Published in final edited form as:
Int Forum Allergy Rhinol. 2012 November ; 2(6): 453–459. doi:10.1002/alr.21058.

Inter-rater Agreement of Nasal Endoscopy in Patients with a Prior History of Endoscopic Sinus Surgery

Edward D. McCoul, MD, MPH[1], **Timothy L. Smith, MD, MPH**[2], **Jess C. Mace, MPH**[2], **Vijay K. Anand, MD**[1], **Brent A. Senior, MD**[3], **Peter H. Hwang, MD**[4], **James A. Stankiewicz, MD**[5], and **Abtin Tabaee, MD**[6]

[1]Department of Otolaryngology—Head and Neck Surgery, Weill Cornell Medical College, New York, NY

[2]Department of Otolaryngology—Head and Neck Surgery, Oregon Health and Science University, Portland, OR

[3]Department of Otolaryngology—Head and Neck Surgery, University of North Carolina, Chapel Hill, NC

[4]Department of Otolaryngology—Head and Neck Surgery, Stanford University, Palo Alto, CA

[5]Department of Otolaryngology—Head and Neck Surgery, Loyola University, Chicago, IL

[6]Department of Otolaryngology, Beth Israel Medical Center—Albert Einstein College of Medicine, New York, NY

Abstract

OBJECTIVE—Nasal endoscopy is an important part of the clinical evaluation of patients with chronic rhinosinusitis. However, its objectivity and inter-rater agreement have not been well studied, especially in patients who have previously had sinus surgery.

METHODS—Patients with a history of endoscopic sinus surgery for chronic rhinosinusitis were prospectively enrolled from a tertiary rhinology practice. Fourteen endoscopic nasal examinations were recorded using digital video capture software. Each patient also underwent computerized tomography (CT) and completed the Sinonasal Outcome Test (SNOT-22). Blinded review of inflammatory and anatomic findings for each video was independently performed by 5 academic rhinologists at separate institutions. Comparisons were performed using the unweighted Fleiss' kappa statistic (K_f) and the prevalence- and bias-adjusted kappa (PABAK).

RESULTS—There were no significant correlations between age, Lund-Mackay score or SNOT-22 score. Inter-rater agreement was variable across the characteristics studied. Mean PABAK was excellent for the assessment of polyps (K_f =0.886); moderate for the assessments of middle turbinate (MT) integrity (K_f =0.543), MT position (K_f =0.443), maxillary sinus patency (K_f =0.593) and ethmoid sinus patency (K_f =0.429); fair for discharge (K_f =0.314), synechiae (K_f =0.257) and middle meatus patency (K_f =0.229); and poor for MT mucosal changes (K_f =0.148) and uncinate process (K_f =0.126).

Corresponding Author: Abtin Tabaee, MD, Department of Otolaryngology, Beth Israel Medical Center, 10 Union Square East, Suite 4J, New York, NY 10003, T- 212-844-8494, F- 212-844-6672, ATabaee@chpnet.org.

Sources of funding, financial disclosures, conflicts of interest : Partially supported by grant funding from the National Institute on Deafness and Other Communication Disorders, one of the National Institutes of Health (R01DC005805; PI: Smith). Public clinical trial registration (http://www.clinicaltrials.gov) ID# NCT00799097.

Presentation: Accepted for oral presentation at American Rhinologic Society Spring Meeting, San Diego, CA. April 19—20, 2012.

CONCLUSIONS—The current study was notable for variability in the inter-rater agreement among the inflammatory and anatomic attributes that were examined. Further standardization of nasal endoscopy with regard to interpretation may improve the reliability of this procedure in clinical practice.

Keywords

endoscopy; rhinosinusitis; paranasal sinuses; surgery; inter-observer variability

INTRODUCTION

Nasal endoscopy is a common technique used by otolaryngologists for the clinical assessment of chronic rhinosinusitis. This procedure allows for close examination of critical anatomic regions including the ostiomeatal complex, sphenoethmoidal recess, and frontal recess. Moreover, the presence of inflammatory findings on nasal examination is an important criterion in the diagnosis of rhinosinusitis as defined by the American Academy of Otolaryngology.[1] Other favorable aspects of nasal endoscopy include patient tolerance, safety, repeatability and rapid skill acquisition. Despite the widespread use of nasal endoscopy, few grading schema have been proposed to summarize the endoscopic findings, and these have generally focused on the assessment of anatomically-intact sinonasal tracts.[2–4] In addition to the potentially subjective process of describing anatomic and mucosal findings, the presence of disrupted anatomy in the post-surgical patient may prove especially challenging. Endoscopy is routinely used to inspect the patency of surgically-opened sinuses and to evaluate patients for the need for revision surgery. Nonetheless, the objectivity and validity of nasal endoscopy in evaluating patients who have had prior sinus surgery has not been well studied.

Recognizing the potential for error and/or variability is important when interpreting the results of any examination or diagnostic test. Although nasal endoscopy is considered an objective procedure, the reproducibility of findings and the agreement between observers may be associated with significant variability.[5] The ostiomeatal complex region is often regarded as fundamentally important to the pathogenesis of chronic rhinosinusitis, yet is also a region of intricate anatomy whose subtle physical findings may be subject to variable diagnostic certainty. Therefore, the clinical assessment of this region and the associated sinuses was identified as a potentially important area of investigation.

The aim of this study was to determine the inter-rater agreement of nasal endoscopy in patients who have had prior endoscopic sinus surgery (ESS). The identification and interpretation of specific areas of anatomic abnormalities and inflammatory pathology may be useful in determining the need for revision sinus surgery. In addition, endoscopy allows the clinician to follow the course of chronic rhinosinusitis findings in patients with prior sinus surgery irrespective of the need for revision surgery. Establishment of the reliability of nasal endoscopy for these purposes may be an important consideration in refining the diagnostic evaluation in patients with chronic rhinosinusitis.

METHODS

Study Design

Approval for this study was obtained from the Institutional Review Board of Weill Cornell Medical College prior to patient enrollment. Patients were prospectively enrolled from October 2010 to March 2011 from the clinical practice of one of the senior investigators (VKA). Consecutive patients, 18 years of age or older with a prior history of at least one ESS for chronic rhinosinusitis, were included. Each patient completed the 22-item Sinonasal

Outcome Test (SNOT-22), prior to examination, to provide an assessment of symptom burden.[6] Each patient subsequently underwent computerized tomography (CT) of the paranasal sinuses within 1 month of the nasal endoscopy as part of routine clinical care. The CT images were reviewed and interpreted according to the Lund-Mackay scale and a severity value was assigned by a single study author (EDM).[7] The patient's sex and age were recorded as well.

Each study participant underwent diagnostic nasal endoscopy at the time of first presentation. The nasal cavity was prepared by topical decongestion with 1% phenylephrine. All examinations were performed by a single clinician (VKA). Endoscopy was performed with the patient in seated position using a 0-degree rigid endoscope (Storz, Tuttlingen, Germany). All examinations were digitally videographed (D-Scope EDB Advanced V4.5 Workstation (Medical Digital Developers, LLC, New York, NY). The recordings were edited using commercially available media software (Windows Live Movie Maker 2011, Microsoft Corp, Redmond, WA). Each side of the examination (right or left) was considered as a separate video. A grading instrument with 10 separate items encompassing inflammatory and anatomic issues relevant for chronic sinusitis in patients with a prior history of ESS was developed for the purpose of this study (Table 1).

A total of 28 patients provided consent and were enrolled for study. Examinations were performed unilaterally in 14 cases and bilaterally in 14 cases, for a total of 42 separate videos. Preliminary review was conducted by a single investigator (EDM) using the data collection instrument. For each video, the adequacy of visualization was noted for each of the 10 items. The purpose of this review was to minimize the occurrence of non-responses during formal data collection. Twenty-eight videos were excluded that did not permit adequate visualization of all 10 items of interest. These exclusions were secondary to multiple factors, including the presence of scarring or other anatomic barriers to visualization, the exclusive use of a 0-degree endoscope to visualize all structures, and the real-time process of clinical endoscopy, wherein the operator obtains split-second visual information that may be too brief to be adequately captured for second-party viewing. Following this review, 14 videos were included for study. The videos were then de-identified and assigned a numeric identifier using a random number generator (Excel 2007, Microsoft).

Five independent, academic rhinologists reviewed the videos (TLS, BAS, PHH, JAS, AT). Each reviewer was provided with a CD-ROM containing the 14 separate video clips identified only by a number. Each reviewer was also given an instruction sheet and 14 data sheets, one for each video. The reviewers were encouraged to pause, rewind or replay the videos as much as necessary to make their assessments. For each video, the reviewers were instructed to select only one answer for each of the 10 questions. The data sheets were then returned by postal mail to the senior study investigator.

Statistical Analysis

Descriptive statistics were used for all patient characteristics and SNOT-22 including means (standard deviation) and score range. Spearman's correlation coefficient was used to assess non-linear associations between age, CT score, and SNOT-22 survey responses. Inter-rater agreement was assessed using unweighted Fleiss' kappa statistic (K_f) for nominal data from each unilateral video examination (n=14), as well as asymptotic standard errors (ASE) and two-sided 95% confidence intervals (CI).[8] Fleiss' kappa is an compilation measure to determine the degree of agreement between all raters above the level expected by chance and values range from −1.0 to 1.0, where 0.0 is reflective of the null hypothesis, ("agreement expected by chance"), 1.0 indicates "perfect agreement" and −1.0 indicates perfect disagreement.[9] P-values for each kappa statistic were calculated to determine if inter-

rater agreement was due to chance observations alone. P-values less than 0.05 denote that the observed rater agreement was not due to chance. In addition, since there are many factors that can influence the magnitude of kappa coefficients, we calculated the proportion of observed and expected rater agreement to supplement each kappa statistic. Additional reliability estimates included interclass correlations between all observer pairs for each grading scale attribute using two-tailed Spearman's rank coefficients. Conventional interpretations of kappa values are included in Table 2.[10]

Multi-rater kappa coefficients are complex and influenced by prevalence and bias.[11] Mean prevalence and observer bias indices were calculated to further discern discrepancy found between kappa values and prevalence of observed rater agreement. Due to the categorical nature of certain grading scale options, endoscopic characteristics with three or more response category options were collapsed into dichotomous measures. Any attribute that a rater was unable to assess during each video examination was removed from prevalence or bias index calculations. Prevalence and bias indices were evaluated and averaged for each distinct pairs of investigators (n=10). Mean prevalence-adjusted, bias-adjusted kappa (PABAK) values were calculated to further account for incongruity between unadjusted kappa values and the percentage of observed agreement for all endoscopy characteristics.[12] Statistical analyses were accomplished using commercially available statistical (SPSS v. 19.0, SPSS, Chicago, IL) and database spreadsheet software (Excel 2007, Microsoft Inc., Redmond, WA).

RESULTS

Patient and disease demographics

A total of 8 male and 6 female adult study patients with chronic rhinosinusitis were included for study. Mean (SD) subjects age was 51.9 (15.5) years of age [age range: 21 – 71] and all obtained coronal plane CT imaging per the standard of care for chronic rhinosinusitis with mean Lund-Mackay scores of 6.93 (3.95) units [score range: 2 – 14]. All patients completed SNOT-22 symptom questionnaires with mean score 36.9 (22.1) units [score range: 8 – 72]. No significant non-linear correlations between age, CT score, and SNOT-22 scores were found (all $p>0.185$).

Reviewer Responses and Reliability Estimations

The frequencies of total observer responses for each subject (n=70) and the percentage of missing observer values are listed in Table 1. Out of 350 total responses among all observers, 24 (6.9%) were "unable to assess." All Fleiss' kappa values, ASE, 95% CI, Z-scores, and corresponding p-values are listed in Table 3. The percentage agreement between independent reviewers was varied across each attribute of the endoscopic grading scale with strongest agreement found for assessment of sinonasal polyps within the middle meatus. With the exception of middle turbinate (MT) mucosal changes ($p=0.178$) and status of the uncinate process ($p=0.125$), all corresponding p-values support the alternative hypothesis that the kappa level of agreement was not due to observational chance. For all grading scale attributes the prevalence of observed agreement was higher than expected agreement (Table 4). Interclass correlation coefficients most frequently showed statistically significant agreement between possible observer pairs for MT integrity (8 of 10, all $p \leq 0.036$), polyps within the middle meatus (9 of 10; all $p \leq 0.008$) and discharge (8 of 10; all $p \leq 0.049$) (Table 5).

For PABAK calculations, middle turbinate (MT) integrity was recategorized as "MT intact" or "MT partially/completely resected", MT mucosal changes were dichotomized into "normal" or "abnormal", the UP into "completely resected" or "partially resected/intact/

nearly-intact", middle meatus patency into "patent" or "partially/completely obstructed", the ethmoid sinus cavity as "patent" or "partially/completely obstructed", maxillary sinus antrostomy into "present, patent" or "present, stenotic/not present", and discharge into "absent" or "clear, thin/thick, purulent". The mean prevalence and bias indices for all endoscopic attributes are listed in Table 6. Higher average prevalence index and lower bias index values typically indicate more substantial observer agreement even if kappa statistics are relatively low. After adjustment for both mean indices, K_f values increased for 9 of 10 endoscopic exam attributes.

DISCUSSION

Determination of inter-rater agreement is an important component of the validation process for any diagnostic tool. However, the study of inter-rater agreement for diagnostic nasal endoscopy has been the subject of limited study. Annamalai et al[3] reported on the inter-rater agreement of endoscopic findings for patients with a variety of nasal complaints as recorded by two independent reviewers. That study, which used the Lund-Kennedy scoring system, found very good agreement for the presence of polyps and discharge and moderate agreement for edema. They also found that a significant proportion of patients had no abnormal findings on nasal endoscopy, which reflects the fact that a diagnosis of rhinosinusitis was not an inclusion criterion. Raithatha et al[5] described inter-rater agreement among 5 independent rhinologists for the examination of patients with a diagnosis of chronic rhinosinusitis. That study found several areas of significant agreement and other areas of partial agreement, which supported the overall precision and reproducibility of nasal endoscopy among clinicians.

The present study reports the inter-rater agreement for endoscopy findings in patients with a prior history of ESS for chronic rhinosinusitis as determined by 5 independent reviewers. The results indicate variability in the level of agreement among the different categories. Each of these examination components describes a potentially important finding related to chronic rhinosinusitis and the anatomic adequacy of prior surgery. The latter category has significant implications for the need for future surgery. The strongest agreement was found for the presence of polyps within the middle meatus, which is likewise an important indicator of the presence of disease in a location that may affect the function of adjacent paranasal sinuses. This agrees with the findings of the study by Raithatha et al,[5] which found the highest level of agreement for nasal polyps (K_f=0.693). Agreement was substantially more limited for middle meatus patency and presence of the uncinate process, both of which describe anatomic findings that may be substantially altered in the post-ESS patient. Poor-to-fair agreement was also found for findings related to mucosal pathology, including MT mucosal changes, discharge, and synechiae involving the middle meatus. These findings contrast with the study by Raithatha et al,[5] which found moderate agreement for both MT mucosal changes and nasal discharge. This discrepancy suggests that, compared to examination of the anatomically intact sinonasal tract, identification and description of inflammatory findings may be more challenging in the post-ESS patient.

Several interpretations of these study findings are possible. Although visual inspection is the fundamental aspect of nasal endoscopy, useful information may also be obtained by palpation, displacement, and the selective use of angled endoscopes. The lack of haptic feedback and other cues may limit the ability of a reviewer to accurately interpret the endoscopic findings. An additional consideration may relate to bias created by the surgical philosophy of the interpreting clinician. The interpretation may also have been hindered by the lack of universally-accepted terminology to describe the presence and degree of anatomic and inflammatory findings, particularly in the setting of prior sinus surgery. This highlights the need for establishment of standardized definitions and a validated instrument

for reporting the findings of nasal endoscopy. Such a tool would ideally include all essential parts of the examination and specify the presence and degree of pathologic findings. Limitations in the video capture system, video editing and playback methods may have also limited the visualization of the examination.

Surprisingly few grading scales for nasal endoscopy findings have been proposed in the literature. Lund and Kennedy[1] developed a ten-point scale that recorded the presence and severity of inflammatory sinonasal findings. This system was initially described for use on patients with intact sinonasal anatomy, although it has also been applied for the comparison of pre- and post-surgical patients.[13] Moreover, this system has never been systematically validated. Wright and Agrawal[4] proposed a system to describe the endoscopic assessment of sinus cavities after sinus surgery. While this system was subjected to limited validation, it has not been widely adopted for clinical use. For each of these grading scales, clinical utility is limited by uncertainty about reliability and applicability to subjects outside of the study group. The development of a grading scale for endoscopic evaluation after sinus surgery may be challenged by a lack of consensus about anatomical terms and variability in the extent of surgery performed among practitioners.

The instrument used in the present study was not intended as a validated endoscopic grading scale for clinical use, but rather to provide a standardized means of comparing examination findings between observers in a research setting. The items on the scoring sheet include measures of sinus inflammation in addition to structural issues related to the post-surgical anatomy. The latter categories reflect how patent the sinuses are in the opinion of each reviewer, and have possible implications for demonstrating agreement on the need for revision surgery. Cumulatively, these different categories of endoscopic findings seek to query how individual rhinologists view the inflammatory sinus findings and structural post-surgical findings in a cohort of patients with a prior history of ESS and chronic rhinosinusitis. Inter-rater reliability may be improved in the future by the application of standardized terminology among clinicians. Further study of nasal endoscopy using a validated scoring system would help to illuminate this relatively unexplored area of interest.

The present study supports a role for the digital recording and transmission of endoscopic findings that has been described in previous studies.[5,14] Digital capture preserves the visual fidelity of the original examination and enables repeated viewing, pausing and magnification of the examination images. Of the videos selected after initial review, only 6.9% of all responses were “unable to assess,” suggesting that digital video recording was beneficial to interpretation. Furthermore, digital media allows the participation of multiple clinicians in the review and interpretation of endoscopy findings. The use of this technology as an adjunct to nasal endoscopy in certain cases may improve the diagnostic yield and help to refine the subsequent treatment.

Several limitations of the present study bear mention. The cohort of patients enrolled for study was small; however, significant p-values were obtained for the majority of kappa values, which suggests the study sample size was adequate. Furthermore, because the Fleiss’ kappa statistic is sensitive to both attribute prevalence and homogeneity of the sample, the PABAK was used to adjust for these differences. The use of 5 independent reviewers likely contributed to a more accurate estimate of reliability; however, all reviewers were academic rhinologists, which may have introduced bias that would not have been present had the reviewers been general otolaryngologists or other sub-specialists. In addition, as stated previously, although all patients had prior ESS, the degree and extent of surgery was subject to wide variability. Finally, the inclusion of each examination item was made by expert opinion and was not previously validated. Further studies are required to evaluate the interpretation and standardization of nasal endoscopy for post-surgical patients.

CONCLUSIONS

Nasal endoscopy for the evaluation of the post-ESS patient is subject to variable degrees of inter-rater reliability. Moderate-to-good agreement was found for certain categories of endoscopic findings such as polyps, middle turbinate position and integrity, and ethmoid and maxillary sinus patency. Other categories showed relatively little agreement, including findings related to sinonasal inflammation. Cumulatively, these findings support the need for standardization and validation of nasal endoscopy as a diagnostic tool.

References

1. Rosenfeld RM, Andres D, Bhattacharyya N, Cheung D, Eisenberg S, Ganiats TG, et al. Clinical practice guidelines: Adult sinusitis. Otolaryngol Head Neck Surg. 2007; 137:S1–S31. [PubMed: 17761281]
2. Lund VJ, Kennedy DW. Quantification for staging sinusitis. The Staging and Therapy Group. Ann Otol Rhinol Laryngol Suppl. 1995; 167:17–21. [PubMed: 7574265]
3. Annamalai S, Davis J, et al. How subjective is nasal endoscopy: a study of interrater agreement using the Lund and Mackay scoring system. Am J Rhinol. 2004; 18:301–3. [PubMed: 15586801]
4. Wright E, Agrawal S. Impact of perioperative systemic steroids on surgical outcomes in patients with chroic rhinosinusitis with polyposis: evaluation with the novel perioperative sinus endoscopy (POSE) scoring system. Laryngoscope. 2007; 117(Suppl 115):1–28. [PubMed: 18075447]
5. Raithatha R, Anand VK, Mace JC, Smith TL, Schaberg MR, Nyquist GG, Hwang PH, Senior BA, Stankiewicz JA, Tabaee A. Inter-rater agreement of nasal endoscopy for chronic rhinosinusitis. Int Forum Allergy Rhinol. 2012 Jan 17. [Epub ahead of print].
6. Hopkins C, Gillett S, Slack R, Lund VJ, Browne JP. Psychometric validity of the 22-item Sinonasal Outcome Test. Clin Otolaryngol. 2009; 34:447–54. [PubMed: 19793277]
7. Lund VJ, Mackay IS. Staging in rhinosinusitis. Rhinology. 1993; 107:183–4. [PubMed: 8140385]
8. Fleiss JL, Nee JCM, Landis JR. Large sample variance of kappa in the case of different sets of raters. Psychological Bulletin. 1979; 86(5):974–77.
9. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull. 1971; 76(5):378–382.
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33:159–174. [PubMed: 843571]
11. Sim J, Wright CC. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. Phys Ther. 2005; 85:257–268. [PubMed: 15733050]
12. Byrt T, Bishop J, Carlin JB. Bias, prevalence, and kappa. J Clin Epidemiol. 1993; 46(5):423–429. [PubMed: 8501467]
13. Mace J, Michael YL, Carlson NE, Litvack JR, Smith TL. Correlations between endoscopy score and quality of life changes after sinus surgery. Arch Otolaryngol Head Neck Surg. 2010; 136:340–6. [PubMed: 20403849]
14. Tabaee A, Johnson P, Anand VK. Office-based digital video endoscopy in otolaryngology. Laryngoscope. 2006; 116:1523–5. [PubMed: 16885766]

Table 1

Distribution of responses using the study grading instrument. Frequency of observer responses are given for each attribute of the endoscopic examination in patients with a history of endoscopic sinus surgery for chronic rhinosinusitis.

Attributes:	Total Scores (n=70)
A. Middle Turbinate (MT) Position	**n (%)**
○ MT does not obstruct MM	54 (77.1%)
○ MT obstructs the MM	16 (22.9%)
○ Unable to assess	0
B. Middle Turbinate Integrity	
○ MT Intact	25 (35.7%)
○ MT partially resected	34 (48.6%)
○ MT completely/near-completely resected	11 (15.7%)
○ Unable to assess	0
C. Middle Turbinate (MT) Mucosal Changes	
○ Normal	23 (32.9%)
○ Edema of MT	45 (64.3%)
○ Polypoid MT	1 (1.4%)
○ Unable to Assess	1 (1.4%)
D. Uncinate Process (UP)	
○ Completely Resected UP	35 (50.0%)
○ Partially Resected UP	17 (24.3%)
○ Intact or Nearly-intact UP	6 (8.6%)
○ Unable to Assess	12 (17.1%)
E. Middle Meatus Patency	
○ Patent	39 (55.7%)
○ Partially Obstructed	29 (41.4%)
○ Completely Obstructed	2 (2.9%)
○ Unable to Assess	0
F. Ethmoid Sinus Cavity	**n (%)**
○ Patent	21 (30.0%)
○ Partially Obstructed	44 (62.9%)
○ Completely Obstructed	5 (7.1%)
○ Unable to Assess	0
G. Maxillary Sinus Antrostomy	
○ Present, patent	49 (70.0%)

Attributes:	Total Scores (n=70)
○ Present, Stenotic	6 (8.6%)
○ Not Present	4 (5.7%)
○ Unable to Assess	11 (15.7%)
H. Synechiae Involving Middle Meatus	
○ Absent	31 (44.3%)
○ Present	39 (55.7%)
○ Unable to Assess	0
I. Polyps Within Middle Meatus	
○ Absent	62 (88.6%)
○ Present	8 (11.4%)
○ Unable to Assess	0
J. Discharge	
○ Absent	27 (38.6%)
○ Clear, Thin	26 (37.1%)
○ Thick, Purulent	17 (24.3%)
○ Unable to Assess	0

Table 2

Conventional interpretation of Fleiss' kappa statistic (K_f) (after Landis & Koch[10]).

K_f Range	Categorized Level of Agreement
<0.00 – 0.20	Poor / slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Strong / substantial
0.81 – 1.00	Almost perfect

Table 3

Unadjusted Fleiss' kappa statistic (K_f) values for each component of nasal endoscopy of patients with a history of endoscopic sinus surgery for chronic rhinosinusitis.

Attribute	K_f	ASE	95% CI [LL, UL]	Z-value	P-value
A. Middle Turbinate Position	0.311	0.085	[0.146, 0.477]	3.684	0.0002
B. Middle Turbinate Integrity	0.486	0.063	[0.362, 0.611]	7.668	<0.0001
C. Middle Turbinate Mucosal Changes	0.104	0.077	[−0.047, 0.256]	1.347	0.1781
D. Uncinate Process	0.083	0.054	[−0.023, 0.189]	1.534	0.1249
E. Middle Meatus Patency	0.171	0.078	[−0.018, 0.324]	2.195	0.0281
F. Ethmoid Sinus Cavity	0.215	0.070	[−0.078, 0.353]	3.070	0.0021
G. Maxillary Sinus Antrostomy	0.323	0.058	[0.209, 0.436]	5.574	<0.0001
H. Synechiae Involving Middle Meatus	0.247	0.085	[0.082, 0.413]	2.926	0.0034
I. Polyps Within Middle Meatus	0.718	0.085	[0.552, 0.883]	8.492	<0.0001
J. Discharge	0.334	0.060	[0.215, 0.453]	5.522	<0.0001

K_f, Fleiss' kappa statistic; ASE, asymptotic standard error; CI, confidence interval; LL, lower limit of 95% confidence interval; UL, upper limit of 95% confidence interval

Table 4

Prevalence of observed and expected agreement for each attribute of the endoscopic examination.

Attribute	P_o	P_e
A. Middle Turbinate Position	75.7	64.7
B. Middle Turbinate Integrity	68.6	38.8
C. Middle Turbinate Mucosal Changes	57.1	52.2
D. Uncinate Process	40.0	34.6
E. Middle Meatus Patency	57.1	48.3
F. Ethmoid Sinus Cavity	60.0	49.0
G. Maxillary Sinus Antrostomy	67.9	52.5
H. Synechiae Involving Middle Meatus	62.9	50.7
I. Polyps Within Middle Meatus	94.3	79.8
J. Discharge	56.4	34.6

Po, observed agreement (%); Pe, expected agreement (%)

Table 5

Frequency of significant non-linear correlation for discrete observer pairs (n=10) for each endoscopy characteristic.

	MTP	MTI	MTMC	UP	MMP	ESC	MSA	SIMM	PwMM	DIS
Frequency of significant agreement	3/10 (30%)	7/10 (70%)	0/10 (10%)	1/10 (10%)	3/10 (30%)	4/10 (40%)	4/10 (40%)	3/10 (30%)	9/10 (90%)	8/10 (80%)

MTP, middle turbinate position; MTI, middle turbinate integrity; MTMC, middle turbinate mucosal changes; UP, uncinate process; MMP, middle meatus patency; ESC, ethmoid sinus cavity; MSA, maxillary sinus antrostomy; SIMM, synechiae involving middle meatus; PwMM, polyps within middle meatus; DIS, discharge. Frequencies represent the number of observer pairs who report significant agreement ($p<0.05$) for each eattribute.

Table 6

Prevalence-adjusted, bias-adjusted Fleiss' kappa statistic (PABAK) values for each component of nasal endoscopy of patients with a history of endoscopic sinus surgery for chronic rhinosinusitis.

	Mean Prevalence Index	Mean Bias Index	Mean PABAK	Agreement Level
A. Middle Turbinate Position	0.543	0.186	0.443	Moderate
B. Middle Turbinate Integrity	0.286	0.100	0.543	Moderate
C. Middle Turbinate Mucosal Changes	0.359	0.285	0.148	Poor/slight
D. Uncinate Process	0.392	0.390	0.126	Poor/slight
E. Middle Meatus Patency	0.271	0.357	0.229	Fair
F. Ethmoid Sinus Cavity	0.414	0.271	0.429	Moderate
G. Maxillary Sinus Antrostomy	0.662	0.124	0.593	Moderate
H. Synechiae Involving Middle Meatus	0.257	0.300	0.257	Fair
I. Polyps Within Middle Meatus	0.771	0.043	0.886	Almost perfect
J. Discharge	0.229	0.186	0.314	Fair