ORIGINAL ARTICLE

Variability in nomenclature of benign laryngeal pathology based on video laryngoscopy with and without stroboscopy

Chau, H.N.,* Desai, K.,† Georgalas, C.,* & Harries, M.†

*Department of ENT, Whipps Cross University Hospital, London, and †Department of ENT, The Royal Sussex County Hospital, Brighton, UK

Accepted 15 July 2005
Clin. Otolaryngol. 2005, 30, 424–427

Objectives: To assess the extent of interobserver variability in the nomenclature of benign laryngeal pathology based on evaluation of video-endoscopies with and without stroboscopy.
Design: Eight video clips of benign laryngeal conditions were viewed by 35 non-voice specialist ENT surgeons. The surgeons viewed the clips in groups of varying sizes with no discussion between them and were asked to make only one diagnosis for each lesion.
Setting: Specialist voice clinic in the department of ENT at The Royal Sussex County Hospital in Brighton, UK.
Participants: Participating ENT surgeons were all either registrars or consultants working at different centres in UK, recruited by the author. None were voice specialists.
Main outcome measure: Interobserver agreement was measured using kappa statistics.
Results: Variation was widespread with only two of the eight cases (25%) showing agreement of over 75%. Agreement could be analysed statistically as moderate at best ($\kappa = 0.5$ with a 95% confidence interval from 0.5 to 0.6). The seniority of the laryngologist was also analysed with consultants and senior trainees (specialist registrar years 4–6) having better agreement than junior trainees (specialist registrar years 1–3).
Conclusions: The generally accepted optimum treatment for different benign laryngeal pathologies varies substantially. However, our results shows a significant high-level interobserver variability in their diagnosis by non-voice specialists, thus reducing the reliability of outcome data and treatment recommendations. It is therefore important to try and lower this interobserver variability, possibly by widespread use of improved diagnostic technology, stricter/more universally accepted definitions and supervised training of junior doctors in a voice clinic environment.

The clinical diagnosis of benign laryngeal pathology is important, as this primarily determines the next step in the treatment. Unfortunately, it is not uncommon for a group of clinicians to use different nomenclature for the same laryngeal lesion. This makes it difficult to evaluate and compare the reports of different pathological vocal conditions and their treatment. Variation between clinicians in interpretation and naming of benign laryngeal lesions has previously been demonstrated, by Dikkers,[1] in a study using photographs. Our study looks at whether this interobserver variability in nomenclature is also the case, even with the aid of moving images and stroboscopy.

Correspondence: H.N. Chau, MRCS, 207 Colfax Building, 23 Plumbers Row, London E1 1EQ, UK, Tel.: 07931 737552; e-mail: hanychau@hotmail.com.

Materials and Methods

Videos

Videos of various benign laryngeal conditions in patients, who attended the specialist voice clinic in the department of ENT at The Royal Sussex County Hospital in Brighton, UK, were examined retrospectively. The clinical examination, video recording and diagnosis were carried out by one laryngologist (MH). However, for the purpose of this study, the correct diagnosis was not important, as we were actually looking at interobserver agreement/variability rather than right or wrong answers.

Both white halogen and stroboscopic light were used in the clinical examination. A rigid Hopkins 70 degree endoscope was used in most cases, although a flexible endoscope was also used for the white light and the laryngostroboscopic examination in a few cases in which the patient could not tolerate the rigid endoscope. Eight video clips of benign laryngeal conditions were selected and transferred

on to CD. The clips were on a media player so that they could be replayed or put on pause for closer inspection.

Surgeons

These clips were then shown to 35 ENT surgeons who were all either registrars or consultants working in UK, recruited on voluntary basis at their hospitals or at academic meetings. Consultants with a subspecialty interest in the voice were excluded. The surgeons viewed the clips in groups of varying sizes with no discussion between them. They were given as much time as they wanted to repeat or pause each of the eight clips. There was no list of potential diagnoses to choose from, and only one diagnosis was permitted for each lesion.

Results

As mentioned above, the correct diagnosis was not important. For the purpose of this study, the most prevalent diagnosis for each of the eight lesions was interpreted as the 'correct' answer. Thus, the percentage agreement for the most prevalent diagnosis in each case is given in Table 1. In only two cases (25%) was there an agreement of >75% and in two cases (25%) an agreement of only 40%.

Table 1. Most prevalent diagnosis for each lesion

Lesion number	Most prevalent diagnosis	Percentage agreement
1	Nodule	68
2	Reinke's oedema	94
3	Granuloma	40
4	Cyst	57
5	Polyp	77
6	Nodule	57
7	Reinke's oedema	40
8	Nodule	62

From these data, six main diagnoses were observed: nodules (lesions 1, 6 and 8), Reinke's oedema (lesions 2 and 7), granuloma (lesion 3), cyst (lesion 4), polyp (lesion 5) and 'other diagnoses' (including no answer). Thus, to look at the total percentage agreement, we can tabulate the data (Table 2).

One way of measuring this interobserver agreement is using kappa statistics. Kappa varies from 0 to 1 (the higher the kappa value, the better the agreement). The results from this study showed $\kappa = 0.528$ with a 95% confidence interval from 0.458 to 0.599. Thus, the strength of agreement, in our study, is considered to be only moderate at best.

It was also interesting to see how the level of training affected the results. We separated the observers into two groups, either junior trainee (specialist registrar year 1–3) or consultant/senior trainee (specialist registrar year 4–6). Unfortunately, seven of the 35 observers did not indicate their training status. The results for 28 observers are shown in Table 3.

In six out of the eight cases, the agreement was higher in the consultant/senior trainee group. This group also used the diagnosis of granuloma and more cyst, and the 'other' diagnosis much less ($P = 0.16$, not significant). If we take the most prevalent as the 'correct' diagnosis, then we can see that there is a significant difference between the two groups. The consultant/senior trainee group showed better agreement and more 'correct' answers, 73% compared with 60% correct answers in the junior trainee group. On chi-square testing this was significant ($P = 0.044$).

Discussion

A detailed search of the literature showed only one other study looking at interobserver variability in the diagnosis

Table 2. Percentage agreement for each diagnosis.

	Observer diagnosis						
Categories	Nodules	Reinke's	Granuloma	Cyst	Polyp	Others	Total (% agreement)
Nodules	66	3	1	6	4	25	105 (63)
Reinke's	0	47	0	7	3	13	70 (67)
Granuloma	4	0	14	2	12	3	35 (40)
Cyst	8	0	0	20	2	5	35 (57)
Polyp	3	0	0	1	27	4	35 (77)
Others	0	0	0	0	0	0	0
Total	81	50	15	36	48	50	280

Number of observed agreements: 174 (62.1% of the observations).
Number of agreements expected by chance: 55.3 (19.7% of the observations).

Table 3. Comparison between junior trainees (specialist registrar year 1–3) and consultant/senior trainees (specialist registrar year 4–6)

Number of different diagnosis[a]	Percentage of correct diagnosis[a]	Number of different diagnosis[b]	Percentage of correct diagnosis[b]	Lesion
2	81.8	3	70.6	1. Nodule
1	100	2	88.2	2. Reinke's
3	63.6	4	29.4	3. Granuloma
4	72.7	4	47.1	4. Cyst
1	100	3	76.5	5. Polyp
5	54.5	4	70.6	6. Nodule
4	36.4	4	47.1	7. Reinke's
3	72.7	4	47.1	8. Nodule

[a]Consultant/senior trainee.
[b]Junior trainee.

Table 4. Suggested diagnostic criterion for simple glottic lesions (after Dikkers and Schutte 1991)

Cyst: A unilateral lesion with a smooth surface, immobile during phonation, usually on the middle third of the vocal fold and often of a yellowish fluid-like appearance.
Reinkes oedema: A condition with a unilateral or bilateral bleach-white swelling of the vocal fold, filled with fluid, is sessile, and very mobile during phonation.
Polyp: A unilateral lesion on the anterior third of the vocal fold, often on the free edge, sessile or pedunculated, and very mobile when pedunculated – a pedunculated polyp on the free edge can sometimes be heard popping through the glottis during the initiation of a phonation.
Nodules: Small lesions occurring on both sides of the larynx, strictly symmetrical on the border of the anterior and middle third of the vocal folds, and usually immobile during phonation – they can be divided, with the aid of stroboscopy, into the early spindle type (soft), or more chronic, white, cone-like form (hard).

of benign vocal cord lesions.[1] In this study by Dikkers, uniformity in clinical diagnosis was evaluated by means of showing slides to groups of ENT surgeons. Although the actual agreement between doctors in the total of the 45 slides used in this study is not documented, it was observed that in only five of 45 slides (11%) were there an agreement of more than 75%. In our study, two of eight (25%) video clips showed an agreement of more than 75%. It is likely that the use of moving images and stroboscopy in our study aided clinicians in their diagnosis and increased the overall agreement. It is interesting that stroboscopy considered the standard in the diagnosis of benign mucosal lesions of the vocal cord, which is not commonly used by the majority of otolaryngologists.[2] Stroboscopic evaluation has been found useful in differentiating between polyps and cysts, and between vocal fold nodules and unilateral lesions with contralateral reactive changes.[3,4] It may be that under-utilization of stroboscopy, and consequent misdiagnosis, contributes to the obscurity of the relationship between specific intervention and diagnoses, thus preventing preferences and recommendations on management from emerging.

Apart from the factors mentioned above, interobserver variability is also likely to be because of confusion in the definitions used for these lesions. Although there is no definite common terminology for vocal cord lesions at present, variations of the descriptions proposed by Dikkers appear to be the most commonly used (Table 4).[1,5]

If the nature of the lesion is such that it cannot be classified according to these definitions, then it should be described with emphasis on location on the vocal fold (divided into three parts), uni or bilaterality, colour, mobility during phonation (using both white light and stroboscopy) and on the size of its attachment to the fold. This is important, as there will be some lesions that can only be diagnosed by histology,[6] in which case it has been shown that providing good clinical information influences the diagnosis made by pathologist.[3]

Although our study made use of moving images and stroboscopy, the results still only showed moderate agreement ($\kappa = 0.528$) in diagnoses/nomenclature. This would not be of any great significance if there were no therapeutic consequences of interobserver disagreement. However, the generally accepted optimum treatment for these diagnoses differs substantially.[2,7] Thus, patients with vocal nodules who might respond to speech therapy may be undergoing unnecessary surgery, and conversely, the surgical removal of cysts or papillomas may be delayed by inappropriate referral to speech therapy which is already overburdened. Speech therapists themselves may lose confidence in the laryngologists' referrals in light of such a large variation in diagnosis and terminology.

The results from our study showed significantly better agreement in the consultant/senior trainee group, and one would assume that agreement would be even higher for surgeons with a subspecialty interest in the voice. However, this raises the question of whether junior doctors are suitable for diagnosing and managing benign laryngeal pathology. Great emphasis in training is put on the early recognition of carcinoma of the larynx and many may see their role as excluding malignancy only. Thus, it may be that all patients with a hoarse voice should ideally be seen in a voice clinic with better equipment and the input of a speech therapist.

This paper makes a good case for review of the diagnosis and clinical terminology of benign laryngeal pathology.

We believe that interobserver variability is too high and attempts should be made to lower this by using improved diagnostic technology, stricter/more universally accepted definitions and supervised training of junior doctors in a voice clinic environment. This should in turn lead to useful outcome data and treatment recommendations.

Conflict of interest

None declared.

References

1 Dikkers F.G. & Schutte H.K. (1991) Benign lesions of the vocal folds: uniformity in assessment of clinical diagnosis. *Clin. Otolaryngol.* **16,** 8–11

2 Sulica L. & Behrman A. (2003) Management of benign vocal fold lesions: a survey of current opinion and practice. *Ann. Otol. Rhinol. Laryngol.* **112,** 827–833

3 Shohet J.A., Courey M.S., Scott M.A. *et al.* (1996) Value of videostroboscopic parameters in differentiating true vocal fold cysts from polyps. *Laryngoscope* **106,** 19–26

4 Colton R.H., Woo P., Brewer D.W. *et al.* (1995) Stroboscopic signs associated with benign lesions of the vocal folds. *J. Voice* **9,** 312–325

5 Poels P.J., de Jong F.I. & Schutte H.K. (2003) Consistency of the preoperative and intraoperative diagnosis of benign vocal fold lesions. *J. Voice* **17,** 425–433

6 Loire R., Bouchayer M., Cornut G. *et al.* (1988) Pathology of benign vocal fold lesions. *Ear Nose Throat J.* **67,** 357–368, 360–362

7 Dikkers F.G. (1991) Intraobserver variation in diagnosis of benign non-neoplastic lesions of vocal folds. *Lancet* **337,** 866