Laryngoscopy Accuracy: Live Physician Recall vs. Post-Exam Video Review

Avery E. Schnell BS[1], Mark A. Vecchiotti MD [2], Alexander P. Marston MD [1,2], Andrew R. Scott MD[1,2]

[1]Tufts University School of Medicine, Boston, MA, 02116

[2]Dept. Of Otolaryngology, Tufts Medical Center, Tufts University School of Medicine, Boston, MA 02116

INTRODUCTION

Children are sent to the Otolaryngology clinic for laryngoscopy, which is done with limited cooperation and may miss subtle findings. This study aims to assess the difference in accuracy and diagnostic yield between real-time physician recall and post-exam video review to determine if recording laryngoscopies should be standardized, especially in young children.

METHODS

Recruitment was in the pediatric Otolaryngology clinic at Tufts between 8/2021-5/2022. Eligibility included age 0-18 and a scheduled laryngoscopy. A physician performed laryngoscopy and filled a questionnaire live to identify rhinitis, sinusitis, turbinate hypertrophy, adenoiditis, adenoid hypertrophy, velopharyngeal insufficiency, lingual tonsil hypertrophy, glossoptosis, laryngomalacia, vocal fold nodules, laryngeal edema, and vocal fold immobility. The questionnaire was repeated during video review. Discrepancies between the live and video review were recorded to assess intrarater reliability.

Characteristic	Total
Sex Assigned at Birth	
Female, No. (%)	27 (34.6)
Male, No. (%)	51 (65.4)
Race	
Asian/Pacific, No. (%)	4 (5.1)
Black/African, No. (%)	5 (6.4)
Caucasian, No. (%)	57 (73.1)
Unknown/Not Reported, No. (%)	11 (14.1)
Other, No. (%)	1 (1.3)
Ethnicity	
Hispanic or Latino, No. (%)	11 (14.1)
Not Hispanic or Latino, No. (%)	65 (83.3)
Unknown/Not Reported, No. (%)	2 (2.6)

Figure 1. Demographics for Patients Enrolled

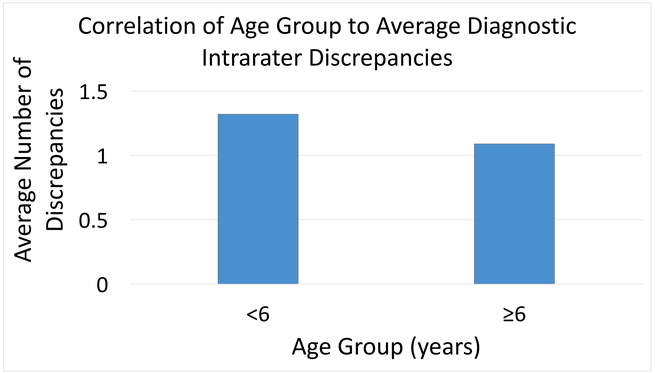

Figure 2. Correlation of Age Group to Diagnostic Discrepancy

Intrarater Changes in Pathology	Age Group <6	Age Group ≥6
Chronic rhinitis, No. (%)	6 (10.3)	3 (8.1)
Sinusitis, No. (%)	0 (0)	2 (5.4)
Turbinate hypertrophy, No. (%)	7 (12.1)	2 (5.4)
Chronic adenoiditis, No. (%)	0 (0)	4 (10.8)
Adenoid hypertrophy, No. (%)	2 (3.4)	0 (0)
Adenoid hypertrophy grade, No. (%)	4 (6.9)	3 (8.1)
Velopharyngeal insufficiency, No. (%)	1 (1.7)	2 (5.4)
Lingual tonsil hypertrophy, No. (%)	2 (3.4)	3 (8.1)
Glossoptosis, No. (%)	1 (1.7)	3 (8.1)
Laryngomalacia, No. (%)	5 (8.6)	1 (2.7)
Vocal fold nodules, No. (%)	7 (12.1)	3 (8.1)
Laryngeal edema, No. (%)	10 (17.2)	5 (13.5)
Vocal Fold Immobility, No. (%)	5 (8.6)	2 (5.4)
Additional diagnoses, No. (%)	8 (13.8)	4 (10.8)
Total, No. (%)	58 (100)	37 (100)

Figure 3. Intrarater Changes in Pathology by Age Group

RESULTS

77 eligible patients enrolled from 8/2021-5/2022. 57.1% were age <6; 42.9% were age ≥6. Using an independent two-tailed t test, the average change in intrarater reliability in age <6 (1.32+/-1.459) was not significantly different than age ≥6 (1.09+/-1.234). Post-hoc power analysis was .114 (d=.166, a=.05, N=77). Comparing the physicians, there was a statistically significant difference in intrarater reliability using one-way ANOVA (p=.021). Using Tukey post hoc test, Dr. A (.65+/-.875) had significantly lower intrarater reliability than Dr. B (1.69+/-1.635, p=.019). There were no significant differences between Dr. A and Dr. C (1.08+/-1.115, p=.521) or Dr. B and Dr. C (p=.199). Post hoc power analysis was .725 (d=.099, a=.05, N=20, 32, and 25 for Dr. A, B, and C).

CONCLUSIONS AND DISCUSSION

Archiving video laryngoscopy footage has the potential to benefit both patients and providers in various ways. Archived footage can be used to educate and counsel patients on their pathology. For providers, footage can be referred to when a report or note is written in a delayed fashion or peer input is desired. If needed, it allows for potential intraoperative reference later, and it can be used for comparison when serial laryngoscopies are performed over time. The purpose of this study was to determine if, on top of these benefits, recorded laryngoscopy also improved diagnostic accuracy during exams that must be performed quickly and are limited by patient cooperation. The results show there was no significant difference in intrarater reliability between age <6 and ≥6. Due to low power, a larger sample size and more research can better stratify differences in intrarater reliability by age. There were significant differences in intrarater reliability amongst physicians, which may be due to certain physicians focusing on performing the exam and relying on their ability to analyze the footage afterwards rather than diagnosing in real-time. This difference could inform further research into physician diagnostic accuracy and related factors. Overall, data is scarce on the accuracy of laryngoscopy review done live versus by video, and more data is needed to standardize video review.