Published in final edited form as:

Laryngoscope. 2015 October ; 125(10): 2259–2265. doi:10.1002/lary.25240.

Impact of educational intervention on the inter-rater agreement of nasal endoscopy interpretation

Patrick Colley, MD¹, Jess C. Mace, MPH², Madeleine R. Schaberg, MD, MPH¹, Timothy L. Smith, MD, MPH², and Abtin Tabaee, MD¹

¹Department of Otolaryngology – Head and Neck Surgery, Mount Sinai Health Systems. New York, New York USA

²Department of Otolaryngology – Head and Neck Surgery, Oregon Health and Science University. Portland, Oregon USA

Abstract

OBJECTIVE—Nasal endoscopy is integral to the evaluation of sinonasal disorders. However, prior studies have shown significant variability in the inter-rater agreement of nasal endoscopy interpretation amongst practicing rhinologists. The objective of the current study is to evaluate the inter-rater agreement of nasal endoscopy amongst otolaryngology residents from a single training program at baseline and following an educational intervention.

METHODS—11 otolaryngology residents completed nasal endoscopy grading forms for 8 digitally recorded nasal endoscopic examinations. An instructional lecture reviewing nasal endoscopy interpretation was subsequently provided. The residents then completed grading forms for 8 different nasal endoscopic examinations. Inter-rater agreement amongst residents for the pre- and post-lecture videos was calculated using the unweighted Fleiss' kappa statistic (Kf) and intra-class correlation agreement (ICC).

RESULTS—Inter-rater agreement improved from a baseline level of fair (Kf range 0.268–0.383) to a post-educational level of moderate (Kf range 0.401–0.547) for nasal endoscopy findings of middle meatus mucosa, middle turbinate mucosa, middle meatus discharge, sphenoethmoid recess mucosa, sphenoethmoid recess discharge and atypical lesions (ICC, $p < 0.001$). The baseline level of agreement for evaluation of nasal septum deviation was poor/fair and did not improve following educational intervention.

CONCLUSIONS—This study demonstrates a limited baseline level of inter-rater agreement of nasal endoscopy interpretation amongst otolaryngology residents. The inter-rater agreement for the majority of the characteristics that were evaluated improved after educational intervention. Further study is needed to improve nasal endoscopy interpretation.

Corresponding author. Abtin Tabaee, MD, Beth Israel Medical Center, Department of Otolaryngology, 10 Union Square East, Suite 4J, New York, NY 10003, T: 212-844-8458 F: 212-844-6672, atabaee1@gmail.com.

Financial disclosures – The other authors (PC, MS, AT) have no other sources of funding, financial disclosures associated with this manuscript.

Conflict of interest - none

Presentation – Poster at the American Rhinologic Society – Combined Otolaryngologic Society Meeting. Las Vegas, NV. May 16 and 17 2014.

Keywords

Nasal endoscopy; chronic sinusitis; otolaryngology; residency; inter-rater agreement

INTRODUCTION

Nasal endoscopy is a primary diagnostic tool in the evaluation of patients with sinonasal disorders. When used alone, or coupled with computed tomography, nasal endoscopy provides objective evaluation of disease related findings, including various forms of sinusitis, rhinitis, and neoplastic lesions. Positive characteristics include clinical utility, repeatability, patient tolerance, favorable safety profile, and ability to digitally record the examination. Despite its fundamental role in rhinology, a standardized method of interpreting the procedure findings is lacking. Previous authors have developed systems for scoring and describing the endoscopic examination^{1,2}. However, many findings on nasal endoscopy are not absolute and rely on the judgment of the endoscopist including an understanding of normal anatomy, and an ability to identify deviations from that standard. The potential for subjectivity and differences of opinion amongst clinicians in interpreting nasal endoscopy negatively impacts its precision^{3,4}.

Additionally, the methodology of teaching nasal endoscopy at the resident level is generally neither uniform nor systematic. Trainees develop competency during residency through instruction by more experienced specialists including faculty and senior residents. Repetition of the procedure and pattern recognition in the clinical setting solidify the educational process. To date, no standardized method of teaching nasal endoscopy has been developed, potentially resulting in variability in the educational process.

Theoretical concerns about the precision of nasal endoscopy interpretation have been borne out by prior studies demonstrating disparate inter-rater agreement amongst rhinologists^{3, 4}. Standardization of both endoscopy interpretation and the associated educational process may improve the precision and utility of the procedure. The goals of the current study are to evaluate inter-rater agreement of nasal endoscopy at the resident level at baseline and following an educational intervention.

MATERIALS AND METHODS

Institutional Review Board approval was obtained prior to the study. Diagnostic nasal endoscopy was performed in a consecutive cohort of patients undergoing sinonasal surgery by the senior author (AT). Inclusion criteria included age greater than 18 years and no prior history of sinonasal surgery. All primary sinonasal diagnosis and surgery types were included. The nasal endoscopy procedure was performed following induction of anesthesia and preparation of the nasal cavity with topical decongestion with oxymetazoline soaked pledgets. Nasal endoscopy video recordings were performed using a 0-degree rigid endoscope, 3 chip charge-couple device (CCD) high definition camera and Aida recording system (Karl Storz, Tuttlingen, Germany). Recordings were edited using commercially available media software (iMovie 2014, Apple Inc, Cupertino, CA). A total of 35 videos were collected for this study. Each video contained examinations of both nasal cavities for a

total of 70 sides. From these, videos representing 16 unique patient sides were included based on adequate visualization of the anatomic sub-sites of interest. A video was excluded if it failed to adequately demonstrate any of the anatomic sub-sites of interest. Editing of the included videos was performed to show only a unilateral nasal endoscopy examination with all of the anatomic sub-sites seen in a single pass.

This study assessed the inter-rater agreement of nasal endoscopy interpretation pre- and post-educational intervention amongst a group of 11 otolaryngology residents from a single training program (New York Eye and Ear Infirmary). A grading instrument consisting of 7 anatomic and inflammatory markers was developed based on prior studies^{3, 4}(Figure 1). At the onset of the study, the residents were given the grading forms and instructed to complete them after viewing a series of videos. No additional study specific instructions or education was given. Eight nasal endoscopy videos were successively shown. The residents were oriented to the side of the nose that was being shown prior to each video. Each video lasted between 15 to 40 seconds and was shown twice without pause. Following the initial videos, a 20 minute instructional lecture with slides was given by the senior author (AT) reviewing nasal endoscopy interpretation including still endoscopy photos of the different findings relevant for the grading instrument: normal bilateral middle meati, normal spheno-ethmoid recesses, inflammatory edema, polypoid changes, clear secretions, muco-purulent secretions, obstructive nasal septum deviation. For the purposes of this lecture, nasal septum deviation was defined as obstructive when it blocked the nasal airway and/or the middle meatus. Eight additional videos were subsequently viewed and graded following the instructional lecture.

Residents from all years of otolaryngology training (post-graduate year [PGY] 2 through 5) participated in the study in September 2013, 2.5 months after the onset of the academic year. The majority of the experience performing nasal endoscopy for residents is obtained through flexible and rigid nasal endoscopy performed in a busy resident clinic. Residents rotate through this clinic for 9 months during their PGY 2 year, 3 months during their PGY 3 year, 3 months during their PGY 4 year and 9 months during their PGY 5 year. In addition to performing nasal endoscopy in clinic, residents gain experience with the sino-nasal examination during two 3-month rotations with an emphasis on operative rhinology during the PGY 3 and PGY 5 year. A variable amount of exposure to rhinology and nasal endoscopy occurs during other clinical rotations that are focused on other sub-specialties.

Statistical Analysis

Descriptive analytics were used to describe the prevalence of pre-intervention and post-intervention score responses for each endoscopic attribute of interest. Inter-rater agreement was assessed using unadjusted, fixed-marginal, Fleiss' kappa statistic (K_f) for both pre-intervention and post-intervention agreement for endoscopic attributes in both the total resident cohort ($n=11$), and separate junior level resident cohort (PGY 2 and 3; $n=5$) and senior level resident cohort (PGY 4 and 5; $n=6$). Fleiss's kappa is a summary statistic to determine the degree of overall agreement between all raters above the level expected by chance alone. Kappa values range from -1.000 to 1.000 where -1.000 is defined as "perfect disagreement", 0.000 is defined as the null hypothesis (agreement expected by chance), and 1.000 indicates "perfect agreement".⁵ P-values for each kappa statistic were calculated to

determine if inter-rater agreement was due to chance observations alone. P-values less than 0.0500 denote that the observed rater agreement was not due to chance. The prevalence of observed agreement (Po) and expected agreement (Pe) was also calculated for each endoscopic attribute to further supplement reported kappa values. Kappa calculations are based on the difference between observed agreement (Po: the degree of agreement actually found) compared to expected agreement (Pe: the degree of agreement expected by chance). The interpretation of K_f values requires the context of both Po and Pe values to determine the precision and accuracy of reported total Kappa estimations. For example, if K_f values are less than reported Pe values, the level of agreement should be considered low as it does not pass the level of agreement through random chance alone⁶. Conventional methods of interpreting kappa values were used with level of agreement categories of “poor/slight”(K_f 0.0–0.20), “fair”(K_f 0.21–0.40), “moderate”(K_f 0.41–0.60), “strong/substantial”(K_f 0.61–0.80), “almost perfect”(K_f 0.81–1.00). Asymptotic standard errors (ASE), two-sided 95% confidence intervals(CI), Z-test statistics, and conventional interpretations of kappa values were reported.^{7,8} All analyses were conducted using a commercially available statistical software(SPSS v.22.0; IBM Corp., Armonk, NY) and standard database spreadsheet software(Excel 2007; Microsoft Inc., Redmond, WA).

Agreement between discrete observer pairings for each endoscopy exam attribute was additionally assessed for both the entire cohort, as well as junior and senior level resident groups, using intra-class correlation(ICC) statistics for both pre-intervention and post-intervention agreement. Two-sided 95% CIs and corresponding p-values are reported for each attribute.

RESULTS

The study cohort consisted of 9 males(56%) and 7 females(44%). The mean age of the patient cohort at the time of nasal endoscopy was 41.1 years (standard deviation 16.3 years). The indication for surgery in the series included chronic rhinosinusitis with nasal polyposis(7 patients, 44%), deviated nasal septum(3 patients, 19%), chronic rhinosinusitis without nasal polyposis and deviated nasal septum(3 patients, 19%), recurrent acute rhinosinusitis with deviated nasal septum(1 patient, 6%), sinonasal inverted papilloma(ipsilateral to video side, 1 patient, 6%) and inferior turbinate hemangioma(contralateral to video side, 1 patient, 6%). The endoscopy was left sided in 10 patients(62%) and right sided in 8 patients(38%). The total number and percentage of pre-intervention and post-intervention responses for all attributes assessed is shown in Table 1. Out of a possible 616 responses, 601 (97.5%) pre-intervention and 603 (97.8%) post-intervention responses were collected. Inter-rater agreement amongst the total resident cohort improved from a baseline level of “fair”(K_f range 0.268–0.38) to “moderate”(K_f range 0.401–0.547) for nasal endoscopy findings of middle meatus mucosa, middle turbinate mucosa, middle meatus discharge, sphenoethmoid recess mucosa, sphenoethmoid recess discharge and atypical lesions following the educational lecture. The baseline level of agreement for evaluation of nasal septum deviation in the total resident cohort was “poor/fair”(K_f 0.084) and did not improve agreement levels following educational intervention(K_f 0.176) (Table 2). In the junior resident cohort, the pre-intervention level of agreement improved from “poor/slight” to “fair” for assessment of the nasal septum(pre K_f 0.105/post

Kf 0.234) and atypical mass lesion(pre Kf 0.086/post Kf 0.341), from “poor/slight” to “moderate” for assessment of the middle meatus mucosa(pre Kf 0.172/post Kf 0.431) and sphenoethmoid mucosa(pre Kf 0.157/post Kf 0.430), from “fair” to “moderate” for assessment of the middle turbinate mucosa(pre Kf 0.395/post Kf 0.503), and from “fair” to “strong/substantial” for assessment of middle meatus discharge(pre Kf 0.360/post Kf 0.646) (Table 3). In the senior resident cohort, the pre-intervention level of agreement improved from “poor/slight” to “strong/substantial” for assessment of the sphenoethmoid recess discharge(pre Kf 0.127/post Kf 0.603), and from “fair” to “moderate” for assessment of middle meatus mucosa(pre Kf 0.325/post Kf 0.537) and middle meatus discharge(pre Kf 0.236/post Kf 0.437). The remaining attributes assessed in the senior resident cohort did not change after the educational lecture including middle turbinate mucosa(pre Kf 0.295/post Kf 0.373), nasal septum(pre Kf 0.075/post Kf 0.077), sphenoethmoid recess mucosa(pre Kf 0.419/post Kf 0.409) and atypical mass lesion(pre Kf 0.480/post Kf 0.515) did not change after the educational lecture(Table 4). For all attributes assessed and for all cohorts the prevalence of observed agreement(Po) was higher than expected agreement(Pe) in both the pre-intervention and post-intervention assessment.

Intra-class correlation coefficients for nearly all attributes evaluated showed statistically significant agreement between discrete observer pairings in the total resident(16 of 16, all p 0.006), junior resident(14 of 16, all p 0.018) and senior resident cohorts(15 of 16, all p 0.012)(Table 5). Assessment of the nasal septum and atypical mass lesions were the only attributes for which the intra-class correlation was not consistently statistically significant. The intra-class correlation coefficient also increased between the pre-intervention and post-intervention assessment for all attributes amongst all cohorts except for the nasal septum in the total and senior resident cohort.

DISCUSSION

The pre-intervention inter-rater agreement among otolaryngology residents in the current study was “fair” or “poor/slight” for all categories evaluated, less than that found among the senior rhinologists in previous studies^{3,4}. However, the inter-rater agreement improved to “moderate” after educational intervention for all categories other than the nasal septum. This finding demonstrates that even with a limited amount of educational intervention, during which the terms used to describe nasal endoscopy findings are defined and standardized, the level of agreement for residents improves to a level comparable to that of more senior clinicians.

If the data is stratified by level of training, senior residents notably had a higher level of pre-intervention agreement than junior residents with 5 categories having a “fair” or “moderate” level of agreement for the senior residents compared to 3 for the junior residents. However, the junior resident cohort experienced greater improvement after the educational intervention and had comparable categories demonstrating “moderate” to “strong/substantial” agreement on the post-intervention testing when compared to the senior residents. The only area that showed no significant improvement after educational intervention was evaluation of the nasal septum. This area improved from “poor/slight” to “fair” for the junior resident cohort but remained unchanged in the total and senior resident

cohorts. This may be due to the loss of perspective associated with recorded unilateral endoscopy videos. The videos shown only included a single side of the endoscopy making it difficult to compare the shape and position of the septum to the other side. This may have also been compounded by the lack of a nasal speculum examination.

Improving the precision of nasal endoscopy interpretation and development of an educational model may be necessary to propagate a common language in nasal endoscopy interpretation. The implications of this process for management of sinonasal disease are profound and include improving the study precision and communication amongst clinicians. Though not the focus of this study, accuracy as it relates to nasal endoscopy has a complimentary importance and reflects the correlation of examination findings to actual disease. Several authors have taken the first step in standardizing nasal endoscopy by proposing grading scales. Lund and Kennedy developed a ten-point scale that allowed for the endoscopic evaluation of polyps, edema, discharge, scarring and crusting throughout different locations in the sinonasal cavity.¹ This system has not been validated but has been correlated with computed tomography (CT) findings in pre- and post-sinus surgery patients. Ryan et al⁹ noted a strong correlation coefficient(0.76) between post-operative Lund-Kennedy endoscopy scores and Lund-Mackay CT scores. Other authors found correlations between this scoring system and patient survey responses. Toros et al¹⁰ found a significant correlation between the Lund-Kennedy scale and total patient reported preoperative($r=0.48$, $p<0.001$) and 12-month postoperative($r=0.63$, $p<0.001$) symptom scores. The perioperative sinus endoscopy scoring system(POSE), proposed by Wright and Agrawal², is a modified version of the Lund-Kennedy scoring system designed specifically to evaluate sinus cavities before and after endoscopic sinus surgery. In their study, the two authors above found that the POSE system demonstrated a moderate correlation($r=0.50$) with postoperative patient survey answers during the 6-month postoperative period. Other authors have developed grading scales including the DIP score(discharge, inflammation, polyps/edema) as a measure of inflammation in chronic rhinosinusitis¹¹ and a staging system for allergic fungal sinusitis¹². Despite the existence of multiple scoring systems, none have been universally accepted by practicing clinicians. The current study utilizes a grading instrument that is based on two prior studies by the senior author investigating nasal endoscopy inter-rater agreement^{3,4}. The instrument attempts to identify and broadly categorize key rhinologic findings.

The inter-rater agreement of nasal endoscopy in previous studies has varied. Annamalai et al¹³ recorded nasal endoscopies from patients with a variety of nasal complaints. Two independent reviewers graded these videos using the Lund-Kennedy scoring system. Using Cohen's kappa, they found inter-rater agreements of "moderate" for edema, "good" for crusting and "very good" for polyps and discharge. Since a diagnosis of rhinosinusitis was not an inclusion criterion, many nasal endoscopies in this study had no abnormal findings. Raithatha et al⁴, in contrast, studied the inter-rater agreement amongst 5 rhinologists using a similar grading form as used in the current study to review nasal endoscopy videos of patients with a diagnosis of chronic rhinosinusitis. They used unweighted Fleiss kappa statistics to compare endoscopy interpretations. Levels of agreement were "almost perfect" for atypical lesions, "substantial" for nasal polyps, "moderate" for nasal discharge and mucosal inflammatory changes of the middle turbinate, and "fair" for edema of the middle

meatus, obstruction by nasal septal deviation and obstruction by the middle turbinate. McCoul et al³ also used 5 rhinologists to review nasal endoscopy videos and reported similar rates of inter-rater agreement when evaluating patients who had undergone endoscopic sinus surgery. Inter-rater agreement was “excellent” for polyps, “moderate” for middle turbinate integrity, middle turbinate position, maxillary sinus patency, and ethmoid sinus patency, “fair” for discharge, synechiae, and middle meatus patency, and “poor” for middle turbinate mucosal changes and uncinate process. While all of these studies varied in their patient selection criteria and outcomes measured, the inter-rater agreement tended to be higher when evaluating more objective findings such as the presence of lesions or discharge. Agreement decreased when evaluating more subjective findings such as mucosal quality or inflammatory edema. This was also found to be the case in our study as the only two categories that had substantial inter-rater agreement both involved the assessment of discharge. The previous studies all involved attending otolaryngologist reviewers, the majority of whom were fellowship-trained rhinologists. Despite their experience and training, there was still variability in many aspects of the nasal endoscopy interpretation. This disagreement amongst senior members of the field highlights the importance of standardizing nasal endoscopy findings.

Our study differs from previous studies in that it compares inter-rater agreement before and after educational intervention. Additionally, the subjects interpreting the endoscopies were residents, rather than practicing clinicians. It is notable that the educational intervention used in this study was only a brief lecture showing examples of specific findings. A programmatic approach to educating residents in nasal endoscopy interpretation could involve a formal curriculum that would include didactic discussion and supervision of multiple endoscopy procedures both in a laboratory and clinical setting. Digital recording technology greatly enhances the efficiency of this process as a library of endoscopic videos can be readily integrated into an educational paradigm.

There are a number of limitations associated with this study including, the sample size of patients, endoscopy videos, and residents. Despite this, significant p-values were obtained for the majority of kappa values which suggests that the sample size was likely adequate. In our study, all reviewers were residents, an intuitive choice for the study of an educational intervention. However, since multiple PGY years were included, a degree of heterogeneity exists given the variable experience with nasal endoscopy. The study did not assess the a priori level of knowledge of nasal endoscopy of the residents. A lack of fundamental knowledge of nasal endoscopy and rhinologic disorders in addition to disparate learning curves amongst residents would potentially affect the results. Furthermore, residency programs differ in their exposure to nasal endoscopy and results could differ based on the residency program. The reported nasal endoscopy categories used in this study have not been previously validated, although similar endoscopy scoring forms have been used in two prior studies^{3,4}. Although the current study demonstrated an increase in the level of agreement following an educational intervention, it did not define the “accuracy” of how the findings were interpreted.

CONCLUSION

Despite the universal usage of nasal endoscopy in clinical practice, controversy continues to exist regarding the interpretation of the examination findings. In the current study, otolaryngology residents demonstrated a limited baseline level of inter-rater agreement which improved after an educational intervention for all categories except nasal septal deviation. The baseline level of agreement was higher amongst senior residents. However, the post-intervention improvement in inter-rater agreement was larger in magnitude amongst junior residents compared to senior residents and resulted in a similar level of agreement in both groups. Further research and collaborative discussion is needed to improve nasal endoscopy interpretation and education.

Acknowledgments

This study was partially supported by a grant from the National Institute on Deafness and Other Communication Disorders (NIDCD), one of the National Institutes of Health, Bethesda, Maryland (R01 DC005805; PI/PD: TL Smith).

Funding/support – Timothy L. Smith, MD, MPH and Jess C. Mace, MPH, are supported by a grant from the NIDCD.

REFERENCES

1. Lund VJ, Kennedy DW. Quantification for staging sinusitis. The Staging and Therapy Group. *Ann Otol Rhinol Laryngol Suppl.* 1995; 167:17–21. [PubMed: 7574265]
2. Wright ED, Agrawal S. Impact of perioperative systemic steroids on surgical outcomes in patients with chronic rhinosinusitis with polyposis: evaluation with the novel Perioperative Sinus Endoscopy (POSE) scoring system. *Laryngoscope.* 2007; 117:1–28. [PubMed: 18075447]
3. 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. [PubMed: 22696506]
4. Raithatha R, Anand VK, Mace JC, et al. Interrater agreement of nasal endoscopy for chronic rhinosinusitis. *Int Forum Allergy Rhinol.* 2012; 2:144–150. [PubMed: 22253129]
5. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull.* 1971; 76:378–382.
6. Sim J, Wright CC. The Kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Phys Ther.* 2005; 85:257–268. [PubMed: 15733050]
7. Fleiss JL, Nee JC, Landis RJ. Large sample variance of kappa in the case of different sets of raters. *Psychol Bull.* 1979; 86:974–977.
8. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33:159–174. [PubMed: 843571]
9. Ryan WR, Ramachandra T, Hwang PH. Correlations between symptoms, nasal endoscopy, and in-office computed tomography in post-surgical chronic rhinosinusitis patients. *Laryngoscope.* 2011; 121:674–678. [PubMed: 21305550]
10. Toros SZ, Bolükbası S, Naıboglul B, et al. Comparative outcomes of endoscopic sinus surgery in patients with chronic sinusitis and nasal polyps. *Eur Arch Otorhinolaryngol.* 2007; 264:1003–1008. [PubMed: 17431658]
11. Durr ML, Pletcher SD, Goldberg AN, Murr AH. A novel sinonasal endoscopy scoring system: the discharge, inflammation, and polyps/edema (DIP) score. *Int Forum Allergy Rhinol.* 2013; 3:66–72. [PubMed: 22927217]
12. Philpott CM, Javer AR, Clark A. Allergic fungal rhinosinusitis - a new staging system. *Rhinology.* 2011; 49:318–323. [PubMed: 21858263]

13. Annamalai S, Davis J, Kubba H. How subjective is nasal endoscopy: a study of interrater agreement using the Lund and Mackay scoring system. *Am J Rhinol.* 2004; 18:301–303. [PubMed: 15586801]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

NASAL ENDOSCOPY GRADING FORM

EXAM NUMBER: _____ RESIDENT INITIALS: _____

SIDE: ☐ RIGHT ☐ LEFT

<p>Sinus- Nasal Polyps: _____</p> <p>0= Absence of sinus-nasal polyps</p> <p>1= Polyps in middle meatus</p> <p>2= Polyps extending beyond middle meatus</p> <p>Middle Meatus Mucosal Edema: _____</p> <p>0= Absent</p> <p>1= Mild edema</p> <p>2= Severe edema</p> <p>Middle Meatus Discharge: _____</p> <p>0= Absent</p> <p>1= Clear, thin</p> <p>2= Thick, purulent</p> <p>Nasal Septum: _____</p> <p>0= No deviation</p> <p>1= Non-obstructing deviation</p> <p>2= Obstructing deviation</p>	<p>Mucosal Changes of Middle Turbinate: _____</p> <p>0= Normal</p> <p>1= Edema</p> <p>2= Polypoid</p> <p>Sphenoethmoid Recess Mucosal Edema: _____</p> <p>0= Absent</p> <p>1= Mild edema</p> <p>2= Severe edema</p> <p>Sphenoethmoid Recess Discharge: _____</p> <p>0= Absent</p> <p>1= Clear, thin</p> <p>2= Thick, purulent</p> <p>Atypical Mass Lesion,</p> <p>Concern for Neoplasm: _____</p> <p>1= No</p> <p>2= Yes</p> <p>Unable to Assess (any portion of exam)</p> <p>If unable to grade a particular part of exam, please comment on which part and why:</p> <p>_____</p>
---	--

Figure 1.

The nasal endoscopy grading instrument used in this study.

Table 1

Prevalence of pre and post-intervention responses using the study grading instrument. Frequency of resident responses are given for each attribute of the endoscopic examination

Attributes:	Total Score Responses n(%)	
	Pre-intervention	Post-Intervention
Mucosal change of middle turbinate		
0=normal	22 (25.0)	25 (29.1)
1=Edema	34 (38.6)	26 (30.2)
2=Polypoid	32 (36.4)	35 (40.7)
Middle meatus mucosa		
0=Normal	35 (41.2)	27 (31.8)
1=Edema (non-polypoid)	22 (25.9)	27 (31.8)
2=Polypoid changes	28 (32.9)	31 (36.5)
Middle meatus discharge		
0=Absent	38 (43.2)	29 (33.3)
1=Clear, thin	35 (39.8)	23 (26.4)
2=Mucoid-purulent	15 (17.0)	35 (40.2)
Nasal septum		
0=No deviation	33 (37.5)	42 (47.7)
1=Non-obstructing deviation	36 (40.9)	32 (36.4)
2=Obstructing deviation	19 (21.6)	14 (15.9)
Sphenoethmoid recess mucosa		
0=Normal	44 (54.3)	41 (49.4)
1=Edema (non-polypoid)	16 (19.8)	15 (18.1)
2= Polypoid changes	21 (25.9)	27 (32.5)
Sphenoethmoid recess discharge		
0=Absent	55 (66.3)	38 (44.2)
1=Clear, thin	19 (22.9)	14 (16.3)
2=Thick, purulent	9 (10.8)	34 (39.5)
Atypical mass lesion, concern for neoplasm		
1=Yes	74 (84.1)	59 (67.0)
2=No	14 (15.9)	29 (33.0)

Table 2

Pre and post-intervention unadjusted Fleiss' Kappa (Kf-m) values and prevalence of observed (Po) and expected (Pe) agreement for all residents (n=11).

Attribute:	Kf-m (ASE)	95% CI (LL, UL)	Z-value (p-value)	Conventional level of agreement	Po/ Pe
Pre - Mucosal change of middle turbinate	0.383(0.038)	(0.308, 0.459)	9.960(<0.0001)	Fair	0.596/0.344
Post - Mucosal change of middle turbinate	0.425(0.037)	(0.353, 0.498)	11.449(<0.0001)	Moderate	0.621/0.340
Pre - Middle meatus mucosa	0.311(0.042)	(0.228, 0.394)	7.374(<0.0001)	Fair	0.552/0.340
Post - Middle meatus mucosa	0.450(0.034)	(0.383, 0.517)	13.132(<0.0001)	Moderate	0.634/0.335
Pre - Middle meatus discharge	0.314(0.048)	(0.221, 0.407)	6.599(<0.0001)	Fair	0.571/0.374
Post - Middle meatus discharge	0.547(0.038)	(0.472, 0.622)	14.302(<0.0001)	Moderate	0.702/0.343
Pre - Nasal septum	0.084(0.042)	(0.002, 0.167)	1.995(0.0460)	Poor/slight	0.409/0.355
Post - Nasal septum	0.176(0.053)	(0.072, 0.279)	3.339(0.0008)	Poor/slight	0.493/0.385
Pre - Sphenothmoid recess mucosa	0.347(0.076)	(0.198, 0.496)	4.567(<0.0001)	Fair	0.625/0.426
Post - Sphenothmoid recess mucosa	0.401(0.061)	(0.282, 0.521)	6.612(<0.0001)	Moderate	0.639/0.396
Pre - Sphenothmoid recess discharge	0.268(0.112)	(0.049, 0.487)	2.401(0.0163)	Fair	0.650/0.522
Post - Sphenothmoid recess discharge	0.526(0.050)	(0.428, 0.624)	10.522(<0.0001)	Moderate	0.707/0.381
Pre - Atypical mass lesion, concern for neoplasm	0.354(0.204)	(-0.046, 0.755)	1.734(0.0828)	Fair	0.827/0.732
Post - Atypical mass lesion, concern for neoplasm	0.424(0.091)	(0.246, 0.602)	4.668(<0.0001)	Moderate	0.746/0.558

Kf-m, Fleiss's fixed-marginal multi-rater kappa statistic. ASE, approximate standard error. CI, confidence interval. LL, lower limit. UL, upper limit. Po, Prevalence of observed agreement. Pe, prevalence of expected agreement.

Table 3

Pre and post-intervention unadjusted Fleiss' Kappa (Kf-m) values and prevalence of observed (Po) and expected (Pe) agreement for junior residents (n=5).

Attribute:	K _{f-m} (ASE)	95% CI (LL, UL)	Z-value (p-value)	Conventional level of agreement	Po/ Pe
Pre - Mucosal change of middle turbinate	0.395(0.082)	(0.235, 0.555)	4.831(<0.0001)	Fair	0.600/0.339
Post - Mucosal change of middle turbinate	0.503(0.086)	(0.334, 0.672)	5.830(<0.0001)	Moderate	0.675/0.346
Pre - Middle meatus mucosa	0.172(0.098)	(-0.020, 0.365)	1.755(0.0793)	Poor/slight	0.488/0.381
Post - Middle meatus mucosa	0.431(0.083)	(0.267, 0.594)	5.172(<0.0001)	Moderate	0.625/0.341
Pre - Middle meatus discharge	0.360(0.106)	(0.151, 0.568)	3.385(0.0007)	Fair	0.613/0.395
Post - Middle meatus discharge	0.646(0.094)	(0.462, 0.830)	6.876(<0.0001)	Strong/substantial	0.775/0.365
Pre - Nasal septum	0.105(0.098)	(-0.087, 0.297)	1.076(0.2821)	Poor/slight	0.438/0.371
Post - Nasal septum	0.234(0.111)	(0.016, 0.452)	2.107(0.0351)	Fair	0.538/0.396
Pre - Sphenothmoid recess mucosa	0.157(0.138)	(-0.113, 0.428)	1.139(0.2545)	Poor/slight	0.538/0.451
Post - Sphenothmoid recess mucosa	0.430(0.128)	(0.180, 0.680)	3.370(0.0007)	Moderate	0.700/0.474
Pre - Sphenothmoid recess discharge	0.361(0.175)	(0.017, 0.704)	2.058(0.0396)	Fair	0.688/0.511
Post - Sphenothmoid recess discharge	0.404(0.095)	(0.218, 0.589)	4.265(<0.0001)	Moderate	0.625/0.371
Pre - Atypical mass lesion, concern for neoplasm	0.086(0.376)	(-0.650, 0.822)	0.228(0.8195)	Poor/slight	0.800/0.781
Post - Atypical mass lesion, concern for neoplasm	0.341(0.150)	(0.047, 0.634)	2.277(0.0228)	Fair	0.700/0.545

Table 4

Pre and post-intervention unadjusted Fleiss' Kappa (Kf-m) values and prevalence of observed (Po) and expected (Pe) agreement for senior residents (n=6).

Attribute:	K _{f-m} (ASE)	95% CI (LL, UL)	Z-value (p-value)	Conventional level of agreement	Po/Pe
Pre - Mucosal change of middle turbinate	0.295(0.072)	(0.154, 0.436)	4.100(<0.0001)	Fair	0.542/0.350
Post - Mucosal change of middle turbinate	0.373(0.066)	(0.243, 0.502)	5.646(<0.0001)	Fair	0.583/0.336
Pre - Middle meatus mucosa	0.325(0.074)	(0.180, 0.470)	4.938(<0.0001)	Fair	0.554/0.339
Post - Middle meatus mucosa	0.537(0.065)	(0.409, 0.664)	8.259(<0.0001)	Moderate	0.692/0.334
Pre - Middle meatus discharge	0.236(0.079)	(0.081, 0.391)	2.998(0.0027)	Fair	0.517/0.367
Post - Middle meatus discharge	0.437(0.065)	(0.309, 0.564)	6.718(<0.0001)	Moderate	0.625/0.334
Pre - Nasal septum	0.075(0.073)	(-0.068, 0.218)	1.022(0.3060)	Poor/slight	0.400/0.352
Post - Nasal septum	0.077(0.084)	(-0.088, 0.241)	0.916(0.3598)	Poor/slight	0.433/0.386
Pre - Sphenothmoid recess mucosa	0.419(0.107)	(0.209, 0.628)	3.912(<0.0001)	Moderate	0.658/0.412
Post - Sphenothmoid recess mucosa	0.409(0.082)	(0.248, 0.569)	4.968(<0.0001)	Moderate	0.625/0.366
Pre - Sphenothmoid recess discharge	0.127(0.161)	(-0.188, 0.443)	0.789(0.4303)	Poor/slight	0.592/0.532
Post - Sphenothmoid recess discharge	0.603(0.087)	(0.432, 0.773)	6.903(<0.0001)	Strong/substantial	0.758/0.392
Pre - Atypical mass lesion, concern for neoplasm	0.480(0.248)	(-0.006, 0.966)	1.932(0.0532)	Moderate	0.842/0.695
Post - Atypical mass lesion, concern for neoplasm	0.515(0.148)	(0.225, 0.805)	3.476(0.0005)	Moderate	0.792/0.570

Table 5

Intra-class correlation coefficient (ICC) results for all residents (n=11), junior residents (n=5) and senior residents (n=6).

Attribute:	Pre-Intervention		Post-Intervention	
	ICC(p-value)	95% CI	ICC(p-value)	95% CI
Mucosal change of middle turbinate- all residents	0.524(<0.001)	(0.281, 0.832)	0.738(<0.001)	(0.504, 0.935)
Mucosal change of middle turbinate- junior residents	0.604(<0.001)	(0.296, 0.880)	0.669(<0.001)	(0.374, 0.905)
Mucosal change of middle turbinate- senior residents	0.400(0.001)	(0.126, 0.775)	0.596(<0.001)	(0.309, 0.874)
Middle meatus mucosa- all residents	0.578(<0.001)	(0.298, 0.899)	0.704(<0.001)	(0.439, 0.938)
Middle meatus mucosa- junior residents	0.414(0.006)	(0.070, 0.847)	0.594(<0.001)	(0.284, 0.876)
Middle meatus mucosa- senior residents	0.552(<0.001)	(0.243, 0.874)	0.657(<0.001)	(0.380, 0.899)
Middle meatus discharge- all residents	0.532(<0.001)	(0.288, 0.836)	0.835(<0.001)	(0.652, 0.962)
Middle meatus discharge- junior residents	0.549(<0.001)	(0.236, 0.857)	0.870(<0.001)	(0.695, 0.968)
Middle meatus discharge- senior residents	0.559(<0.001)	(0.269, 0.858)	0.751(<0.001)	(0.506, 0.932)
Nasal septum- all residents	0.242(<0.001)	(0.069, 0.622)	0.162(0.006)	(0.023, 0.527)
Nasal septum- junior residents	0.188(0.070)	(-0.047, 0.628)	0.373(0.004)	(0.079, 0.767)
Nasal septum- senior residents	0.257(0.012)	(0.024, 0.673)	-0.061(0.706)	(-0.144, 0.235)
Sphenoethmoid recess mucosa- all residents	0.612(<0.001)	(0.221, 0.985)	0.716(<0.001)	(0.427, 0.956)
Sphenoethmoid recess mucosa- junior residents	0.286(0.018)	(0.016, 0.708)	0.637(<0.001)	(0.335, 0.893)
Sphenoethmoid recess mucosa- senior residents	0.603(<0.001)	(0.316, 0.877)	0.693(<0.001)	(0.425, 0.912)
Sphenoethmoid recess discharge- all residents	0.554(<0.001)	(0.219, 0.949)	0.899(<0.001)	(0.769, 0.978)
Sphenoethmoid recess discharge- junior residents	0.673(<0.001)	(0.379, 0.906)	0.689(<0.001)	(0.400, 0.912)
Sphenoethmoid recess discharge- senior residents	0.305(0.005)	(0.056, 0.711)	0.863(<0.001)	(0.692, 0.965)
Atypical mass lesion, concern for neoplasm- all residents	0.398(<0.001)	(0.175, 0.756)	0.478(<0.001)	(0.240, 0.807)
Atypical mass lesion, concern for neoplasm- junior residents	0.106(0.179)	(-0.093, 0.544)	0.422(0.001)	(0.118, 0.795)
Atypical mass lesion, concern for neoplasm- senior residents	0.551(<0.001)	(0.262, 0.854)	0.566(<0.001)	(0.277, 0.861)

ICC, single measure intra-class correlation using a two-way mixed model