Reliability and Accuracy of Remote Fiberoptic Nasopharyngolaryngoscopy in the Pediatric Population

Lauren E. Miller, MD, MBA1, Adva Buzi, MD2, Ashley Williams, CRNP2, Rachel S. Rogers, MS3, Angel G. Ortiz, BS3, Kellye O. Jones-Ho, CRNP2, and Lisa M. Elden, MD2

Abstract

Introduction: Telemedicine is an increasingly prevalent component of medical practice. In otolaryngology, there is the potential for telemedicine services to be performed in conjunction with device use, such as with a nasolaryngoscope. This study evaluates the reliability of remote examinations of the upper airway through an iPhone recording using a coupling device attached to a nasopharyngolaryngoscope (NPL).

Methods: A prospective, blinded study was performed for pediatric patients requiring an NPL during an office visit. The NPL was performed using a coupling device attached to a smartphone to record the examination. A second, remote otolaryngologist then evaluated the recorded examination. Both otolaryngologists evaluated findings of anatomic sites including nasopharynx, oropharynx, base of tongue, larynx including subsites of epiglottis, arytenoids, aryepiglottic folds, false vocal cords, true vocal cords, patency of airway, and diagnostic impression, all of which were documented through a survey. Results of the survey were evaluated through inter-rater agreement using the \( \kappa \) statistic.

Results: Forty-five patients underwent an NPL, all of which were included in the study. The average age was 4.9 years. The most common complaint requiring NPL was noisy breathing (n = 16). The inter-rater agreement for overall diagnosis was 0.74 with 80% percent agreement, rated as “good.” Other anatomic subsites with “good” or better inter-rater agreement were nasopharynx (0.75), oropharynx (0.75), and true vocal cords (0.71), with strong percentage agreement of 89%, 91%, and 87%, respectively. Both users of the adaptor found the recording setup to run smoothly.

Conclusion: A telemedicine device for NPL use demonstrates strong diagnostic accuracy across providers and good overall evaluation. It holds potential for use in remote settings.

Keywords
nasolaryngoscopy, telemedicine, diagnostic accuracy

Introduction

Telemedicine continues to evolve across health care with new technology and devices, allowing expanded patient access to medical providers. Particular to otolaryngology, previous studies have shown improvements in patient care through a variety of telemedicine uses, including telehealth consultations,1,2 overnight flap checks,3 tele-emergency room triage for nasal trauma,4 and by utilizing devices such as otoscopes,5 otorhinoendoscopes,6 and laryngoscopes7,8 to provide information to experts at remote destinations. However, certain components of an in-person, live examination are hard to replicate in remote telemedicine, such as optimal visualization, the ability to ask questions, or the gestalt of the full clinical picture. Thus, it is essential that these novel...
telemedicine instruments be validated to ensure reliability and diagnostic accuracy.

A telemedicine-based nasopharyngolaryngoscopy (tele-NPL) system has proven its utility in a variety of situations, such as for consultations in family medicine and other primary care clinics, as well as for resident teaching and educational purposes during inpatient consultations. To date, there have been no published studies validating the diagnostic accuracy of flexible fiber-optic NPL recordings. In this study, we examined the inter-rater reliability of the diagnostic accuracy for remote examination of NPLs.

**Methods**

**Patient Demographics**

Patients at the Children’s Hospital of Philadelphia whose outpatient visit warranted an NPL examination were considered for participation. There were no exclusion criteria, as long as the presenting symptom related to an anatomic site evaluated in the study. If consent was obtained, the otolaryngologist responsible for the patient’s care at the clinic would perform the NPL with a coupling device (MobileOptx, Philadelphia, Pennsylvania) attached to an iPhone (Apple, Cupertino, California; Figure 1). The attachment directly connects to the eyepiece of the scope on one side and redirects images to the iPhone camera on the other, allowing for direct visualization and recording. Following the examination, the video was securely transmitted electronically to the remote otolaryngologist, who was blinded to the patient’s history, audio, and findings of the live examination. The remote examiner was provided basic information about the patient including age, sex, and chief complaint. Blinding was performed to not guide or encourage the second reader of the examination to comment solely on specific anatomic parts or agree with the live reviewer’s clinical gestalt (often voiced audibly) on the video recording. The otolaryngologists involved in the study alternated between performing the live examinations and reviewing video recordings sent remotely. The remote examinations were not performed immediately after the live examinations and as such, there was no change in management if there were discrepancies between the exam evaluations. Videos did not contain identifiable health information and recordings were stored on a password-protected file and computer.

**Survey**

A diagnostic survey was created for this study that both the live and remote otolaryngologists completed following their respective examinations. The survey designed incorporated feedback from multiple otolaryngologists regarding question format, meaningful answer choices, intuitive design, and general understanding. The survey broadly covered the following sections: final NPL diagnosis, patency of airway, nasopharynx, oropharynx, base of tongue, epiglottis, arytenoids, aryepiglottic (AE) folds, false vocal cords (FVC), and true vocal cords (TVC). The specific subsections included on the survey were nasopharynx—normal or adenoid hypertrophy (with a degree of adenoid obstruction as secondary analysis); oropharynx—normal or tonsillar hypertrophy (with a degree of tonsil obstruction as secondary analysis); base of tongue—normal, large lingual tonsil, or other; epiglottis—normal, edematous, tubular, lesion, obstructing, or other; arytenoids—normal, edematous, redundant, red, or obstructing; AE folds—normal, tight, redundant, or red; FVCs—normal, edematous, mass, lesion, malacia (in the setting of laryngomalacia); TVC—normal, partially obstructed, not visible, edematous, mass, nodules, paresis, or paralysis (Figure 2). The examiners could check more than one box in the multiple-choice options in the larynx section given the possibility of presentations. The diagnostic category was in a free-response format. The recordings included audio but were muted upon remote review.

**Statistical Analysis**

Cohen \( \kappa \) was used to measure the amount inter-rater agreement in NPL survey results between live and remote otolaryngologists, for each outcome. \( \kappa \) correlation coefficients were calculated along with 95% confidence intervals. The \( \kappa \) rating scale of \(<0.2\), poor; \(0.2\) to \(0.4\), fair; \(0.4\) to \(0.6\), moderate; \(0.6\) to \(0.8\), good; and \(0.8\) to \(1.0\), strong was utilized. To provide a sense of observed agreement in addition to the proportion of agreement
beyond that expected by chance, which is gained from Cohen’s kappa, an additional measure for percentage of agreement was also calculated as the number of agreement scores divided by the total number of scores. Statistical analyses were performed using SAS 9.4 (SAS Institute, Cary, NC, USA). This study was approved by the institutional review board at the Children’s Hospital of Philadelphia under protocol number 16-013553.

Results

Demographics

A total of 45 patients underwent an NPL with remote evaluation and were included in the study. The average age was 4.9 years and the range was 4 weeks to 17 years. Twenty-eight (62%) patients were male. The indications for the NPL included noisy breathing (16), hoarseness (13), chronic cough (8), recurrent croup (2), dysphagia (2), congestion (1), chronic throat clearing (1), snoring (1), and difficulty breathing (1). Thirty-four of the 45 (76%) NPLs were performed by 1 otolaryngologist with the remaining 11 (24%) of the 45 performed by the other (Table 1).

Survey Results

The strongest inter-rater agreements were in the groups of diagnosis (0.74), nasopharynx (0.75), and oropharynx (0.75). The most agreed upon diagnoses were laryngomalacia (n = 14), normal examination (n = 9), vocal cord nodules (n = 4), and GERD/reflux changes (n = 3). The inter-rater agreement for the remaining subsites were base of tongue: 0.48; epiglottis: 0.43; arytenoids: 0.34; AE folds: 0.20; FVC: 0.34; TVC: 0.71; and patency of airway: 0.38. There was also a strong percentage agreement across most of the subsites, with 7 of the 10 sites analyzed having over an 80% agreement. The

Table 1. Demographics of Sample Who Underwent Tele-Nasopharyngolaryngoscopy.

| Demographic                        | Number             |
|------------------------------------|--------------------|
| Average age                        | 4.9 years (4 weeks to 17 years) |
| Male (%)                           | 28 (62%)           |
| Live examination recorded by author L.M.E. | 34 (76%)           |
| Live examination recorded by author A.B. | 11 (24%)           |
| Chief complaint                    | n (%)              |
| Noisy breathing                    | 16 (36)            |
| Hoarseness                         | 13 (29)            |
| Chronic cough                      | 8 (18)             |
| Recurrent croup                    | 2 (4)              |
| Dysphagia                          | 2 (4)              |
| Congestion                         | 1 (2)              |
| Chronic throat clearing            | 1 (2)              |
| Snoring                            | 1 (2)              |
| Difficulty breathing               | 1 (2)              |

*n = 45.
percentages were diagnosis: 80%; nasopharynx: 89%; oropharynx: 91%; base of tongue: 96%; epiglottis: 73%; arytenoids: 53%; AE folds: 64%; FVC: 89%; TVC: 87%; and patency of airway: 93% (Table 2).

Several of the subsites had further categorization. For example, in the nasopharynx section, there were 14 examinations in agreement for enlarged adenoids. Of these 14, 5 (36%) cases matched in a degree of adenoid obstruction, while the remaining 9 (74%) cases had differing degrees of obstruction (eg, <30% vs 30% to 60% obstruction). For the oropharynx, 8 examinations were in agreement of the examination demonstrating enlarged tonsils. All these 8 evaluations were concordant in describing the enlarged tonsils in either the 1 to 2+ range or 3 to 4+ range for the same examinations. Eight evaluations for the epiglottis subsite were in concordance with being abnormal. Of these, 5 were in agreement that the examination demonstrated a tubular epiglottis. For arytenoids, 18 agreed in the abnormal examination with 4 agreeing on the abnormality; the others held a combination of edematous, redundant, and obstructing descriptions. For AE folds, all 4 were abnormal and were all in agreement on the abnormality (tight). For FVC, 2 were agreed to be abnormal. For TVCs, 11 were abnormal and 8 were agreed upon abnormalities. For the diagnosis subcategory, 32 had abnormal examinations and 27 of these examinations were in agreement. Those examinations that were not in agreement mostly reflected differences in interpretations regarding the diagnosis of gastroesophageal reflux disease (GERD) versus laryngomalacia (Table 3). Both otolaryngologists were satisfied with the logistics of attaching the adaptor to an iPhone in clinic for video recording and transmission.

### Discussion

This study describes the utility of a tele-NPL system and its diagnostic accuracy between live and remote evaluations of the recordings. Overall, there was a good inter-rater agreement and high percentage agreement in the examinations for most of the subsites. In particular, there appear to be several areas where this telemedicine service is particularly helpful, such as for issues arising from the nasopharynx and oropharynx as well as patency of the airway. These results hold promise for remote diagnostic utility in clinical practice.

Our findings are similar to another study by Lozada et al that analyzed agreement between the resident and attending in using mobile recordings performed by residents, which demonstrated an inter-rater agreement of 0.75 and overall percent agreement of 89%. While the main purpose of this study was for application in teaching utility versus remote telemedicine and monitoring, the overall results are similar and encouraging with regard to generalizability to the adult population as well as feasibility in both remote and academic practices.

Several of our subsites had an overall low agreement, most notably the arytenoids, AE folds, and TVCs, which all had inter-rater agreement values of 0.35 or less, although these subsites maintained modest percentage agreements in our data (53%, 64%, and 89%, respectively). One reason for low inter-rater agreement but high percent agreement could be attributed to lower prevalence within diagnostic groups for each measure. Additionally, these lower values may be attributed to subjective survey wording. For example, in the arytenoid category, only 4 of the 18 that were categorized as abnormal agreed in cause of abnormality, while the remaining mismatched across edematous, redundant, and obstructing descriptions. It is possible that these words were overly subjective, and perhaps additional refinement of word choices may prove beneficial for optimizing agreement and diagnostic accuracy in these subsites. Finally, these subsite sections of the survey had more (at least 4) answer choices, which spread out over the total data of the study and did not provide a large number of data points for each of these answer choices. These subsites thus rely more on percentage agreement to assess reliability. Despite the subsites with a low inter-rater agreement, 94% of diagnoses that were stated as abnormal were either in complete agreement or were stated to be laryngomalacia versus GERD, 2 diagnoses that often occur coincidently in the pediatric patient population. Because of this, the variability in disagreement across the subsites does not appear to hold substantial clinical significance across our data; however, a more robust sample size would prove helpful to fully understand this finding.

This study has limitations. The live NPL examinations and recordings were performed by 2 academic pediatric otolaryngologists at 1 medical center. While we switched the role of the otolaryngologist to remove some bias, more recordings from a greater number of otolaryngologists would enhance the robustness of this study. Further, with similar backgrounds and likely similar evaluations of NPLs, the generalizability is lower. Given that this study only evaluated pediatric patients, it cannot be generalizable to the adult population; however, Brant et al found strong inter-rater agreement across similar subsites in the adult population. Some of the diagnoses, such as GERD, can be difficult to diagnose objectively and may be disputed across clinicians. While our study only had n = 3 of such diagnosis, it does hold as a limitation or possible confounder in our data. Finally, the survey we designed and utilized in this study is not

| Table 2. Overall Inter-Rater Agreement by κ Statistic and Percentage Agreement for Patients Who Underwent Tele-Nasopharyngolaryngoscopy. |
|---------------------------------------------------------------------------------------------------------------|
|                                                                                                               |
| **Statistic** | **95% Confidence Interval** | **Percentage Agreement** |
| Diagnosis     | 0.74                       | 0.59-0.89                     | 80%                  |
| Nasopharynx   | 0.75                       | 0.56-0.96                     | 89%                  |
| Oropharynx    | 0.75                       | 0.51-0.97                     | 91%                  |
| Base of tongue| 0.48                       | 0.0-1.0                      | 96%                  |
| Larynx—epiglottis | 0.43         | 0.16-0.69                     | 73%                  |
| Larynx—arytenoids | 0.34         | 0.16-0.51                     | 53%                  |
| Larynx—aryepiglottis folds | 0.20    | 0.0-0.42                      | 64%                  |
| Larynx—false vocal cord | 0.34    | 0.0-0.70                     | 89%                  |
| Larynx—true vocal cord | 0.71    | 0.52-0.91                     | 87%                  |
| Patency of airway | 0.38    | 0.0-0.91                     | 93%                  |
validated. There is no validated survey available for evaluation of NPL findings to date, and as such, this was the best option available at the time of the study.

This telemedicine model has several obvious applications. Utilizing remote video analysis within otolaryngology residency programs for trainees to perform and send educators their findings for verification reduces morbidity, reduces patient cost, and more efficiently utilizes expensive resources. Further, this telemedicine model allows for multiple remote clinicians to view a video, allowing for more intellectual discussion and consults for difficult cases which can help to improve outcomes. Finally, primary care physicians who do not have local access to specialists could be trained to perform an NPL and use this tool.

Although NPL has been proven to be an important diagnostic tool, it is most often used in the clinical setting to confirm and stage airway and throat diagnoses at the time of the examination. The primary goal of the paper was to determine the usefulness of the video recordings and to validate the interpretation of findings by 2 independent experts based on the video findings alone. In doing so, we found that the reliability of diagnosis was good (independent of audio and clinical history), but we found discrepancies at subsites that could improve with adding clinical information (such as symptoms or history of GERD) or by adding audio information (whereby the examiner could not only visualize findings consistent with edema but appreciate hoarseness or airway noises). Overall, our study confirms that videos of findings from NPL are useful as tools to relay information from one clinician to another but do not supplant the addition of audio/clinical information. More practical uses are to video NPL may help clinicians in triaging referrals to tertiary centers, reduce cost, as well as the need for multiple scopes by different experts to aid in clinical diagnosis. In addition, given that video NPL recordings are routinely uploaded into the electronic medical charts, it is important for clinicians to understand that the overall diagnostic reliability is good, but limitations in reliability of more subtle subgroup findings exist which could impact patient management strategies. Future studies would be important to validate a tool to rate/evaluate the findings based on anatomical findings.

### Conclusion

A telemedicine model for NPL demonstrates strong diagnostic accuracy across providers and good overall evaluation. It holds potential for use in remote settings to help clinicians evaluate patients with a complaint that warrants an NPL but also to help determine who should be referred to a tertiary center.

### 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.

### ORCID iD

Lauren E. Miller https://orcid.org/0000-0002-8693-8424

### References

1. Yulzari R, Bretler S, Avraham Y, Sharabi-Nov A, Even-Tov E, Gilbey P. Mobile technology-based real-time teleotolaryngology care facilitated by a nonotolaryngologist physician in an adult population. *Ann Otol Rhinol Laryngol*. 2018;127(1):46-50.

2. Kohlert S, Murphy P, Tse D, Liddy C, Afkham A, Keely E. Improving access to otolaryngology-head and neck surgery expert advice through eConsultations. *Laryngoscope*. 2018;128:350-355.

3. Alemi AS, Seth R, Heaton C, Wang SJ, Knott PD. Comparison of video and in-person free flap assessment following head and neck free tissue transfer. *Otolaryngol Head Neck Surg*. 2017;156(3):1035-1040.

4. Barghouthi T, Glynn F, Speaker RB, Walsh M. The use of a camera-enabled mobile phone to triage patients with nasal bone injuries. *Telemed J E Health*. 2012;18(3):150-152.

5. Moberly AC, Zhang M, Yu L, et al. Digital otoscopy versus microscopy: How correct and confident are ear experts in their diagnoses? *J Telemed Telecare*. 2018;24(7):453-459. 1357633X17708531.
6. Wu CJ, Wu SY, Chen PC, Lin YS. An innovative smartphone-based otorhinoendoscope and its application in mobile health and teleotolaryngology. *J Med Internet Res*. 2014;16(2):e71.
7. Lozada KN, Morton K, Stepan K, Capo J, Chai RL. The clinical impact of bedside fiberoptic laryngoscopic recording on a tertiary consult service. *Laryngoscope*. 2018;12(8):818-822.
8. Liu YF, Kim CH, Bailey TW, et al. A prospective assessment of nasopharyngolaryngoscope recording adaptor use in residency training. *Otolaryngol Head Neck Surg*. 2016;155(4):710-713.
9. Wilkins T, Gillies RA, Getz A, Zimmerman D, Kang L. Nasolaryngoscopy in a family medicine clinic: indications, findings, and economics. *J Am Board Fam Med*. 2010;23(2):591-597.
10. Zoumalan R, Maddalozzo J, Holinger LD. Etiology of stridor in infants. *Ann Otol Rhinol Laryngol*. 2007;116(1):329-334.
11. Brant JA, Leahy K, Mirza N. Diagnostic utility of flexible fiberoptic nasopharyngolaryngoscopy recorded onto a smartphone: a pilot study. *World J Otorhinolaryngol*. 2018;4:135-139.
12. Lechien JR, Akst LM, Hamdan AL, et al. Evaluation and management of laryngopharyngeal reflux disease: state of the art review. *Otolaryngol—Head Neck Surg (United States)*. 2019;160(5):762-782. doi:10.1177/0194599819827488.