

Standardizing Laryngeal Cleft Evaluations: Reliability of the Interarytenoid Assessment Protocol

Steven Coppess, JD, MBA¹, Reema Padia, MD², David Horn, MD², Sanjay R. Parikh, MD², Andrew Inglis, MD², Randall Bly, MD², John Dahl, MD, PhD, MBA², Daniel Dudley¹, and Kaalan Johnson, MD²

Otolaryngology—
 Head and Neck Surgery
 2019, Vol. 160(3) 533–539
 © American Academy of
 Otolaryngology–Head and Neck
 Surgery Foundation 2018
 Reprints and permission:
sagepub.com/journalsPermissions.nav
 DOI: 10.1177/0194599818806283
<http://otojournal.org>



Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

Abstract

Objective. While the Benjamin-Inglis classification system is widely used to categorize laryngeal clefts, it does not clearly differentiate a type I cleft from normal anatomy, and there is no widely accepted or validated protocol for systematically evaluating interarytenoid mucosal height. We sought to propose the interarytenoid assessment protocol as a method to standardize the description of the interarytenoid anatomy and to test its reliability.

Study Design. Retrospective review of endoscopic videos.

Setting. Pediatric academic center.

Subjects and Methods. The interarytenoid assessment protocol comprises 4 steps for evaluation of the interarytenoid region relative to known anatomic landmarks in the supraglottis, glottis, and subglottis. Thirty consecutively selected videos of the protocol were reviewed by 4 otolaryngologists. The raters were blinded to identifying information, and the video order was randomized for each review. We assessed protocol completion times and calculated Cohen's linear-weighted κ coefficient between blinded expert raters and with the operating surgeon to evaluate interrater/intrarater reliability.

Results. Median age was 4.9 years (59 months; range, 1 month to 20 years). Median completion time was 144 seconds. Interrater and intrarater reliability showed substantial agreement (interrater $\kappa = 0.71$ [95% confidence interval (CI), 0.55–0.87]; intrarater mean $\kappa = 0.70$ [95% CI, 0.59–0.92/rater 1, 0.47–0.85/rater 2]; $P < .001$). Comparing raters to the operating surgeon demonstrated substantial agreement (mean $\kappa = 0.62$; 95% CI, 0.31–0.79/rater 1, 0.48–0.89/rater 2; $P < .001$).

Conclusion. The interarytenoid assessment protocol appears reliable in describing interarytenoid anatomy. Rapid completion times and substantial interrater/intrarater reliability were demonstrated. Incorporation of this protocol may provide important steps toward improved standardization in

the anatomic description of the interarytenoid region in pediatric dysphagia.

Keywords

laryngeal cleft, type I laryngeal cleft, interarytenoid, dysphagia, interarytenoid assessment protocol

Received May 14, 2018; revised August 21, 2018; accepted September 20, 2018.

Traditionally, the Benjamin-Inglis and Evan's classification systems have been used for diagnosis of type I laryngeal cleft (T1C) but do not clearly identify the superior anatomic limit for T1C.^{1–4} The International Pediatric Otolaryngology Group: Consensus guidelines on the diagnosis and management of type I laryngeal clefts highlights a key decision point in the evaluation process depending on diagnosis of T1C or deep interarytenoid groove vs normal anatomy.¹ Furthermore, there is not consistency in distinguishing T1C from a deep interarytenoid groove.¹ This leads to possible controversy when classifying minor laryngeal clefts and leaves us without a clear understanding of normative anatomy. The lack of a standardized descriptive anatomical classification has led to the use of a variety of evaluation techniques in rigid endoscopic assessment of the interarytenoid region, including a variety of protocols and even novel evaluation devices.⁵ Without a standardized measure of the subtle anatomic differences

¹University of Washington School of Medicine, University of Washington, Seattle, Washington, USA

²Division of Pediatric Otolaryngology–Head and Neck Surgery, Seattle Children's Hospital, Seattle, Washington, USA

This paper was a podium presentation at the American Society of Pediatric Otolaryngology Annual Meeting; April 21, 2018; National Harbor, Maryland.

Corresponding Author:

Kaalan Johnson, MD, Division of Pediatric Otolaryngology–Head and Neck Surgery, Seattle Children's Hospital, MS OA.9.321, PO Box 5371, Seattle, WA 98145-5005, USA.

Email: kaalan.johnson@seattlechildrens.org

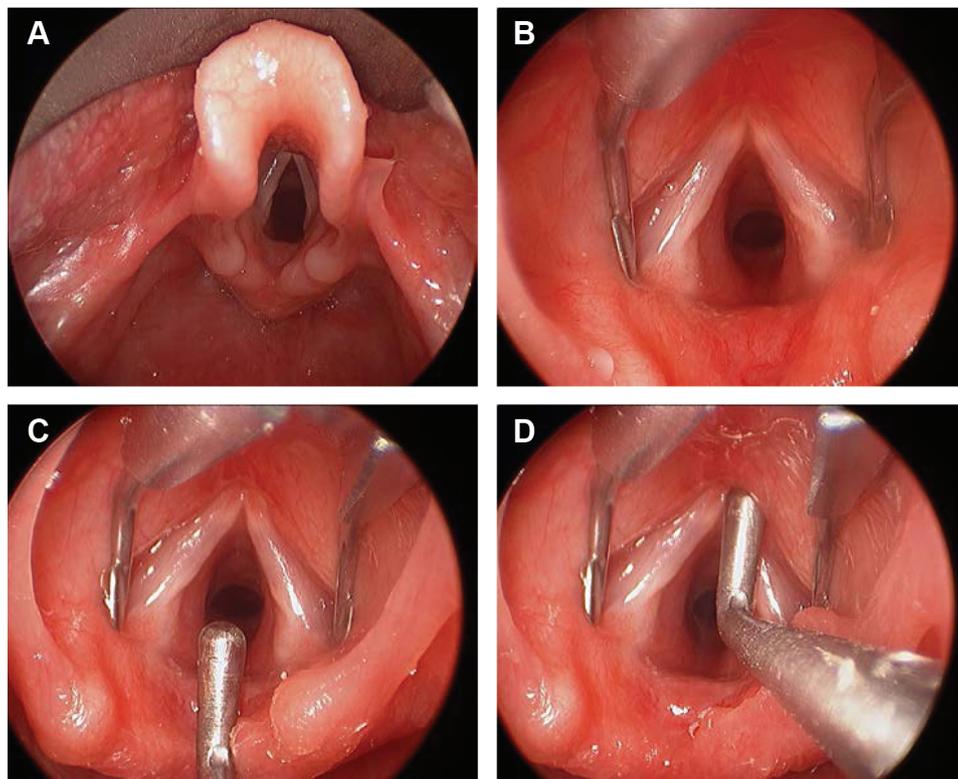


Figure 1A: Step 1 is to establish suspension laryngoscopy.
 Figure 1B: Step 2 is to insert vocal fold distenders, generally infant spreaders upside down to one click.
 Figure 1C: Step 3 is to palpate the interarytenoid musculature to assess for submucous laryngeal cleft, cricoid defect, and bulk of musculature.
 Figure 1D: Step 4 is to evaluate the depth of the interarytenoid mucosa using a swinging hook method to determine the anatomic reference.

Figure 1. Interytenoid assessment protocol.

seen in TIC, we will not be able to establish a correlation between anatomy and other measures of dysphagia, such as fluoroscopic swallow studies and fiberoptic endoscopic evaluations of swallowing, as well as response to therapies such as dietary modifications or operative interventions such as injection laryngoplasty or laryngeal cleft repairs.⁶

Previous work at our institution demonstrated similar benefit to interarytenoid augmentation in patients with or without a TIC, raising the question of whether understanding the complex physiology of dysphagia may be more important than purely the anatomy, but those types of questions cannot be fully explored without a clear definition of the anatomy.⁷ Our proposed interarytenoid assessment protocol (IAAP) provides a standardized descriptive anatomical classification of the interarytenoid region by using specific evaluation techniques. This study tests the reliability of the proposed IAAP. Reliability is assessed by testing the following hypotheses: (1) that there is a significant difference in scoring of the IAAP between different raters evaluating a video of the same procedure, (2) that there is a significant difference in scoring of the IAAP by a rater evaluating a video of the same procedure on 2 different occasions, and (3) that there is a significant difference in scoring of the

IAAP between raters evaluating a video of the procedure being performed and documented by an operating surgeon.

Materials and Methods

This study used retrospective review of endoscopic videos of pediatric patients listed in a Seattle Children's Hospital otolaryngology database (SCH OTO Database). Institutional review board (IRB) approval was obtained at the Seattle Children's Hospital (IRB No. STUDY00000815).

IAAP

The protocol was developed with the goal of providing standardized, efficient, and comprehensive evaluations of the interarytenoid region in patients. The protocol was created using a modified Delphi approach, with input from the attending otolaryngologist authors at Seattle Children's Hospital.⁸ The proposed steps of the IAAP are as follows:

- Step 1: Insert and suspend laryngoscope.** This should be performed to optimize a Cormack-Lehane grade I exposure if possible (**Figure 1A**).⁹
- Step 2: Insert laryngeal distending forceps.** Infant distenders are best used for children up to

approximately age 5 years, and then adult distenders may be used in older patients.⁹ The larynx should be distended to allow adequate visualization of the height of the interarytenoid mucosa, while not overdistending to avoid distorting the anatomy. Appropriate distension will position the vocal folds at an approximate 30-degree angle—most pediatric patients only require distension to the first “click” after touching the tissue (**Figure 1B**).

Step 3: Palpate interarytenoid area. Perform palpation using a nerve hook, right-angle laryngeal probe, or right-angle laryngeal measuring stick. Palpate and observe bulk of interarytenoid muscle anteroposterior and craniocaudal extent. Palpate surface of cricoid to observe for clefts or defect. Check for submucosal laryngeal clefts (**Figure 1C**).

Step 4: Objective assessment of interarytenoid mucosal height (IAMH). With the right hand carefully braced on the patient’s face or the laryngoscope to prevent movement in or out, use the right-angle laryngeal probe to swing anterolaterally from a resting place in contact with but not distorting the interarytenoid mucosa to an anterolateral point in line with the midpoint of the membranous true vocal folds (TVFs) to determine anatomical reference to the following laryngeal landmarks: (a) above false vocal folds (FVFs), (b) at FVFs, (c) in ventricle, (d) at TVFs, and (e) below TVFs (**Figure 1D**).

Reliability Testing

Sample selection. Reliability of the IAAP was tested with review of endoscopic videos from the SCH OTO Database. Due to variation in prevalence of each IAMH and the desire to obtain ratings for each IAMH among the sample, we set a random target number of videos to be collected from each IAMH, as rated at the initial procedure. From the database, only 1 video was available from the IAMH level below TVFs and 2 for at TVFs, so these were all included. A random-number generator was used to select a number between 9 and 12 for the most prevalent levels above FVFs and at FVFs. The in-ventricle target number was set so that the remaining videos obtained would sum to 30 videos, a total sample number felt to be appropriate for reliability testing during statistical consultation. This process was undertaken so that reviewers would not be able to predict the expected number of videos to assign to each IAMH. Thirty videos were then selected consecutively for each IAMH until the target number of videos was obtained. Inclusion criteria for videos required sufficient video quality to visualize the procedure and limited to the target number for each IAMH. A total of 115 records were reviewed for inclusion consecutively, with 32 of the records’ videos excluded due to inadequate visualization of the steps of the IAAP, 8 videos excluded due to poor quality, and 45 records’ videos excluded due to our having already filled

the target number for the respective IAMH. In reliability assessment of raters against the operating surgeon, 1 additional rating was excluded due to the operating surgeon’s report not identifying a specific IAMH.

Reliability was assessed in 2 sessions by attending otolaryngologists at Seattle Children’s Hospital: primarily for 2 naive raters without any training and secondarily for 2 additional raters who received training on the scoring system by providing the IAAP Scoring Sheet (**Figure 2**). The raters who received additional training were also compared with the operating surgeons who performed the protocol. All raters were blind to identifying information and how many videos were selected for each IAMH. The initial raters were only provided with the scoring criteria as listed in step 4 above and no additional instructions or training. The subsequent raters were provided with the IAAP Scoring Sheet with examples of the appearance of the hook at the different IAMHs and the instructions to only evaluate the anatomy as demonstrated by the surgeon without attempting to make independent assessments of the surgical technique. The addition of the IAAP Scoring Sheet for the second 2 raters was to determine if additional education about the IAAP is needed to improve reliability. They each rated the IAMH for the 30 videos on 2 separate occasions separated by at least 27 days, totaling 4 reviews. The videos were edited to only show the IAAPs being performed and no additional procedures. In addition, the order of the videos reviewed by the raters was randomized on both occasions, and the raters did not have access to their first video review when reviewing the second time.

Interrater and intrarater reliability data were assessed using a linearly weighted Cohen’s κ coefficient for each set of reviews. For interrater reliability assessment, we calculated the results based on the initial scorings of each pair of surgeons (eg, the first pair’s initial scorings against each other and the second pair’s initial scorings against each other). For intrarater reliability assessment, we calculated the results based on the scorings of each surgeon (eg, first scoring against second scoring) and reported a mean of the κ coefficient for each pair (eg, mean of surgeons from first pair and second pair, respectively). Assessment of Cohen’s linear weighted κ coefficient scale of agreement was defined with the following categories: <0, less than chance; 0.01 to 0.20, slight; 0.21 to 0.40, fair; 0.41 to 0.60, moderate; 0.61 to 0.80, substantial; and 0.81 to 0.99, almost perfect agreement.¹⁰ We used RStudio statistical software for our analysis, with a cutoff for significance of $P < .05$.¹¹

Start to finish completion times were measured from insertion of the distenders to removal of the measuring device. A total of 9 videos were used for measurement of completion time, due to the remaining 21 videos not including visualization of the start of the insertion of the distenders.

Results

A total of 30 endoscopic videos were collected for patients with a median age of 4.9 years (59 months; range, 1 month to 20 years), where 30% were female (7 years; range, 3.2

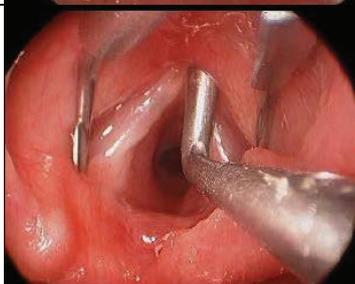
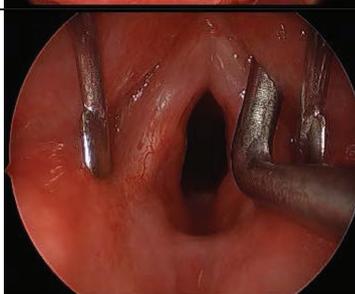
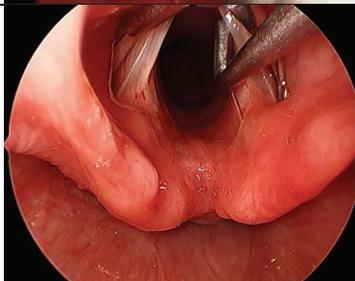
| Level | Sample Picture | Description |
|-----------------------------|---|---|
| Above False Vocal Folds (1) |  | The instrument touches anywhere above the medial aspect of the false vocal fold. |
| At False Vocal Folds (2) |  | The instrument touches the medial aspect of the false vocal fold. |
| In Ventricle (3) |  | The instrument lands anywhere within the ventricle, below the false vocal fold and above the true vocal fold. |
| At True Vocal Folds (4) |  | The instrument touches the medial aspect of the true vocal fold. |
| Below True Vocal Folds (5) |  | The instrument lands below the true vocal fold. |

Figure 2. Teaching guide: definitions of the interarytenoid mucosal height for the interarytenoid assessment protocol.

months to 20 years) and 70% were male (4 years; range, 1.1 months to 13.2 years). Median completion time of the protocol was 144 seconds.

Four fellowship-trained pediatric otolaryngologists participated in the IAAP reliability assessment. In both groups of raters, with and without instruction, interrater reliability and intrarater reliability showed substantial agreement ($\kappa = 0.61-0.80$). Without instruction, the interrater $\kappa = 0.65$ (95%

confidence interval [CI], 0.47-0.82; $P < .001$), and intrarater mean $\kappa = 0.62$ (95% CI, 0.52-0.86/rater 1, 0.34-0.74/rater 2; $P < .001$). With instruction, after incorporating the IAAP Scoring Sheet, interrater $\kappa = 0.71$ (95% CI, 0.55-0.87; $P < .001$) (**Table 1**), and the intrarater reliability also showed substantial agreement, with intrarater mean $\kappa = 0.70$ (95% CI, 0.59-0.92/rater 1, 0.47-0.85/rater 2; $P < .001$) (**Tables 2 and 3**). When comparing 29 videos of the

Table 1. Interrater Reliability: Ratings Comparison with the Interarytenoid Assessment Protocol Scoring Sheet (n = 30).

| Rater 2 | Rater 1, No. | | | | |
|-----------|--------------|--------|-----------|--------|-----------|
| | Above FVF | At FVF | Ventricle | At TVF | Below TVF |
| Above FVF | 11 | 4 | | | |
| At FVF | 1 | 8 | 1 | | |
| Ventricle | | | 3 | | |
| At TVF | | | 1 | 0 | 1 |
| Below TVF | | | | | 0 |

Abbreviations: FVF, false vocal fold; TVF, true vocal fold.

Table 2. Intrarater Reliability (Rater 1): Ratings Comparison with the Interarytenoid Assessment Protocol Scoring Sheet (n = 30).

| Rater 1 | Rater 1, No. | | | | |
|-----------|--------------|--------|-----------|--------|-----------|
| | Above FVF | At FVF | Ventricle | At TVF | Below TVF |
| Above FVF | 14 | 1 | | | |
| At FVF | 3 | 6 | 1 | | |
| Ventricle | | 1 | 2 | | |
| At TVF | | | | 1 | 1 |
| Below TVF | | | | | 0 |

Abbreviations: FVF, false vocal fold; TVF, true vocal fold.

Table 3. Intrarater Reliability (Rater 2): Ratings Comparison with the Interarytenoid Assessment Protocol Scoring Sheet (n = 30).

| Rater 2 | Rater 2, No. | | | | |
|-----------|--------------|--------|-----------|--------|-----------|
| | Above FVF | At FVF | Ventricle | At TVF | Below TVF |
| Above FVF | 7 | 4 | | | |
| At FVF | 1 | 10 | 2 | | |
| Ventricle | | 1 | 4 | | |
| At TVF | | | | 0 | |
| Below TVF | | | | 1 | 0 |

Abbreviations: FVF, false vocal fold; TVF, true vocal fold.

second set of raters with instruction to the operating surgeon's report, the interrater reliability demonstrated substantial agreement, with mean interrater $\kappa = 0.62$ (95% CI, 0.31-0.79/rater 1, 0.48-0.89/rater 2; $P < .001$) (Tables 4 and 5).

Discussion

Our findings demonstrate that the proposed IAAP for evaluation and classification of the interarytenoid area for patients undergoing airway evaluations was possible when attempted in a busy academic practice. The steps outlined in the protocol provide a clear and complete outline for describing the interarytenoid anatomy, which will be critical

Table 4. Interrater Reliability (Rater 1 with Instruction) to Operating Surgeon (n = 29).

| Operating Surgeon | Rater 1, No. | | | | |
|-------------------|--------------|--------|-----------|--------|-----------|
| | Above FVF | At FVF | Ventricle | At TVF | Below TVF |
| Above FVF | 7 | 3 | | | |
| At FVF | 3 | 6 | | | |
| Ventricle | 2 | 2 | 3 | | |
| At TVF | | | 2 | 0 | |
| Below TVF | | | | | 1 |

Abbreviations: FVF, false vocal fold; TVF, true vocal fold.

Table 5. Interrater Reliability (Rater 2 with Instruction) to Operating Surgeon (n = 29).

| Operating Surgeon | Rater 2, No. | | | | |
|-------------------|--------------|--------|-----------|--------|-----------|
| | Above FVF | At FVF | Ventricle | At TVF | Below TVF |
| Above FVF | 11 | | | | |
| At FVF | 1 | 7 | | | |
| Ventricle | 2 | 2 | 3 | | |
| At TVF | | 1 | | 1 | |
| Below TVF | | | | 1 | 0 |

Abbreviations: FVF, false vocal fold; TVF, true vocal fold.

to better differentiate between laryngeal clefts and normal interarytenoid anatomy.

Our findings demonstrate that the proposed IAAP is reliable, with interrater and intrarater reliability showing substantial agreement with little instruction and with the IAAP Scoring Sheet. In the first iteration of reviews, the raters noted they were not clear on which rating to choose for situations that appeared close to the margin of each IAMH. When providing instruction to evaluate the anatomy being demonstrated and the IAAP Scoring Sheet, interrater and intrarater reliability showed improvement in the κ coefficient; however, the agreement remained in the substantial category. While providing the IAAP Scoring Sheet may have been helpful, it did not significantly improve reliability and therefore does not appear to be a requirement for a provider to have before performing the protocol. In addition, comparing raters against the operating surgeon demonstrated substantial agreement, indicating that even when considering potential differences in tactile appreciation or improved procedural depth perception of the operating surgeon, the protocol appears reliable. Interestingly, there was slightly greater variability when comparing blinded video review results with the operating surgeons' results, and the expected IAMHs varied up to 2 categories in certain cases, where the video reviewers only differed among themselves by a maximum of 1 category. In these cases, the operating surgeon always graded the mucosa deeper than the video

reviewers. This could be interpreted as a limitation of the video review compared to palpation or that there may have been some bias involved with the ratings elected by the operating surgeon based on clinical or other factors at the time of the procedure.

Other common grading systems and procedures that are used by otolaryngologists have shown similar κ coefficients when reliability was tested.¹²⁻¹⁵ The need for a more standardized way to measure the interarytenoid region is evident by the ambiguity in the literature of exactly what constitutes a “deep interarytenoid notch/groove” and what is normal anatomy.^{1,14} The IAAP is a technique that accomplishes this while using simple tools that are frequently found in laryngoscopy sets.⁷ Four separate raters were included in the study to allow for more variability in the scoring and help strengthen our results. Giving at least 27 days in between each review and randomizing the order with each review helped limit rater recall bias. Having separate raters perform the rating with the IAAP Scoring Sheet on the second round helped appreciate the effect of the IAAP Scoring Sheet on the IAAP.

A normal IAMH is still unknown and cannot be extrapolated using the data from this study. While other studies have attempted to identify a specific measurement that constitutes a normal IAMH, there is a lack of sufficient evidence to correlate these measurements to a clinical phenotype.^{5,14,15} The IAAP specifically delineates IAMHs above the traditional TIC but includes IAMHs at and below the level of the true vocal folds, which is the current working definition we are using at our institution for a TIC. Future studies that use the IAAP would allow for more accurate comparisons of the IAMH and other independent variables to clinical outcomes, and ultimately a clinically relevant definition for laryngeal cleft could potentially be further refined from our current working definition.

This study was limited by its retrospective nature with all data relying on documentation in a quality improvement database that may have been incomplete. Raters were also observing videos of providers performing the IAAP instead of performing the IAAP themselves, which limits realism. The inability to have the palpable sense during the procedure and to be able to adjust the vantage point with the scope in real time to obtain a better view may have altered the scoring of the videos, although this limitation is mitigated by incorporating the comparison of raters against the operating surgeons and this also showing substantial agreement. In addition, the initial surgeons performing the protocol in the rated videos included attendings and trainees. We anticipate the raters’ review may have been more precise had an experienced surgeon performed every protocol, but we believe the inclusion of trainees in performing the protocol makes the study more generalizable. We did not formally assess feasibility in this study, but the protocol uses tools found in many airway endoscopy trays. Steps 1 to 3 of the protocol involve procedures that are used currently by many otolaryngologists for airway evaluations, and the protocol demonstrated low median completion times of 144 seconds.

Implications for Practice

The IAAP appears reliable in describing interarytenoid anatomy with demonstration of substantial interrater and intrarater reliability. The protocol employs readily available tools and accessible techniques, with rapid completion times. Incorporation of the IAAP may improve standardization in the anatomic description of the interarytenoid region in the study of pediatric dysphagia.

Author Contributions

Steven Coppess, substantial contribution to the concept and design; acquisition, analysis, and interpretation of data; drafting and revising manuscript; **Reema Padia**, substantial contribution to the concept and design; acquisition, analysis, and interpretation of data; drafting and revising manuscript; **David Horn**, substantial contribution to the concept and design, revising manuscript; **Sanjay R. Parikh**, substantial contribution to the concept and design, revising manuscript; **Andrew Inglis**, substantial contribution to the concept and design, revising manuscript; **Randall Bly**, substantial contribution to the concept and design, revising manuscript; **John Dahl**, substantial contribution to the concept and design, revising manuscript; **Daniel Dudley**, substantial contribution to the concept and design; acquisition, analysis, and interpretation of data; drafting and revising manuscript; **Kaalan Johnson**, substantial contribution to the concept and design; acquisition, analysis, and interpretation of data; drafting and revising manuscript.

Disclosures

Competing interests: None.

Sponsorships: Seattle Children’s Hospital & University of Washington.

Funding source: None.

References

1. Yeung JC, Balakrishnan K, Cheng ATL, et al; International Pediatric Otolaryngology Group. Consensus guidelines on the diagnosis and management of type I laryngeal clefts. *Int J Pediatr Otorhinolaryngol*. 2017;101:51-56.
2. Benjamin B, Inglis A. Minor congenital laryngeal clefts: diagnosis and classification. *Ann Otol Rhinol Laryngol*. 1989;98:417-420.
3. Ojha S, Ashland JE, Hersh C, Ramakrishna J, Maurer R, Hartnick CJ. Type 1 laryngeal cleft: a multidimensional management algorithm. *JAMA Otolaryngol Head Neck Surg*. 2014;140:34-40.
4. Evans JN. Management of the cleft larynx and tracheoesophageal clefts. *Ann Otol Rhinol Laryngol*. 1985;94:627-630.
5. Smith RJ, Neville MB, Bauman NM. Interarytenoid notch height relative to the vocal folds: pilot study. *Ann Otol Rhinol Laryngol*. 1994;103:753-757.
6. Chien W, Ashland J, Haver K, Hardy SC, Curren P, Hartnick CJ. Type 1 laryngeal cleft: establishing a functional diagnostic and management algorithm. *Int J Pediatr Otorhinolaryngol*. 2006;70:2073-2079.
7. Horn DL, DeMarre K, Parikh SR. Interarytenoid sodium carboxymethylcellulose gel injection for management of pediatric aspiration. *Ann Otol Rhinol Laryngol*. 2014;123:852-858.

8. Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. *Information Management*. 2004;42:15-29.
9. Maturo S, Hartnick CJ. The vocal fold retractor: a useful tool for diagnosis and treatment of laryngeal and tracheal pathology. *Laryngoscope*. 2010;120:2227-2230.
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.
11. RStudio. *Integrated Development Environment for R*. Boston, MA: RStudio; 2015.
12. McCoul ED, Smith TL, Mace JC, et al. Interrater agreement of nasal endoscopy in patients with a prior history of endoscopic sinus surgery. *Int Forum Allergy Rhinol*. 2012;2:453-459.
13. Chan DK, Liming BJ, Horn DL, Parikh SR. A new scoring system for upper airway pediatric sleep endoscopy. *JAMA Otolaryngol Head Neck Surg*. 2014;140:595-602.
14. Larsen KL, Lange B, Darling P, Jørgensen G, Kjeldsen AD. The validity of nasal endoscopy in patients with chronic rhinosinusitis—an inter-rater agreement study. *Clin Otolaryngol*. 2018;43:144-150.
15. Miglani A, Schraff S, Clarke CY, et al. An aerodigestive approach to laryngeal clefts and dysphagia using injection laryngoplasty in young children. *Curr Gastroenterol Rep*. 2017; 19:60.