing agreement for ordinal data.* Scand J Caring Sci, 2005. **19**(4): p. 427-31.
32. Landis, J.R. and G.G. Koch, *The measurement of observer agreement for categorical data.* Biometrics, 1977. **33**(1): p. 159-74.
33. Altman, D., *Practical statistics for medical research*. 1991, London: Chapman and Hall. 611.
34. Pak, M.W., et al., *How reliable is contact endoscopy of the nasopharynx in patients with nasopharyngeal cancer?* Laryngoscope, 2009. **119**(3): p. 523-7.
35. Schossee, A., et al., *[Evaluation of a Classification Model of Horizontal Vascular Lesions of the Vocal Folds].* Laryngorhinootologie, 2016. **95**(4): p. 245-50.
36. Ni, X.G., et al., *Clinical utility and effectiveness of a training program in the application of a new classification of narrow band imaging for vocal cord leukoplakia: a multicenter study.* Clin Otolaryngol, 2019. **[Epub ahead of print]**.
37. Ni, X.G., et al., *Diagnosis of vocal cord leukoplakia: The role of a novel narrow band imaging endoscopic classification.* Laryngoscope, 2018. **129**(2): p. 429-434.
38. Davaris, N., et al., *Evaluation of Vascular Patterns Using Contact Endoscopy and Narrow-Band Imaging (CE-NBI) for the Diagnosis of Vocal Fold Malignancy.* Cancers (Basel), 2020. **12**(1).

Table 1

Table 1. Characterization of vascular changes according to the classification systems suggested by Ni et al[4], Puxeddu et al.[5], and European Laryngological Society (ELS)[1]

	Vascular Characteristics	Interpretation *
Ni-classification		
I	Interconnected thin, oblique and arborescent vessels. IPCLs almost invisible.	Benign, non-neoplastic lesions
II	Oblique and arborescent vessels with enlarged diameter. IPCLs almost invisible.	Benign, non-neoplastic lesions
III	IPCLs obscured by white mucosa.	Benign, non-neoplastic lesions
IV	IPCLs recognized as small dots. Relatively regular distribution and low density.	Mild or moderate dysplasia
V	V_a: IPCLs solid or hollow, with a brownish, speckled pattern and various shapes. V_b: IPCLs irregular, tortuous, line-like shapes. V_c: IPCLs brownish speckles or tortuous, line-like shapes with irregular distribution.	Severe dysplasia, CIS or invasive carcinoma
Puxeddu-classification		
0	Thin-end regular, subepithelial vessels connecting with a thicker and deeper arborescent vascular network running parallel to the epithelium.	Normal mucosa
I	Increase in number and size of subepithelial vessels, with irregular and sometimes crossing directions.	Inflammation
II	Very thin and short IPCLs running toward the surface, scattered distribution.	Hyperplasia
III	Elongated small vessels in the typical “bobby-pin” shape, but some arborescence appears at the end of the IPCLs.	Mild or moderate dysplasia
IV	IPCLs significantly dilated, with various shapes and a wide range of vascular architectural changes such as corkscrews or tree-like patterns.	High-grade dysplasia, CIS or invasive carcinoma
ELS-guideline		
Longitudinal	Ectasia (Dilated vessels). Meander (Meandering, tortuous vessels). Varicose (Advanced meandering and dilated vessels). Convolute (Organized coil/tangle of vessels). Number of vessels (Increased vessels number). Branches of vessels (Increased branches of vessels). Change of direction (Abrupt change of vessels direction).	Benign, non-neoplastic
Perpendicular	Enlarged vessel loops (Abnormal IPCLs with wide-angled turning points). Dot-like vessel loops (Abnormal IPCLs with narrow-angled turning points). Worm-like vessels (Abnormal vessels with spiral morphology and bizarre course).	Premalignant or malignant

*Interpretation according to author of the original classification-system in question

IPCLs: Intraepithelial Papillary Capillary Loops. CIS: Carcinoma in Situ. Class: classification

Table 2

Table 2. Distribution of ratings by six raters using three separate classification systems for vascular changes visualized in contact endoscopy of glottic lesions.

	Experts			Beginners		
	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5	Rater 6
Ni-classification						
I	22 (18%)	24 (20%)	3 (3%)	58 (48%)	24 (20%)	25 (21%)
II	51 (43%)	29 (24%)	68 (57%)	2 (2%)	39 (33%)	34 (28%)
III	12 (10%)	9 (8%)	15 (13%)	17 (14%)	12 (10%)	13 (11%)
IV	15 (13%)	10 (8%)	0 (0%)	3 (3%)	26 (22%)	9 (8%)
V	20 (17%)	48 (40%)	34 (28%)	40 (33%)	19 (16%)	39 (33%)
Puxeddu-classification						
0	22 (18%)	42 (35%)	4 (3%)	26 (22%)	24 (20%)	31 (26%)
I	51 (43%)	20 (17%)	74 (62%)	53 (44%)	37 (31%)	42 (35%)
II	9 (8%)	23 (19%)	7 (6%)	21 (18%)	15 (13%)	4 (3%)
III	4 (3%)	33 (28%)	11 (9%)	19 (16%)	27 (23%)	12 (10%)
IV	34 (28%)	2 (2%)	24 (20%)	1 (1%)	17 (14%)	31 (26%)
ELS-guideline						
No perpendicular	85 (71%)	62 (52%)	84 (70%)	75 (63%)	73 (61%)	72 (60%)
Perpendicular	35 (29%)	58 (48%)	36 (30%)	45 (38%)	47 (39%)	48 (40%)

(%) percentages of ratings (n=120) for each classification system and each of the six raters
Raters 1-3 are considered experts and 4-6 beginners.

Table 3

Table 3. Interrater reliability and agreement estimates and statistical uncertainty in classification of vascular changes in Contact endoscopy

			All raters (n=6)		**Experts (n=3)**		**Beginners (n=3)**	
	N of grading categories	Categories	Crude agreement	Fleiss´ kappa (95 % CI)	Crude agreement	Fleiss´ kappa (95 % CI)	Crude agreement	Fleiss´ kappa (95 % CI)
Ni-classification	5	I-V	0.48	0.31 (0.26- 0.37)	0.56	0.38 (0.30-0.46)	0.40	0.22 (0.16-0.29)
	2	I-III / IV-V	0.87	0.73 (0.64-0.81)	0.87	0.71 (0.60-0.82)	0.89	0.78 (0.68-0.87)
Puxeddu-classification.	5	0-IV	0.45	0.27 (0.22- 0.31)	0.40	0.19 (0.13-0.26)	0.45	0.27 (0.20-0.35)
	2	0-II / III-IV	0.84	0.63 (0.54-0.72)	0.88	0.71 (0.60-0.82)	0.82	0.56 (0.44-0.68)
ELS-guideline	2	Perpendicular (+/-)	0.86	0.70 (0.63-0.78)	0.87	0.71 (0.60-0.82)	0.88	0.75 (0.66-0.85)