Evaluation of Vocal Fold Motion Abnormalities: Are We All Seeing the Same Thing?

Lyndsay L. Madden and Clark A. Rosen, *Pittsburgh, Pennsylvania*

Summary: Objectives. Flexible laryngoscopy is the principle tool for the evaluation of vocal fold motion. As of yet, no consistent, unified outcome metric has been developed for vocal fold paralysis/immobility research. The goal of this study was to evaluate vocal fold motion assessment (inter- and intra-rater reliability) among general otolaryngologists and fellowship-trained laryngologists.

Study Design. Prospective video perceptual analysis study.

Methods. Flexible laryngoscopic examinations, with sound, of 15 unique patient cases (20 seconds each) were sent to 10 general otolaryngologists and 10 fellowship-trained laryngologists blinded to clinical history. Reviewers were given written definitions of vocal fold mobility and immobility and two video examples. The cases included bilateral vocal fold mobility (six), unilateral vocal fold immobility (five), and unilateral vocal fold hypomobility (four). Five examinations were repeated to determine intra-rater reliability. Participants were asked to judge if there was or there was no purposeful motion, as described by written definitions, for each vocal fold (800 tokens in total).

Results. Twenty reviewers (100%) replied. Both general otolaryngologists and fellowship-trained laryngologists had an overall inter-rater reliability of 95%. Difference in inter-rater reliability between the two groups of raters was negligible: 95% for general otolaryngologists and 97.5% for fellowship-trained laryngologists. There was no variability in intra-rater reliability within either rater group (99%).

Conclusion. Intra- and inter-rater agreement in determining whether the patient had purposeful vocal fold motion on flexible laryngoscopic examination was excellent in both groups. This study demonstrates that otolaryngologists can consistently and accurately judge the presence and the absence of vocal fold motion.

Key Words: Video perceptual analysis–Vocal fold motion–Vocal fold immobility–Vocal cord paralysis–Vocal cord immobility.

INTRODUCTION

Laryngoscopy is the principle diagnostic instrument for the evaluation of every patient with a voice-related complaint.[1] It is often the singular diagnostic instrument upon which treatment is based. In particular, flexible laryngoscopy is the modality of choice for the evaluation of vocal fold motion abnormalities. Despite the primal nature of vocal fold motion assessment for vocal fold paralysis, voice researchers have yet to develop a clear-cut outcome measure for vocal fold paralysis research. Some researchers[2,3] have studied whether certain laryngoscopic findings of vocal fold paralysis can consistently be rated, but none have investigated if otolaryngologists are accurate and consistent in measuring vocal fold motion, purely. Distinguishing whether a patient has vocal fold motion is important in the proactive treatment of patients with vocal fold motion abnormalities.[4,5] The purpose of this study was to determine if general otolaryngologists and fellowship-trained laryngologists can consistently and accurately determine various states of vocal fold motion.

MATERIALS AND METHODS

Approval of this study was obtained from the University of Pittsburgh Institutional Review Board.

Accepted for publication December 15, 2015.

Accepted for poster presentation at American Broncho-esophagological Association (ABEA) 2015, Boston, Massachusetts, April 22–26.

From the University of Pittsburgh Voice Center, Department of Otolaryngology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania.

Address correspondence and reprint requests to Clark A. Rosen, University of Pittsburgh Voice Center, UPMC Mercy Hospital, 1400 Locust St. Bldg B Suite 11500, Pittsburgh, Pennsylvania 15219. E-mail: rosenca@upmc.edu

Journal of Voice, Vol. 31, No. 1, pp. 72–77

0892-1997

http://dx.doi.org/10.1016/j.jvoice.2015.12.009

Selection of examinations

The patients included in the study were patients with laryngeal electromyography who underwent flexible laryngoscopic examinations and were proven to have vocal fold paralysis; patients with clinically relevant, purposeful vocal fold motion; and patients with varying degrees of hypomobility on flexible laryngoscopic examination. These examinations were collected consecutively at our institution. Each case was reviewed and selected based on their video quality and whether they had an laryngeal electromyography (LEMG) data. Each patient with no purposeful motion had laryngeal electromyographically proven vocal fold paralysis; however, this was not an inclusion criterion for patients with vocal fold hypomobility as this is not standard practice at our institution. Each LEMG was interpreted by a single board-certified electromyographer. Each patient with purposeful motion was chosen by the authors after reviewing the video-recorded examination. All recordings were made using a distal chip-tip, flexible transnasal laryngoscope (ENF-VT2 Distal Chip Tip, Olympus America Inc., Center Valley, PA, USA). Twenty-second video clips, with sound, from 15 unique examinations were chosen. Each patient was recorded performing the following common tasks for the assessment of vocal fold motion: quiet respiration, sustained phonation, and alternation between phonation and a sniff. The selections included case types of six examinations that had bilateral mobile vocal folds, five with unilateral vocal fold paralysis, and four with varying degrees of unilateral hypomobility. The sample size chosen was based off of a statistical paper by Donner and Eliasziw,[6] which aids in the planning of reliability studies, as well as literature by Fleiss on the design of clinical experiments.[7–9] Five of the examinations were repeated to assess intra-rater

reliability. These videos were randomly selected from our data set.

Selection of reviewers

Ten fellowship-trained laryngologists and 10 general otolaryngologists, ranging in years of practice from 1 to 23 (Appendix A), were asked to participate in the study. Each reviewer received a PowerPoint (Mac Version 2011, Microsoft Corporation, Redmond, WA, USA) presentation with one individual case per slide. Reviewers were given a rating form that included written definitions of purposeful vocal fold motion and no purposeful vocal fold motion (Appendix B), and two video examples of each were also included in the PowerPoint presentation. Results from each rating sheet were entered into Excel spreadsheets (Mac Version 2011, Microsoft Corporation) for analysis of the data.

Rating of examinations

Two laryngoscopic criteria were measured by each rater in a binary fashion: purposeful vocal fold motion or no purposeful vocal fold motion. Vocal fold mobility, for this study, was defined as purposeful, adduction and/or abduction of the vocal folds on clinical examination. Vocal fold immobility was defined as no active or voluntary adduction and/or abduction of the vocal fold on clinical examination.[10] Reviewers were asked to rate each examination sequentially and could repeat them as needed. They were blinded to the clinical history of the patient.

Statistical analysis

Inter-rater reliability was determined using the Fleiss kappa statistical metric.[7,8] Using this statistical measure, values near zero represent poor agreement and values close to one represent good agreement.

Intra-rater reliability was determined simply by using the number of points of agreement divided by the total. This is reported as the percent agreement among raters. Raw data for each of the scored cases are included in Appendix C.

RESULTS

Twenty examiners returned the survey, for a 100% response rate. The summed result of general otolaryngologists and fellowship-trained laryngologists resulted in an overall inter-rater reliability of 95%. Difference in inter-rater reliability between the two groups of raters was negligible: 95% for general otolaryngologists and 97.5% for laryngologists. There was essentially no variability in intra-rater reliability within either rater group (99%) (Table 1 and Table 2). Overall, there were only four tokens scored differently from the *a priori* designation of motion status. These cases were as follows: case 3—left, case 12—right, case 17—right, and case 17—left.

TABLE 1.
Overall Inter-rater Reliability for Both Groups

N	Mean	Standard Deviation	Inter-rater Reliability
20	99.5	1.3	95.0

TABLE 2.
Inter-rater Reliability for General Otolaryngologists and Fellowship-trained Laryngologists

Fellowship Training *vs.* General otolaryngologist (ENT)

Fellowship Training	N	Mean	Standard Deviation	Inter-rater Reliability
No	10	99.3	1.7	95.0
Yes	10	99.8	0.8	97.5

Of the cases with vocal fold hypomobility, case 3—left, case 9—right and left, case 15—right and left (cases 9 and 15 were repeated cases), and case 19—left, only one token was scored discordantly: case 3—left. Of note, case 9 was an examination with bilateral vocal fold hypomobility.

DISCUSSION

The management of patients with vocal fold motion impairment begins with proper diagnosis, based frequently upon a flexible transnasal laryngoscopic examination. Overall, both fellowship-trained laryngologists and general otolaryngologists demonstrated a very high rate of concordance of determining purposeful vocal fold motion or no purposeful vocal fold motion.

Of the discordant items, some assumptions can be made as to why they were scored differently. The discordant cases were case 3—left, case 12—right, case 17—right, and case 17—left.

Case 3 included an examination with the *a priori* label of vocal fold hypomobility on the left and a mobile vocal fold on the right with a vocal process granuloma. The presence of the right vocal process granuloma could have dissuaded the rater, a fellowship-trained laryngologist, into choosing a different response from that of his or her colleagues.

Case 12 included an examination with the *a priori* label of left vocal fold immobility and a right hypomobile vocal fold on abduction, with a narrowed glottis. The rater, a general otolaryngologist, rated that neither vocal fold had purposeful motion. The rater could have interpreted a narrowed glottis as bilateral vocal fold immobility.

Finally, in case 17, the rater, a general otolaryngologist, rated that the right vocal fold was mobile and the left was immobile. This was opposite from what the *a priori* assessment and the other reviewers' assessment were for this case. This could be attributed to human error and transposition of the responses inadvertently. Interestingly, if we were to exclude these outlying cases from the data analysis, the number of consistent and correct responses would have been 100% for each group.

Overall, this study definitely demonstrates that raters can consistently rate motion or no motion, but in analyzing our study, we may have been biased in favor of greater inter-rater agreement by the inclusion of sound in the videos sent to raters. This information may have provided hints to blinded reviewers of who were being studied for their video perceptual analysis. This study design included sound because we wanted to replicate normal clinical care, as much as possible.

Another possible area of bias lies in the educational background information of each rater. In future studies, it may be