LARYNGOLOGY

Probe-based confocal laser endomicroscopy in detecting malignant lesions of vocal folds

Endomicroscopia confocale laser tramite sonda nel rilevamento delle lesioni maligne delle corde vocali

M. GONCALVES1, M. AUBREVILLE2, S.K. MUELLER1, M. MAIER2, H. IRO1, C. BOHR3

1 Department of Otorhinolaryngology, Head and Neck Surgery, Friedrich-Alexander-Universität Erlangen-Nürnberg, University Hospital, Erlangen, Germany; 2 Pattern Recognition Lab, Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany; 3 Department of Otorhinolaryngology, Head and Neck Surgery, Universität Regensburg, University Hospital, Regensburg, Germany

SUMMARY

Probe-based confocal laser endomicroscopy (CLE) is an innovative technique for real-time, non-invasive analysis of the surface epithelium. While being successfully used for diagnosis by experts, this method has not yet been established in clinical routine, partly due to the lack of standards and criteria for classifying various lesions. Our aim was to determine the diagnostic value and inter-rater reliability of CLE in detecting malignant lesions of the vocal cords. 58 video sequences were extracted from the probe-based CLE (GastroFlex probe with a Cellvizio® laser system) examinations of 3 patients with squamous cell carcinomas and 4 patients with benign alterations of the vocal folds. Two ENT surgeons, who were blinded to the histological result, were asked to identify the sequences representing a carcinoma. We showed an accuracy, sensitivity, specificity, PPV and NPV of 91.38-96.55%, 100%, 87.8-95.2%, 77.27-89.47% and 100%, respectively, with an inter-rater reliability of k = 0.89 (“almost perfect agreement”). Probe-based CLE is a promising method for diagnosis and assessment of vocal fold lesions in vivo. Our results suggest that, with adequate training, the diagnostic value of this technique can be improved and potentially provide important information during oncological surgery.

KEY WORDS: Confocal laser endomicroscopy • Laryngeal cancer • Non-invasive histological imaging • Vocal folds

Introdution

The vocal folds are the most frequent location of laryngeal cancer, accounting for more than two-thirds of all cases. More than 90% are classified as squamous cell carcinomas (SCC). The diagnosis is provided by biopsy and histopathological assessment. Leucoplaikia, erythroplakia and papillomatosis are the macroscopically visible changes from which SCC usually originate. Up to 50% of leukoplakias show no dysplasia or invasive carcinoma in subsequent histological examination. The unnecessary excision or biopsy of these lesions could have a negative impact on voice due to scarring of the vocal cords. A number of optical imaging methods, such as confocal laser endomicroscopy (CLE), narrow-band imaging, fluorescence endoscopy and optical coherence tomography (OCT) have been suggested as having the potential to im-
prove the laryngoscopic analysis of mucosal lesion with white light 8-13.

Probe-based CLE is a novel technique that enables the imaging of cell outlines at the surface of a lesion with a magnifying power of up to 1000. The method requires administration of fluorescein as a contrast agent, which accumulates in the intercellular spaces but not in the nuclei 14. As they differ from other imaging techniques, the acquired images require the clinician or pathologist to have special training 15. In the last few years, probe-based CLE has been intensively studied in gastroenterology and has expanded its application to other areas such as the head and neck, pulmonology and urology 16-21. The aim of this study was to assess the diagnostic value of probe-based CLE in identifying malignant lesions of the vocal cords in comparison to the accepted gold standard of histopathological examination.

Materials and methods

This research was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). The local ethics committee approved the study and all patients gave written informed consent. The study was carried out at the ENT Department of a tertiary level university hospital. Between July and October 2015, seven patients (three women, four men; average age 56.7 ± 5.8 years) underwent microlaryngoscopic examination under general anaesthesia. All patients had a suspected unilateral lesion of unknown nature in the vocal cords. Consequently, they all had an indication for this procedure. After documentation of findings under white light microscopy, 5 ml fluorescein (Fluorescein Alcon 10%, Alcon PHARMA GmbH, Freiburg, Germany) was administered intravenously and the vocal cords were scanned by probe-based CLE (GastroFlex probe with Cellvizio laser system, Mauna Technologies, Paris, France). The images were taken within five minutes of intravenous fluorescein injection, as the image quality was expected to deteriorate after eight minutes 22. The probe was placed on the vocal cords under direct vision. A biopsy was subsequently performed in the area of interest. The video recordings were analysed and compared with the histological results. 58 representative CLE video sequences (3,224 images) of healthy vocal cords, benign lesions (hyperplasia, hyperkeratosis, polyps, cysts) and malignant lesions were extracted and presented independently to two medical professionals (blinded examiners) for assessment. Of these 58 representative video sequences, 17 showed malignancies (29.3%) and 41 benign lesions or healthy normal mucosa (70.7%). The examiners were two ENT specialists who had undergone training and certification as provided online by Cellvizio 23. At the present time, there is no certification programme available for head and neck lesions, but there is one for lesions in the oesophagus. Since the epithelium of these regions is for the most part similar and the prevalence of squamous cell carcinomas is also comparable (over 90% for both), we used the training programme designed for the oesophagus to help us learn how to classify the lesions in the vocal cords 24. The blinded examiners had to identify video sequences showing malignancy. The histological findings were regarded as the reference standard for subsequent statistical analysis.

Results

The accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and accuracy were calculated with 95% confidence intervals for each examiner. Inter-rater reliability/agreement was tested statistically using Cohen’s kappa (Cohen’s kappa coefficient). The values were interpreted according to the widely accepted Landis and Koch classification (25). Agreement with values of χ between 0 and 0.20 was regarded as slight, between 0.21 and 0.40 adequate, between 0.41 and 0.60 moderate, between 0.61 and 0.8 substantial, and between 0.81 and 1.0 almost perfect. Statistical analysis was carried out using SPSS version 25.0 (IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 25.0. Armonk, New York, United States of America).
Confocal laser endomicroscopy in detecting malignant lesions of the vocal folds

91.38-96.55%, 100%, 87.8-95.2%, 77.27-89.47% and 100%, respectively, for the two examiners when identifying the video sequences with malignant alterations (Table I).

Inter-rater reliability was tested with Cohen’s kappa (k) and evaluated according to the Landis and Koch classification as well as Fleiss’s criteria. The two examiners obtained a k value of 0.89, which is to be interpreted as almost perfect or excellent agreement. Figure 1 shows a typical image of healthy mucosa and malignant changes. Of the 58 video sequences, 2 sequences representing healthy mucosa were incorrectly classified by both examiners independently as representing malignant changes (false positives, Fig. 2). Additionally, examiners E1 and E2 disagreed on three other sequences, representing false-positive findings for examiner E1 (Fig. 3).

Table I. Diagnostic metrics. Accuracy, sensitivity, specificity, value and negative, negative predictive value with corresponding 95% confidence intervals (95% CI).

<table>
<thead>
<tr>
<th></th>
<th>E1 (95% CI)</th>
<th>E2 (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>91.38% (81.02-97.14%)</td>
<td>96.55% (88.09-99.58%)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>100.00% (80.49-100.00%)</td>
<td>100.00% (80.49-100.00%)</td>
</tr>
<tr>
<td>Specificity</td>
<td>87.80% (73.80-95.92%)</td>
<td>95.12% (83.47-99.40%)</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>77.27% (59.93-88.55)</td>
<td>89.47% (68.75-97.05%)</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

E1: examiner 1; E2: examiner 2.

**Discussion**

In this study, probe-based CLE showed very good results in identifying malignant lesions. We obtained an accuracy, sensitivity, specificity, PPV, NPV of 91.38-96.55%, 100%, 87.8-95.2%, 77.27-89.47% and 100%, respectively. We also showed excellent inter-observer agreement, as suggested by Cohen’s kappa statistics. This represents an improvement in the results compared to a previous study of our group which suggested that examinations of the vocal cords would only show a fair agreement between observers. We attribute the improvement in the results to two methodical changes in our work. First, the examiners underwent certified training provided by the Cellvizio Academy in the interpretation of the CLE images and, most importantly, of the video sequences, as also suggested by Oetter et al. As the images obtained by this
technique differ from the other methods usually applied in daily routine, specific training is required.

The quality of the images was variable, as seen in Figures 2 and 3 depicting healthy mucosa. This should not, however, be regarded as a limitation of the study as it represents the expected set-up in the operating theatre. Some of these lesions have a friable surface, resulting in slight bleeding of the mucosa, which could impair the quality of the images. Additionally, the rugged surface of a tumour makes it more difficult to assure proper contact between the cell surface and the probe.

Even though our results suggest reliability and a high degree of certainty in identifying malignant lesions in the vocal folds, we did not demonstrate that CLE can add diagnostic value to microlaryngoscopic examination with white light, as there was already a high suspicion of malignancy. Further studies with a larger number of patients will have to be carried out to address this question. The three patients with SCC only underwent biopsy due to the suspicion of advanced glottic carcinoma and therefore there was no indication for cordectomy Type I using transoral laser microsurgery (TLM).

The penetration depth of probe-based (pCLE) is limited to about 60 µm, thus providing a two-dimensional visualisation of the most superficial layer of the lesion. Due to this fact, it is usually not possible to differentiate between in situ carcinoma and invasive carcinoma, since the stromal invasion cannot be demonstrated. This low penetration depth of the probe is possibly also the reason why benign

Fig. 2. Example of a false-positive finding. Both examiners independently assessed this sequence as representing malignant changes. Note the relatively poor quality of this CLE image with three capillaries filled with erythrocytes (frame 085, Fig. A). The second image represents a later part of the sequence (frame 119, Fig. B) showing many small, similar cells without clear aberrant pleomorphism.

Fig. 3. Example of an image, on which the examiners disagreed, representing a false positive of examiner E1. The quality of the image is also of poor quality, but the polygonal aspect suggests a benign lesion.
lesions of the vocal folds, such as polyps, cannot be adequately differentiated from healthy mucosa. The histopathological changes of these benign lesions are usually found in the lamina propria under a healthy superficial layer of epithelium. For this reason, we opted for a two-category question for this study on malignant/non-malignant changes. Additional information about deeper layers could possibly be provided by combining pCLE with other endoscopic techniques such as optic coherence tomography or narrow band imaging (NBI). NBI provides information about the examined areas by evaluating surrounding perpendicular and longitudinal vascular pattern changes. Newest reports on the diagnostic value of predicting malignancy using NBI show an accuracy, sensitivity, specificity, PPV, NPV of 96%, 100%, 95%, 88%, 100%, respectively, which appear very similar to our results. The assumption that the contralateral vocal fold, when inconspicuous to examination with white light during microlaryngoscopy, represents healthy mucosa must be seen as a limitation of this study. A biopsy of the contralateral vocal fold to fully exclude epithelial changes would, however, not be ethically acceptable.

Some groups provide their own “training programme” before showing the pCLE images to be analysed to the test examiners. This bears the risk of bias with respect to interpretation and makes comparison between studies more difficult. Because there is no specific training programme for head and neck cancer, we had to opt for the use of the training programme available for the oesophagus, despite its limitations when extrapolated to the mucosa of the larynx. Moore et al. reported an accuracy, sensitivity, specificity, PPV, and NPV of 100% using a similar methodical approach: 29 offline images and 6 video sequences. Oetter et al. investigated the value of pCLE in the classification of lesions of the oral mucosa and suggested a classification and scoring system to facilitate the interpretation of these images by examiners without prior experience in this technique. The suggested scoring system enabled a sensitivity of over 95% and a specificity of 89% and showed excellent agreement between examiners.

The contrast agent (fluorescein), administrated by i.v. injection prior to the examination, accumulates in the intercellular spaces, thereby enabling the imaging of the cell outlines as well as the small capillaries, but does not accumulate in nuclei. Visualisation of nuclei is, however, essential for diagnosis and grading of malignant lesions in head and neck cancer. The most commonly used contrast agent for the nuclei in CLE is acriflavine. When administrated topically, acriflavine passes the cell membrane and binds strongly to the acidic constituents of the nucleus, thus enabling the staining of the superficial labels of the epithelium. It allows the differentiation of epithelial cells, goblet cells and other pathological patterns.

The potential of staining agents of the nuclei in the head and neck region was examined by Linxweiler et al. in formalin-fixed samples in 2016. Acriflavine showed that it stained the nuclei while suppressing the autofluorescence of collagen fibres, only marginally improving the margin detection of tumours in the head and neck. Because of this only modestly positive net effect, the authors do not recommend the use of acriflavine. As formaldehyde changes the nuclear proteins and tissue autofluorescence, these results cannot be directly transferred to the examination with CLE in vivo. Recently, an improved version of acriflavine, acrinol, was suggested as an alternative topical contrast agent for the nucleus. This contrast agent has shown minimal mucosal irritation, being mostly excreted in stools while still showing increased nuclear density and prominent abnormalities in carcinoma cells. Further studies on the toxicity of acrinol will, however, be required before it can be routinely applied in vivo.

Motivated by the difficulty in reliable and reproducible image interpretation, algorithms for the automatic classification of CLE images have been recently emerging. The application of deep learning algorithms to CLE images, as described by Aubreville et al. based on Convolutional Neural Networks, was able to correctly recognise CLE images of oral SCC with an accuracy of 88.3%, a sensitivity of 86.6% and a specificity of 90%.

To confirm the robustness of this model the group applied the algorithm to our dataset of the vocal cords and obtained an accuracy of 90.7%. The quality of the video sequences can be also diminished due to motion artefacts that are usually caused by slight movements of the probe during examination. Although these motion artefacts are relatively easy for a human examiner to overlook, they represent a relevant interfering factor in automatic analysis. Detection of motion artefacts demonstrated that the performance can be improved by pattern recognition algorithms that recognise malignant changes. For comparison, the results presented in this study gained by two trained and certified examiners show an accuracy of 91.38-96.55%. We find both these results very encouraging and worth pursuing in future studies. If the detection of SCC through pCLE proves to be reliable in further studies, this novel technique could help to better define surgical margins in real time (e.g. TLM) and allow a more selective use of biopsy in the follow-up of patients who underwent TLM, as recurrence incidence was shown to be around 15%.
Conclusions

In this study, we showed that malignant lesions of the vocal cords can be reliably and accurately differentiated from healthy epithelium with an accuracy, sensitivity, specificity, PPV, and NPV of 91.38-96.55%, 100%, 87.8-95.2%, 77.27-89.47% and 100%, respectively. This suggests the potential of the non-invasive diagnosis of SCC in vivo. This is particularly important in the vocal folds, as a biopsy can cause scarring with irreversible damage to the vocal folds. Further development of staining of the nuclei and especially the optimisation of training programs for the human examiner as well as the deep learning algorithms that constitute the core of the automatic classification of CLE images are very promising and should be the focus of further investigation.

Conflict of interest statement

None declared.

References

Confocal laser endomicroscopy in detecting malignant lesions of the vocal folds


Received: February 17, 2018 - Accepted: April 10, 2018