

Inter- and Intra-Rater Reliability of Laryngeal Sensation Testing with the Touch Method During Flexible Endoscopic Evaluations of Swallowing

Annals of Otolaryngology, Rhinology & Laryngology
1–7

© The Author(s) 2020


Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0003489419901145

journals.sagepub.com/home/ao



James C. Borders, MS¹ , Meredith B. O’Dea, MS¹, Edel McNally, MS¹, Elizabeth Norberg, MS¹, Merertu Kitila, MS¹, Michael Walsh, MEd¹, Rui Liu, MS², and Jessica M. Pisegna, PhD, MEd^{1,3}

Abstract

Objectives: Sensation is an integral component of laryngeal control for breathing, swallowing, and vocalization. Laryngeal sensation is assessed by elicitation of the laryngeal adductor reflex (LAR), a brainstem-mediated adduction of the true vocal folds. During Flexible Endoscopic Evaluations of Swallowing (FEES), the touch method can be used to elicit the LAR to judge laryngeal sensation. Despite the prevalence of this method in clinical practice and research, prior studies have yet to examine inter- and intra-rater reliability.

Methods: Four speech-language pathologists rated 125 randomized video clips for the presence, absence, or inability to rate the LAR. Fifty percent of video clips were re-randomized and re-rated 1 week later. Raters then created guidelines and participated in formal consensus training sessions on a separate set of videos. Ratings were repeated post-training.

Results: Overall inter-rater reliability was fair ($\kappa=0.22$) prior to training. Pre-training intra-rater reliability ranged from fair ($\kappa=0.35$) to almost perfect ($\kappa=0.89$). Inter-rater reliability significantly improved after training ($\kappa=0.42$, $P<.001$), though agreement did not reach prespecified acceptable levels ($\kappa\geq 0.80$). Post-training intra-rater reliability ranged from moderate ($\kappa=0.49$) to almost perfect ($\kappa=0.85$).

Conclusion: Adequate inter-rater reliability was not achieved when rating isolated attempts to elicit the LAR. Acceptable within-rater reliability was observed in some raters 1 week after initial ratings, suggesting that ratings may remain consistent within raters over a short period of time. Limitations and considerations for future research using the touch method are discussed.

Keywords

laryngeal sensation, laryngeal adductor reflex, touch method, reliability, FEES

Background

Sensation is an integral component of laryngeal control for airway protection, breathing, swallowing, and vocalization, and precise sensorimotor integration is required for guiding and modulating these movements and reflexes.^{1–4} Laryngeal sensation can be assessed by elicitation of the laryngeal adductor reflex (LAR), a brainstem-mediated adduction of the true vocal folds.⁵ The afferent branch of the LAR is comprised of the internal branch of the superior laryngeal nerve, which receives sensory input from mechanoreceptors and chemoreceptors in the laryngeal mucosa.⁵ This afferent input provides sensory feedback to central neural circuits in the medulla with subsequent elicitation of the efferent component of the reflex through the recurrent laryngeal nerve, which

ultimately results in a brief, bilateral contraction of the thyroarytenoid muscle.^{5–7} An intact LAR requires sensorimotor integrity at both peripheral and central neural

¹Boston Medical Center, Boston, MA, USA

²Department of Communication Sciences and Disorders, MGH-Institute of Health Professions, Boston, MA, USA

³Department of Otolaryngology, School of Medicine, Boston University, Boston, MA, USA

Corresponding Author:

Jessica M. Pisegna, PhD, MEd, Director of Research and Clinic, Department of Otolaryngology, Speech-Language Pathology, Boston Medical Center, 830 Harrison Ave, Suite 1400, Moakley Building, Boston, MA 02118, USA; Assistant Professor of Otolaryngology, Otolaryngology—Head & Neck Surgery, School of Medicine, Boston University, BCD Building, 5th floor, 800 Harrison Ave, Boston, MA 02118, USA.
Email: jessica.pisegna@bmc.org

structures, including sensory afferents, interneurons, and motor signaling.

Laryngeal sensation has been examined in patients with gastroesophageal reflux,⁸⁻¹⁰ obstructive sleep apnea,¹¹⁻¹⁴ stroke,^{15,16} acute respiratory failure,¹⁷ tracheostomy,¹⁸ Parkinson's disease,^{2,19} amyotrophic lateral sclerosis,²⁰ head and neck cancer,^{19,21} partial laryngectomy,²² chronic cough and paradoxical vocal fold movement,²³ and pediatric populations with varying medical diagnoses.^{24,25} In the dysphagia literature, an association between laryngeal sensory deficits and aspiration,^{15,17,26,27} pharyngolaryngeal secretions,^{17,28} and pneumonia^{29,30} has been documented. However, methodological variability exists across studies, including various operational definitions of laryngeal sensation (eg, self-report, LAR, cough, or swallow), as well as location (eg, arytenoids, epiglottis, aryepiglottic folds) and type of stimuli administered (eg, air pulse or touch method).

The LAR is often included as part of Flexible Endoscopic Evaluations of Swallowing (FEES) with two methods: the air pulse³¹ and touch method.³² The air pulse method, also called FEES with sensory testing (FEESST), applies pressure and duration-controlled air pulses to the laryngeal mucosa through a working channel of a flexible endoscope.³¹ Though normative values have been established with the air pulse method,³³ its clinical use is currently lacking since this specialized equipment is no longer commercially manufactured.

Tactile stimulation of laryngeal structures, known as "the touch method," is routinely used in clinical practice since it does not require specialized equipment beyond an endoscope. During laryngeal sensation testing, the endoscope is advanced to make contact with an arytenoid. Once contact is made, the endoscope is partially retracted to view the presence or absence of the LAR. However, there are inherent methodological limitations, including variability of pressures both within and between endoscopists,³⁴ inconsistency in the location of tactile stimulation of the endoscope, and poor visualization due to obstruction from residue or secretions. Additional patient factors, such as volitional vocal fold adduction during testing or poor exam tolerance, can limit one's ability to assess the LAR. Despite these limitations, prior studies report on laryngeal sensation outcomes using the touch method without any reported reliability measures.^{15-17,19,29,35,36}

This study sought to examine the inter- and intra-rater reliability of clinician ratings of the LAR during laryngeal sensation testing with the touch method. Following assessment of baseline reliability, additional guidelines were developed and consensus training was performed before post-training reliability ratings were completed. We hypothesized that raters would achieve acceptable inter- and intra-rater reliability ($\kappa \geq 0.80$) after training.

Material and Methods

Study Design

This prospective study examined inter- and intra-rater reliability for ratings of the LAR during the touch method of laryngeal sensation testing both before and after training. Videos of laryngeal sensory testing were retrospectively obtained from a database of FEES videos performed as standard of care exams on dysphagic patients seen in the ambulatory and inpatient setting of a large urban hospital. Inclusion criteria included laryngeal sensation testing during FEES as either an inpatient or outpatient evaluation. Exclusion criteria included videos without a clear view of the larynx at rest, or suspected unilateral or bilateral vocal fold immobility identified during phonatory speech tasks on FEES ("eee—sniff"), which elicited maximum vocal fold abduction and adduction. In these standard of care FEES videos, laryngeal sensation testing was performed at the end of the FEES and involved brief tactile stimulation of each arytenoid with the tip of the endoscope. Since videos were obtained from FEES performed during routine clinical care, tactile stimulation duration and pressure was not controlled or recorded. The decision to touch the arytenoids was based on research suggesting that this area contains the highest density of mechanoreceptors compared to other laryngeal subsites, such as the epiglottis.³⁷

Laryngeal Sensory Testing Videos

Videos were examined and obtained by a research assistant. First, the entire laryngeal sensation testing exam was obtained, then each video was divided into isolated clips for each individual tactile stimulation of an arytenoid. A total of 125 video clips were obtained for pre- and post-training reliability, as well as a fifty video clips for the training session.

Pre-Training Reliability Ratings

Four speech pathologists with clinical experience performing FEES and laryngeal sensation testing with the touch method individually performed baseline and post-training LAR ratings. REDCap (Research Electronic Data Capture), a secure web-based data collection tool, was used to collect ratings for reliability.³⁸ Videos were put in random order by a research assistant and raters were blinded to patient demographics. Two exemplar video clips of a present LAR and one video clip of an absent LAR were provided to raters at the beginning of each rating session. The exemplar videos were chosen by consensus from clinicians who were not raters. Both frame-by-frame analysis and real-time observation to judge the LAR was encouraged. Raters were instructed to judge the LAR as present only when adduction of the arytenoids or true vocal folds was visualized. An

Table 1. Guidelines Used During Training Sessions and Post-Training Reliability Ratings.

Rating	Observation
Present:	<ul style="list-style-type: none"> • Robust and immediate adduction of the arytenoids and/or vocal folds after tactile stimulation.
Absent:	<ul style="list-style-type: none"> • No movement of the arytenoids and/or vocal folds despite tactile stimulation.
Unable to rate:	<ul style="list-style-type: none"> • Adduction of the vocal folds occurred before the endoscope touched the arytenoid. • The endoscope did not make contact with the arytenoid. Contact is defined when the entire screen is occluded during an attempt. • After the endoscope touched the arytenoid, the larynx could not be visualized due to obstruction of the view from bolus residue or secretions. • The patient swallowed, coughed, or gagged after tactile stimulation of the arytenoid. • Movement of arytenoids or vocal folds is related to the patient's habitual breathing pattern.

absent LAR was defined as no movement of the arytenoids or true vocal folds. No criteria for the inability to rate the LAR or specific rating time frame was provided. Intra-rater reliability was performed 1 week after the initial rating session on a randomly selected 50% of video clips.

Training Session

The raters met to discuss and develop rating guidelines after baseline reliability ratings were completed (see Table 1). Once guidelines were established, the four raters individually judged 50 video clips, which were different from pre-training videos. The first author (JCB) served as a facilitator to guide rating discussions and disagreements were resolved in real-time by consensus after watching the video clip again. Two training sessions, each lasting approximately 2 hours, were completed by all raters.

Post-Training Reliability Ratings

The same four speech pathologists were instructed to individually rate video clips from the same, re-randomized database of videos from the pre-training rating session. Raters were provided guidelines developed at the training session, as well as exemplar videos. Intra-rater reliability was again performed 1 week later on a randomly selected 50% of video clips.

Statistical Analyses

Statistical analyses were performed in R³⁹ with the *rel* open-source package.⁴⁰ In order to examine inter-rater reliability across four raters with a categorical outcome (present, absent, unable to rate the LAR), Fleiss' kappa (κ) was computed pre- and post-training. Intra-class correlation coefficients (ICC) were calculated to assess overall intra-rater reliability for pre- and post-training ratings. Estimates were obtained via an intercept only multinomial logistic mixed effects model. Cohen's κ was used to examine inter- and intra-rater reliability between dyads and within each rater. Fleiss and Cohen's κ values were interpreted as follows:

values ≤ 0 as indicating no agreement, 0.01-0.20 as none to slight, 0.21-0.40 as fair, 0.41-0.60 as moderate, 0.61-0.80 as substantial, and 0.81-1.00 as almost perfect agreement.⁴¹ An a priori Fleiss' κ value of 0.80 was set for acceptable reliability. The Fleiss κ statistic was compared between pre- and post-training using a linearization method to compare correlated agreement coefficients.⁴²

Results

Patient Demographics for Pre- and Post-Training Videos

Laryngeal sensory testing videos were obtained from the FEES of 48 patients. The majority of exams were from inpatients (54%) without nasogastric tubes (83%). The mean age of patients was 59 years (SD= 16 years). Primary medical diagnoses included neurologic (33%), head and neck cancer (29%), respiratory (17%), esophageal (11%), and spinal cord injury (10%). 0.2 mL of 4% atomized lidocaine was administered (75%) at the discretion of the clinician and laryngeal abnormalities (eg, erythema, edema, ulcerative tissue) were appreciated in nearly half of patients (46%). On average, the clinician performing the FEES with sensory testing performed three attempts (range= 1-5) of tactile stimulation per exam. In total, 125 video clips (isolated attempts at eliciting the LAR) were included for reliability ratings. Raters had an average of 8 years of experience performing FEES with the touch method of sensory testing (range= 1-25 years). Specifically, rater A had 1 year of experience, rater B had 2 years of experience, rater C had 12 years of experience, and rater D had 25 years of experience.

Pre-Training Inter- and Intra-Rater Reliability

Reliability analyses between rater dyads are reported in Table 2. Fair inter-rater reliability was achieved between all raters (Fleiss' κ =0.22, 95% CI 0.17-0.28). An item analysis revealed that raters achieved complete agreement on 30% of the video clips, agreement in three out of four raters on

Table 2. Inter-Rater Reliability.

	Pre-Training	Post-Training
Rater A versus Rater B	0.39	0.57
Rater A versus Rater C	0.27	0.47
Rater A versus Rater D	0.25	0.42
Rater B versus Rater C	0.25	0.44
Rater B versus Rater D	0.26	0.40
Rater C versus Rater D	0.34	0.39
Overall*	0.22	0.42

Note. All values are Cohen's κ , except for overall (Fleiss' κ).
* $P < .001$.

37% of the video clips, and agreement in two out of four raters on 34% of the video clips. On videos where disagreements were limited to only two responses (eg, present or unable to rate), the most common disagreements were between present versus unable to rate (74%), compared to absent versus unable to rate (14%) or present versus absent (12%). Overall intra-rater reliability was moderate (ICC=0.48), ranging from fair (Cohen's $\kappa=0.35$) to almost perfect (Cohen's $\kappa=0.89$) agreement (Table 3).

Post-Training Inter- and Intra-Rater Reliability

After establishing guidelines (Table 1) and additional training, moderate inter-rater reliability was achieved ($\kappa=0.42$, 95% CI 0.37-0.47). Inter-rater reliability showed statistically significant improvements from pre- to post-training ($t=3.46$, $P < .001$); however, post-training reliability did not reach prespecified acceptable levels (Fleiss' $\kappa \geq 0.80$). An item analysis revealed that raters achieved complete agreement on 34% of clips, agreement in three out of four raters on 41%, and agreement in two out of four raters on 25%. On videos with disagreements limited to two responses, the most common disagreements were between present versus unable to rate (46%), compared to absent versus unable to rate (32%) or present versus absent (22%). Overall intra-rater reliability was moderate (ICC = 0.47), ranging from moderate (Cohen's $\kappa=0.49$) to almost perfect (Cohen's $\kappa=0.85$) agreement.

Discussion

This study examined the inter- and intra-rater reliability for judging the LAR with the touch method of laryngeal sensation testing during FEES. Despite establishing guidelines for rating the presence or absence of the LAR in a large sample of clinical videos and participating in two training sessions, we were unable to achieve acceptable reliability between raters. The consensus training sessions did significantly improve reliability compared to pre-training ratings, though post-training reliability did not meet prespecified acceptable levels. Sufficient intra-rater reliability was

achieved in some raters, suggesting that ratings may remain consistent within raters over a short period of time. However, the accuracy of these ratings remains unclear.

Though studies have not formally reported reliability for the touch method, a recent study by Kaneoka et al³⁴ reported raw data on judgments of the LAR from two separate raters. Disagreements were resolved by a third rater. In this study, LAR ratings during the touch method on healthy adults were reported across 48 trials. When trial-by-trial ratings were compared between raters, moderate reliability was achieved (Cohen's $\kappa=0.44$, 95% CI 0.05-0.83, 85% agreement). Though not formally reported in the original study, this inter-rater reliability value closely aligns with our findings (Fleiss' $\kappa=0.42$; Cohen's κ range=0.39-0.57).

There are considerations inherent in the administration and interpretation of the touch method of laryngeal sensation testing which might further explain the low levels of reliability found in this study. The touch method requires the clinician to briefly touch the arytenoid and then withdraw the endoscope so that adduction of the arytenoids or vocal folds can be visualized. Given that the LAR occurs within a short time frame, this swift endoscopic movement must ensure that both adequate contact and visualization of the glottis is achieved. This method of sensation testing requires instruction beyond what is typically required during a FEES. Additional patient factors, such as vocal fold adduction during testing from coughing, speaking, or hyperfunctional breathing patterns, can further obfuscate rating the LAR. These factors likely contributed to the high proportion of disagreement in this study between rating the LAR as "present" versus "unable to rate" in both pre- and post-training sessions. Videos were deliberately chosen from routine clinical exams in order to include attempts in suboptimal conditions. Raters potentially had different internal criteria to judge whether an attempt was not sufficient to rate as "present," despite providing explicit guidelines post-training. This underscores the complexity of rating the LAR under suboptimal conditions and the importance of obtaining multiple attempts to elicit the LAR with this method. Future studies examining LAR reliability would benefit from additional guidelines to better identify low quality attempts as unable to be rated.

Another method of laryngeal sensation testing, the air pulse method, reconciles many of these limitations. This method allows for quantification of the severity of sensory deficits based on the intensity of the air pulse stimulus, defined as either normal (<4 mmHg), moderate (4-6 mmHg), or severe (>6 mmHg). The air pulse method provides a consistent, clear view of the larynx during testing and does not require movement of the endoscope. However, there are some limitations with the air pulse method, including device generated noise, poor stimulus reproducibility, limited stimulus range, and a lack of commercially manufactured equipment.⁴³ Though equipment has since been developed to

Table 3. Intra-Rater Reliability.

	Pre-Training			Post-Training		
	Cohen's κ	95% CI	% Agreement	Cohen's κ	95% CI	% Agreement
Rater A	0.53	0.34-0.71	71%	0.65	0.49-0.82	77%
Rater B	0.89	0.79-0.99	94%	0.85	0.73-0.97	90%
Rater C	0.73	0.55-0.91	87%	0.61	0.44-0.78	74%
Rater D	0.35	0.09-0.61	71%	0.49	0.30-0.68	68%

address these limitations,⁴⁴⁻⁴⁶ a major benefit of the touch method is that it can be performed during routine FEES exams and does not require specialized equipment.

Few studies have directly compared the air pulse and touch methods. Kaneoka and colleagues¹⁹ showed that the air pulse method identified laryngeal sensory deficits at greater frequency than the touch method in healthy adults and patients with head and neck cancer and Parkinson's disease. Additionally, these investigators found that the air pulse method was not correlated with abnormal penetration-aspiration scale scores,⁴⁷ whereas the touch method did show an association. However, it should be noted that the distribution of penetration-aspiration scale scores in that study's sample did not include more severe scores (>4) or patients who aspirated. In another study, Cuellar and Harvey¹⁶ examined both methods' ability to predict sensory responses to airway invasion and pharyngeal residue. Though the authors conclude that LAR testing alone is unable to predict sensory impairments, the authors did not include a spontaneous swallow as a normal sensory response to penetration despite research suggesting that stimulation above the level of the true vocal folds results in a swallow.^{6,7} Future studies comparing the two methods in patients with severe deficits in airway protection are necessary to better elucidate their clinical utility.

There are several limitations of the study's design that warrant discussion. First, raters assessed the LAR during a single attempt of tactile stimulation of an arytenoid. Though this design provided precise analysis of rater judgments, this is not common in clinical practice. During FEES, clinicians often perform multiple attempts to elicit the LAR. Once a single positive response is elicited, the LAR is judged as at least unilaterally present. It is unclear whether prior studies have rated the LAR during single attempts or over the course of an entire exam. We suspect that reliability would improve if raters were provided the entire laryngeal sensation testing exam, as opposed to isolated attempts. The inclusion of additional types of reflexes (eg, cough, swallow, gag) would likely improve reliability, though these responses do not consistently occur in healthy adults during laryngeal sensation testing compared to the LAR.³⁴ Similarly, including the entire exam rather than a clip of a single attempt may provide more context to the raters to bolster confidence and reliability in ratings. Secondly,

laryngeal sensation testing was performed at the end of FEES. In case the patient did not tolerate the endoscopic exam well, bolus administration was prioritized in the clinical exams and therefore sensory testing was carried out at the end of the exam. It is possible that there was residue from bolus trials, which might obscure one's view during testing. Clinicians also rely on proprioceptive feedback during sensation testing to determine if the endoscope adequately contacted the arytenoid. This type of feedback was not available to raters. Thirdly, guidelines were formulated based on clinical experience with the touch method. The decision was made to define a present LAR as "robust" (ie, more complete, quick adduction) in order to better delineate unclear cases, though there are likely variations in adduction patterns of the arytenoids or true vocal folds depending on the integrity of the recurrent laryngeal nerve, peripheral trauma such as edema, or potential subtle variations in a normal LAR. Future research using two endoscopes to confirm the presence or absence of the LAR (ie, one endoscope providing the touch stimulus and the other observing in home position) is needed to formulate more accurate judgments and establish detailed and rigorous guidelines to improve reliability with this method. Finally, it should be noted that absolute agreement is an unforgiving standard and the a priori threshold for "acceptable" reliability (Fleiss' $\kappa \geq 0.80$) might be unrealistic given the inherent complexity and limitations of the touch method. Future investigations might benefit from methodologies to resolve discrepancies between raters, such as a consultation from a third rater or consensus panel.

Conclusion

This study was unable to demonstrate adequate reliability between raters when judging the LAR with the touch method. Sufficient intra-rater reliability was achieved in some raters, suggesting that ratings may remain consistent within raters over a short period of time. Reliability reporting and specific definitions when performing and judging the LAR are encouraged to promote transparency, reproducibility, and sound methodological practices. Future studies examining reliability of the touch method of laryngeal sensation testing are required to better understand the method's utility in clinical practice and research.

Authors' Note

Portions of this manuscript were presented at the European Society for Swallowing Disorders in Vienna, Austria on September 19, 2019.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study was deemed exempt by IRB as it did not constitute research with human subjects.

Informed Consent

The regulatory requirement for a consent form does not apply to exempt research.

ORCID iD

James C. Borders  <https://orcid.org/0000-0002-0291-4266>

References

1. Hammer MJ, Barlow SM. Laryngeal somatosensory deficits in Parkinson's disease: implications for speech respiratory and phonatory control. *Exp Brain Res*. 2010;201(3):401-409.
2. Hammer MJ, Murphy CA, Abrams TM. Airway somatosensory deficits and dysphagia in Parkinson's disease. *J Parkinsons Dis*. 2013;3(1):39-44.
3. Steele CM, Miller AJ. Sensory input pathways and mechanisms in swallowing: a review. *Dysphagia*. 2010;25(4):323-333.
4. de Carlos F, Cobo J, Macías E, et al. The sensory innervation of the human pharynx: searching for mechanoreceptors. *Anat Rec (Hoboken)*. 2013;296(11):1735-1746.
5. Domer AS, Kuhn MA, Belafsky PC. Neurophysiology and clinical implications of the laryngeal adductor reflex. *Curr Otorhinolaryngol Rep*. 2013;1(3):178-182.
6. Jafari S, Prince RA, Kim DY, Paydarfar D. Sensory regulation of swallowing and airway protection: a role for the internal superior laryngeal nerve in humans. *J Physiol*. 2003;550(Pt 1):287-304.
7. Barkmeier JM, Bielamowicz S, Takeda N, Ludlow CL. Modulation of laryngeal responses to superior laryngeal nerve stimulation by volitional swallowing in awake humans. *J Neurophysiol*. 2000;83(3):1264-1272.
8. Phua SY, McGarvey LP, Ngu MC, Ing AJ. Patients with gastro-oesophageal reflux disease and cough have impaired laryngopharyngeal mechanosensitivity. *Thorax*. 2005;60(6):488-491.
9. Phua SY, McGarvey L, Ngu M, Ing A. The differential effect of gastroesophageal reflux disease on mechanostimulation and chemostimulation of the laryngopharynx. *Chest*. 2010;138(5):1180-1185.
10. Aviv JE, Liu H, Parides M, Kaplan ST, Close LG. Laryngopharyngeal sensory deficits in patients with laryngopharyngeal reflux and dysphagia. *Ann Otol Rhinol Laryngol*. 2000;109(11):1000-1006.
11. Nguyen AT, Jobin V, Payne R, Beauregard J, Naor N, Kimoff RJ. Laryngeal and velopharyngeal sensory impairment in obstructive sleep apnea. *Sleep*. 2005;28(5):585-593.
12. Clayton NA, Carnaby GD, Peters MJ, Ing AJ. Impaired laryngopharyngeal sensitivity in patients with COPD: the association with swallow function. *Int J Speech Lang Pathol*. 2014;16(6):615-623.
13. Clayton NA, Carnaby-Mann GD, Peters MJ, Ing AJ. The effect of chronic obstructive pulmonary disease on laryngopharyngeal sensitivity. *Ear Nose Throat J*. 2012;91(9):370-372.
14. Payne RJ, Kost KM, Frenkiel S, et al. Laryngeal inflammation assessed using the reflux finding score in obstructive sleep apnea. *Otolaryngol Head Neck Surg*. 2006;134(5):836-842.
15. Onofri SM, Cola PC, Berti LC, da Silva RG, Dantas RO. Correlation between laryngeal sensitivity and penetration/aspiration after stroke. *Dysphagia*. 2014;29(2):256-261.
16. Cuellar ME, Harvey J. Predictive value of laryngeal adductor reflex testing in patients with dysphagia due to a cerebral vascular accident. *Int J Speech Lang Pathol*. 2019;21(6):593-601.
17. Borders JC, Fink D, Levitt JE, et al. Relationship between laryngeal sensation, length of intubation, and aspiration in patients with acute respiratory failure. *Dysphagia*. 2019;34(4):521-528.
18. Sasaki CT, Suzuki M, Horiuchi M, Kirchner JA. The effect of tracheostomy on the laryngeal closure reflex. *Laryngoscope*. 1977;87(9 Pt 1):1428-1433.
19. Kaneoka A, Krisciunas GP, Walsh K, Raade AS, Langmore SE. A comparison of 2 methods of endoscopic laryngeal sensory testing: a preliminary study. *Ann Otol Rhinol Laryngol*. 2015;124(3):187-193.
20. Amin MR, Harris D, Cassel SG, Grimes E, Heiman-Patterson T. Sensory testing in the assessment of laryngeal sensation in patients with amyotrophic lateral sclerosis. *Ann Otol Rhinol Laryngol*. 2006;115(7):528-534.
21. Maruo T, Fujimoto Y, Ozawa K, et al. Laryngeal sensation and pharyngeal delay time after (chemo)radiotherapy. *Eur Arch Otorhinolaryngol*. 2014;271(8):2299-2304.
22. Schindler A, Ginocchio D, Peri A, Felisati G, Ottaviani F. FEESST in the rehabilitation of dysphagia after partial laryngectomy. *Ann Otol Rhinol Laryngol*. 2010;119(2):71-76.
23. Murry T, Branski RC, Yu K, Cukier-Blaj S, Duflo S, Aviv JE. Laryngeal sensory deficits in patients with chronic cough and paradoxical vocal fold movement disorder. *Laryngoscope*. 2010;120(8):1576-1581.
24. Thompson DM. Abnormal sensorimotor integrative function of the larynx in congenital laryngomalacia: a new theory of etiology. *Laryngoscope*. 2007;117(6 Pt 2 suppl 114):1-33.

25. Ulualp S, Brown A, Sanghavi R, Rivera-Sanchez Y. Assessment of laryngopharyngeal sensation in children with dysphagia. *Laryngoscope*. 2013;123(9):2291-2295.
26. Aviv JE, Spitzer J, Cohen M, Ma G, Belafsky P, Close LG. Laryngeal adductor reflex and pharyngeal squeeze as predictors of laryngeal penetration and aspiration. *Laryngoscope*. 2002;112(2):338-341.
27. Setzen M, Cohen MA, Mattucci KF, Perlman PW, Ditkoff MK. Laryngopharyngeal sensory deficits as a predictor of aspiration. *Otolaryngol Head Neck Surg*. 2001;124(6):622-624.
28. Link DT, Willging JP, Miller CK, Cotton RT, Rudolph CD. Pediatric laryngopharyngeal sensory testing during flexible endoscopic evaluation of swallowing: feasible and correlative. *Ann Otol Rhinol Laryngol*. 2000;109(10 Pt 1):899-905.
29. Kaneoka A, Pisegna JM, Inokuchi H, et al. Relationship between laryngeal sensory deficits, aspiration, and pneumonia in patients with dysphagia. *Dysphagia*. 2018;33(2):192-199.
30. Aviv JE, Sacco RL, Mohr JP, et al. Laryngopharyngeal sensory testing with modified barium swallow as predictors of aspiration pneumonia after stroke. *Laryngoscope*. 1997;107(9):1254-1260.
31. Aviv JE, Kim T, Sacco RL, et al. FEESST: a new bedside endoscopic test of the motor and sensory components of swallowing. *Ann Otol Rhinol Laryngol*. 1998;107(5 Pt 1):378-387.
32. Langmore SE, Schatz K, Olsen N. Fiberoptic endoscopic examination of swallowing safety: a new procedure. *Dysphagia*. 1988;2(4):216-219.
33. Aviv JE, Kim T, Thomson JE, Sunshine S, Kaplan S, Close LG. Fiberoptic endoscopic evaluation of swallowing with sensory testing (FEESST) in healthy controls. *Dysphagia*. 1998;13(2):87-92.
34. Kaneoka A, Pisegna JM, Krisciunas GP, et al. Variability of the pressure measurements exerted by the tip of laryngoscope during laryngeal sensory testing: a clinical demonstration. *Am J Speech Lang Pathol*. 2017;26(3):729-736.
35. Jensen K, Lambertsen K, Grau C. Late swallowing dysfunction and dysphagia after radiotherapy for pharynx cancer: frequency, intensity and correlation with dose and volume parameters. *Radiother Oncol*. 2007;85(1):74-82.
36. Shapira-Galitz Y, Shoffel-Havakuk H, Halperin D, Lahav Y. Association between laryngeal sensation, pre-swallow secretions and pharyngeal residue on fiberoptic endoscopic examination of swallowing. *Dysphagia*. 2019;34(4):548-555.
37. Sinclair CF, Téllez MJ, Ulkatan S. Human laryngeal sensory receptor mapping illuminates the mechanisms of laryngeal adductor reflex control. *Laryngoscope*. 2018;128(11):E365-E370.
38. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) — a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-381.
39. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2018.
40. Lo Martire R. rel: Reliability coefficients. R package version 1.3.1. 2017. <https://CRAN.R-project.org/package=rel>
41. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
42. Gwet KL. Testing the difference of correlated agreement coefficients for statistical significance. *Educ Psychol Meas*. 2016;76(4):609-637.
43. Hammer MJ. Design of a new somatosensory stimulus delivery device for measuring laryngeal mechanosensory detection thresholds in humans. *IEEE Trans Biomed Eng*. 2009;56(4):1154-1159.
44. Giraldo-Cadavid LF, Agudelo-Otalora LM, Burguete J, et al. Design, development and validation of a new laryngo-pharyngeal endoscopic esthesiometer and range-finder based on the assessment of air-pulse variability determinants. *Biomed Eng Online*. 2016;15(1):52.
45. Giraldo-Cadavid LF, Burguete J, Rueda F, et al. Reliability of a laryngo-pharyngeal esthesiometer and a method for measuring laryngo-pharyngeal mechano-sensitivity in a prospectively recruited cohort of patients. *Eur Arch Otorhinolaryngol*. 2017;274(7):2861-2870.
46. Giraldo-Cadavid LF, Burguete J, Rueda F, et al. Accuracy of a laryngopharyngeal endoscopic esthesiometer (LPEER) for evaluating laryngopharyngeal mechanosensitivity: a validation study in a prospectively recruited cohort of patients. *Dysphagia*. 2018;33(1):15-25.
47. Rosenbek JC, Robbins JA, Roecker EB, Coyle JL, Wood JL. A penetration-aspiration scale. *Dysphagia*. 1996;11(2):93-98.