The Edematous and Erythematous Airway Does Not Denote Pathologic Gastroesophageal Reflux

Rachel Rosen, MD, MPH[1], **Paul D. Mitchell, MS**[2], **Janine Amirault, BS**[1], **Manali Amin, MD**[3], **Karen Watters, MD**[4], **Reza Rahbar, MD**[4]

[1]Aerodigestive Center, Division of Gastroenterology and Nutrition, Boston Children's Hospital, Boston, MA;

[2]Clinical Research Center, Children's Hospital Boston, Boston, MA;

[3]DuPage Children's Ear, Nose, and Throat and Allergy, Naperville, IL;

[4]Department of Otolaryngology, Boston Children's Hospital, Boston, MA

Abstract

Objective—To determine if the reflux finding score (RFS), a validated score for airway inflammation, correlates with gastroesophageal reflux measured by multichannel intraluminal impedance (MII) testing, endoscopy, and quality of life scores.

Study design—We performed a prospective, cross-sectional cohort study of 77 children with chronic cough undergoing direct laryngoscopy and bronchoscopy, esophagogastroduodenoscopy, and MII testing with pH (pH-MII) between 2006 and 2011. Airway examinations were videotaped and reviewed by 3 blinded otolaryngologists each of whom assigned RFS to the airways. RFS were compared with the results of reflux testing (endoscopy, MII, symptom scores). An intraclass correlation coefficient was calculated for the degree of agreement between otolaryngologists' RFS. Receiver operating characteristic curves were created to determine the sensitivity of the RFS. Spearman correlation was calculated between the RFS and reflux measurements by pH-MII.

Results—The mean ± SD RFS was 12 ± 4. There was no correlation between pH-MII variables and mean RFS ($|r| < 0.15$). The concordance correlation coefficient for RFS between otolaryngologists was low (intraclass correlation coefficient = 0.32). Using pH-metry as a gold standard, the positive predictive value for the RFS was 29%. Using MII as the gold standard, the positive predictive value for the RFS was 40%. There was no difference in the mean RFS in patients with (12 ± 4) and without (12 ± 3) esophagitis ($P = .9$). There was no correlation between RFS and quality of life scores ($|r| < 0.15$, $P > .3$).

Conclusions—The RFS cannot predict pathologic gastroesophageal reflux and an airway examination should not be used as a basis for prescribing gastroesophageal reflux therapies.

As part of the evaluation for patients with cough, hoarseness, croup, and other respiratory symptoms, otolaryngologists frequently perform a direct laryngoscopy and bronchoscopy

Reprint requests: Rachel Rosen, MD, Aerodigestive Center, Division of Gastroenterology and Nutrition, Boston Children's Hospital, 300 Longwood Ave, Boston, MA 02115. Rachel.rosen@childrens.harvard.edu.

(DLB) and, when a red, inflamed larynx is seen, otolaryngologists frequently implicate gastroesophageal reflux disease (GERD) as a cause for the laryngeal findings. As a result, otolaryngologists have become one of the main prescribers of proton pump inhibitors for adults and children despite a lack of evidence suggesting causality.[1]

Adult studies show an inconsistent relationship between acid reflux and laryngeal findings and these studies are limited by their use of pH-metry, which lacks the sensitivity needed to detect proximal reflux and nonacid reflux events.[2,3] Overcoming this limitation, multichannel intraluminal impedance (MII) with pH (pH-MII) offers the same strengths as pH-metry but offers the advantage that the catheters can accurately detect both full column reflux and acid and nonacid reflux events. As a result, at many institutions, pH-MII has become the gold standard tool to measure full column reflux, which has the potential to reach the oropharynx and larynx. There is only a single adult study using pH-MII to determine the relationship between laryngeal findings and full column acid and nonacid reflux events.[3] In this study of 142 healthy volunteers who underwent both pH-MII and laryngoscopy, there was no correlation between reflux events and the appearance of the airway scored using the reflux findings score (RFS), a validated scoring system of laryngeal findings suggestive of GERD.[3] There are no studies in symptomatic adults or any studies in pediatrics.

Determining if there is a significant, positive relationship between the appearance of the larynx and gastroesophageal reflux is critical not only for diagnostic purposes but most importantly, to ensure that only patients with documented extraesophageal reflux disease receive proton pump inhibitor (PPI) therapy. This is particularly critical in patients with respiratory symptoms, as PPI use has been associated with increased risk of respiratory infections, which may actually worsen the symptoms for which they are prescribed.[4,5]

It is the goal of this study to determine (1) the relationship between reflux events by pH-MII and the RFS[6]; and (2) the interrater reliability of the RFS among otolaryngologists of different experience levels.

Methods

Between 2008 and 2014, we prospectively recruited 77 consecutive patients (who were not currently receiving acid suppression therapy) undergoing direct laryngoscopy, endoscopy, and pH-MII testing at Boston Children's Hospital for the evaluation of chronic cough. The study was approved by Boston Children's Hospital institutional review board and all parents signed consent forms to participate. Parents of recruited patients completed the Pediatric Quality of Life Inventory (PedsQL) and the PedsQL Gastrointestinal Symptoms Module.[7,8] At the time of endoscopy, DLBs were videotaped and deidentified. Each patient underwent, immediately after the DLB, an esophagogastroduodenoscopy with biopsies and pH-MII placement. Three otolaryngologists (1 junior, 1 midlevel, and 1 senior) independently reviewed the airway videos and assigned each airway a total RFS including all of its 8 subscores. Subscores include subglottic edema (0 = absent, 2 = present), ventricular obliteration (0 = none, 2 = partial, 4 = complete), erythema (0 = none, 2 = arytenoid only, 4 = diffuse), vocal fold edema (0 = none, 1 = mild, 2 = moderate, 3 = severe, 4 = polypoid),

diffuse laryngeal edema (0 = none, 1 = mild, 2 = moderate, 3 = severe, 4 = obstructing), posterior commissure hypertrophy (0 = none, 1 = mild, 2 = moderate, 3 = severe, 4 = obstructing), granuloma presences (0 = absent, 2 = present), and thick mucus presence (0 = absent, 2 = present). The total RFS and the subscores were compared with the pH-MII reflux measurements including the total number of reflux episodes, the number of acid and nonacid reflux events, the percentage of the study time during which the pH<4, the percentage of time reflux was in the proximal and distal esophagus (sum of the bolus clearance times at each site divided by the total study duration), and the percentage of reflux episodes that were full column. The pH portion of the study was considered abnormal if the pH was <4 for >6% of the study duration, and the MII portion was considered abnormal if the total number of reflux episodes were >73.[9,10] For each pH-MII tracing, pH-MII tracings were analyzed by 1 expert reviewer after running automated analysis software.

Total RFS was calculated as the mean of the 3 otolaryngologists, rounded to ensure the optimal cut point would be an integer. A sensitivity analysis was conducted using the mean RFS of the 3 otolaryngologists. Agreement among otolaryngologists was assessed by the concordance correlation coefficient with 95% CI. The correlation of RFS with number of reflux episodes is shown in a scatterplot of normalized RFS by normalized reflux episodes, where the data have been transformed to a standard normal distribution.

An empirical receiver operating characteristic curve was used to characterize the relation of total RFS with abnormal pH-metry from the pH portions of the study and with abnormal numbers of reflux events by the MII portion of the study. An optimal cut point for each curve was determined by Youden's Index and used to determine the positive predictive value.

The association of RFS items with reflux measurements was assessed with the Spearman rank correlation coefficient. Means are presented as means ± SD. All data analysis was performed with SAS v 9.4 (SAS Institute, Cary, North Carolina).

Results

The mean age (±SD) of patients was 6.5 ± 3.7 years. The mean RFS of the 77 patients was 12 ± 3. Eighteen percent of patients had an abnormal MII study, and 28% had an abnormal pH portion of the study. The mean number of total, acid, and nonacid reflux episodes per 24-hour period was 48 ± 27, 25 ± 15, and 23 ± 23 episodes, respectively. The mean percentage time pH <4 per 24 hour study was 4.8 ± 5.6. None of the patients had evidence of erosive esophagitis endoscopically, but 21% had histological evidence of reflux esophagitis, and 5% had evidence of eosinophilic esophagitis. The distribution of the most severe findings for each of the RFS characteristics is shown in Figure 1.

The concordance correlation coefficient was low at 0.32 (95% CI 0.22, 0.42) indicating poor agreement among otolaryngologists regarding the RFS. There was no correlation between any reflux measurement (total number of reflux episodes, number of acid or nonacid reflux episodes, percentage of time reflux was in the distal or proximal esophagus) and the mean RFS (|r| < 0.14, P > .21). Sixty-six patients completed symptom and quality of life

questionnaires; there was no significant correlation between the mean RFS and the PedsQL or the PedsQL Gastrointestinal Symptoms Module ($|r| < 0.27$, $P > .06$). There was no difference the mean RFS in patients with (12 ± 4) and without (12 ± 3) esophagitis ($P = .9$).

There was no relationship between abnormal pH and MII testing and RFS at every experience level as shown in the Table. There is a lack of relationship between the RFS at every experience level and the total reflux burden as shown in Figure 2.

The receiver operating characteristic curves for the RFS based on an abnormal pH probe or MII are shown in Figure 3. Using pH-metry as the gold standard, the sensitivity, specificity, positive predictive value and negative predictive value, and accuracy of the RFS with a cut point of >7 was 95%, 9%, 29%, 83%, and 33%, respectively. Using MII as the gold standard, the sensitivity, specificity, positive predictive value and negative predictive value, and accuracy of the RFS with a cut point of >16 was 14%, 95%, 44%, 83%, and 81%, respectively. Importantly, the positive predictive values of RFS in predicting pathologic reflux by pH-MII are extremely low supporting the idea that the laryngeal appearance cannot diagnose pathologic reflux.

Discussion

Our results show no relationship between airway inflammation and either pathologic reflux measured either by pH probe or MII or full column acid or nonacid reflux events. We also found significant disagreement between otolaryngologists on the airway findings suggesting subjectivity in the examination which limits the utility of the test. This study has several important strengths over the existing literature, including (1) the prospective recruitment of all children undergoing combined laryngoscopy, endoscopy, and pH-MII testing; (2) the use of multiple methods to diagnose reflux (pH-MII, symptom scores and endoscopy); and (3) the inclusion of 3 otolaryngologists of different experience levels blindly scoring the airway.

The implications of this study are great at a time when acid suppression medications such as PPIs are being prescribed at alarmingly high rates in the pediatric population, particularly for the treatment of extraesophageal symptoms.[1,11,12] Based on this study which shows that no component of the airway exam predicts pathologic reflux defined by pH-MII, endoscopy, and symptom scores, PPIs or other acid suppression should not be prescribed based on the airway appearance alone because the appearance does not denote pathologic gastroesophageal reflux. The risk of these medications in the population of children with respiratory symptoms is particularly acute as these medications carry with them an increased risk of respiratory infections including upper respiratory tract infections, bronchitis, and pneumonia all of which may worsen the patient's underlying symptoms.[4,5] The results of this study have significant public health implications as otolaryngologists are currently one of the biggest prescribers of these medications in children.[1]

The results of this study are not only significant from a public health perspective but also from a cost perspective. Currently in the US, the cost of diagnosing and treating adult patients with extraesophageal reflux symptoms is on par with cancer, approximately 50 billion dollars per year, and this staggering number does not even include the pediatric costs.

[13] These costs are driven by 2 factors: the cost of PPIs and the cost of diagnostic testing. Currently in the US, patients undergo a mean of 5 procedures as part of their diagnostic evaluation for extraesophageal reflux disease. The results of our study suggest that DLB has a low positive predictive value for diagnosing GERD, and it should be removed from the list of diagnostic tests for extraesophageal reflux disease.

One possible argument against our findings is that the cut-off values for pathologic reflux using pH-MII may not accurately reflect the amount of reflux needed to alter the appearance of the larynx. However, even when we used each of the reflux values as continuous variables to assess for any association between any amount of reflux and appearance, we still could not find a significant relationship or a better pathologic cut-off point for pathologic reflux. Further supporting our finding that there is a lack of relationship between the RFS and reflux, acid suppression trials fail to consistently improve laryngeal findings; in a meta-analysis by Guo et al,[14] the authors found that the RFS did not predict PPI responsiveness; only 1 out of 4 studies showed a positive relationship between RFS and PPI response again suggesting a lack of a relationship.

Another possible explanation for a lack of relationship between reflux and the RFS in our study is the use of the RFS is flawed; the RFS has only been studied in adults, and studies of the RFS have consistently shown lack of interrater and intrarater reliability in adult studies. [2,3] Our results support the adult literature that there is poor interrater reliability between otolaryngologists, highlighting the limitations of these examinations. We further add to the literature by showing that this insensitivity of the examination is independent of physician experience. Recognizing that there may be limitations to the RFS, we redid our analyses using only the subscore of laryngeal erythema, the most commonly identified abnormality in clinical practice that triggers referral to gastroenterology or PPI prescribing. Even with this subgroup analysis, we still did not see a relationship between erythema and any of the reflux variables. We feel confident, using both the RFS and the erythema score alone, that there is no relationship between any airway finding and reflux events, and the airway examination should not be used as a diagnostic test for reflux.

There are 2 pediatric studies that have addressed this relationship between pathologic reflux and the appearance of the airway. Carr et al[15] compared the visual appearance of the larynx of children with reflux diagnosed by any of a variety of tests (upper gastrointestinal series, scintigraphy, endoscopy, or pH probe). In this study, the authors found the combination of postglottic edema, arytenoid edema, and vocal cord edema had a sensitivity of 75% and specificity of 67% compared with reflux diagnosed by any of 4 methods. This study by Carr et al[15] has several important limitations. First, the tests used to diagnose reflux in this study are insensitive and no longer used to diagnose reflux because of this lack of sensitivity. Second, the scoring system for the airway was not validated and did not include a variable for erythema, which is one of the primary findings on laryngoscopy that prompts referral to a gastroenterologist or a prescription for a PPI. In another pediatric study by Simons et al[16] of 36 children with dysphonia and cough, there were no significant correlations between the RFS and any reflux measurements by esophageal biopsies or by reflux symptom scores, again calling into question the utility of the RFS to diagnose reflux.[16] Unfortunately,

however, only endoscopy and reflux symptoms were used to make the diagnosis, which limits its generalizability.

Although our study adds significantly to the literature, there are several limitations. First, we only included symptomatic patients suggesting that we may have a biased population with a high pretest probability of detecting reflux. However, in our study, only 20%–27% of patients had pathologic reflux (by MII or pH) making this bias much less likely. Furthermore, even if rates of reflux in this population were significantly higher, we would have expected a more significant correlation paralleling our high rates of erythematous airways. Therefore, even with this potential bias, we still could not find a significant relationship.

A second limitation is that the RFS was designed for adults yet we applied it to children, which may or may not be a valid tool in this population. However, even when we redid our analyses with common pediatric subscores, we still failed to determine a relationship between reflux characteristics and these subgroup findings suggesting that even a modification of this score would not improve its sensitivity and validates our original conclusions.

In summary, laryngeal findings should not be used to diagnose pathologic reflux. The corollary to this is that acid suppression medications should not be prescribed based on airway findings alone because the findings do not correlate with any reflux variable.

Acknowledgments

Glossary

DLB	Direct laryngoscopy and bronchoscopy
GERD	Gastroesophageal reflux disease
MII	Multichannel intraluminal impedance
pH-MII	Multichannel intraluminal impedance with pH
PedsQL	Pediatric Quality of Life Inventory
PPI	Proton pump inhibitor
RFS	Reflux finding score

References

1. Nelson SP, Kothari S, Wu EQ, Beaulieu N, McHale JM, Dabbous OH. Pediatric gastroesophageal reflux disease and acid-related conditions: trends in incidence of diagnosis and acid suppression therapy. J Med Econ 2009;12:348–55. [PubMed: 19827992]

2. Chang BA, MacNeil SD, Morrison MD, Lee PK. The reliability of the reflux finding score among general otolaryngologists. J Voice 2015;29:572–7. [PubMed: 26118936]

3. Jette ME, Gaumnitz EA, Birchall MA, Welham NV, Thibeault SL. Correlation between reflux and multichannel intraluminal impedance pH monitoring in untreated volunteers. Laryngoscope 2014;124:2345–51. [PubMed: 24782404]

4. Writing Committee for the American Lung Association Asthma Clinical Research Centers, Holbrook JT, Wise RA, Gold BD, Blake K, Brown ED, et al. Lansoprazole for children with poorly controlled asthma: a randomized controlled trial. JAMA 2012;307:373–81. [PubMed: 22274684]

5. Canani RB, Cirillo P, Roggero P, Romano C, Malamisura B, Terrin G, et al. Therapy with gastric acidity inhibitors increases the risk of acute gastroenteritis and community-acquired pneumonia in children. Pediatrics 2006;117:e817–20. [PubMed: 16651285]

6. Belafsky PC, Postma GN, Koufman JA. The validity and reliability of the reflux finding score (RFS). Laryngoscope 2001;111:1313–7. [PubMed: 11568561]

7. Varni JW, Burwinkle TM, Seid M, Skarr D. The PedsQL 4.0 as a pediatric population health measure: feasibility, reliability, and validity. Ambul Pediatr 2003;3:329–41. [PubMed: 14616041]

8. Varni JW, Bendo CB, Nurko S, Shulman RJ, Self MM, Franciosi JP, et al. Health-related quality of life in pediatric patients with functional and organic gastrointestinal diseases. J Pediatr 2015;166:85–90. [PubMed: 25241177]

9. Shay S, Tutuian R, Sifrim D, Vela M, Wise J, Balaji N, et al. Twenty-four hour ambulatory simultaneous impedance and pH monitoring: a multicenter report of normal values from 60 healthy volunteers. Am J Gastroenterol 2004;99:1037–43. [PubMed: 15180722]

10. Vandenplas Y, Rudolph CD, Di Lorenzo C, Hasall E, Liptak G, Mazur L, et al. Pediatric gastroesophageal reflux clinical practice guidelines: joint recommendations of the North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition (NASPGHAN) and the European Society for Pediatric Gastroenterology, Hepatology, and Nutrition (ESPGHAN). J Pediatr Gastroenterol Nutr 2009;49:498–547. [PubMed: 19745761]

11. Slaughter JL, Stenger MR, Reagan PB, Jadcherla SR. Neonatal histamine-2 receptor antagonist and proton pump inhibitor treatment at United States children's hospitals. J Pediatr 2016;174:63–70, e3. [PubMed: 27131401]

12. Chai G, Governale L, McMahon AW, Trinidad JP, Staffa J, Murphy D. Trends of outpatient prescription drug utilization in US children, 2002–2010. Pediatrics 2012;130:23–31. [PubMed: 22711728]

13. Francis DO, Rymer JA, Slaughter JC, Choksi Y, Jiramongkolchai P, Oqbeide E, et al. High economic burden of caring for patients with suspected extraesophageal reflux. Am J Gastroenterol 2013;108:905–11. [PubMed: 23545710]

14. Guo H, Ma H, Wang J. Proton pump inhibitor therapy for the treatment of laryngopharyngeal reflux: a meta-analysis of randomized controlled trials. J Clin Gastroenterol 2015;50:295–300.

15. Carr MM, Nagy ML, Pizzuto MP, Poje CP, Brodsky LS. Correlation of findings at direct laryngoscopy and bronchoscopy with gastroesophageal reflux disease in children: a prospective study. Arch Otolaryngol Head Neck Surg 2001;127:369–74. [PubMed: 11296043]

16. Simons JP, Rosen CA, Casselbrant ML, Chi DH, Schaitkin BM, Rubinstein EN, et al. Comparison of pediatric voice outcome survey, reflux symptom index, reflux finding score, and esophageal biopsy results. Arch Otolaryngol Head Neck Surg 2008;134:837–41. [PubMed: 18711057]
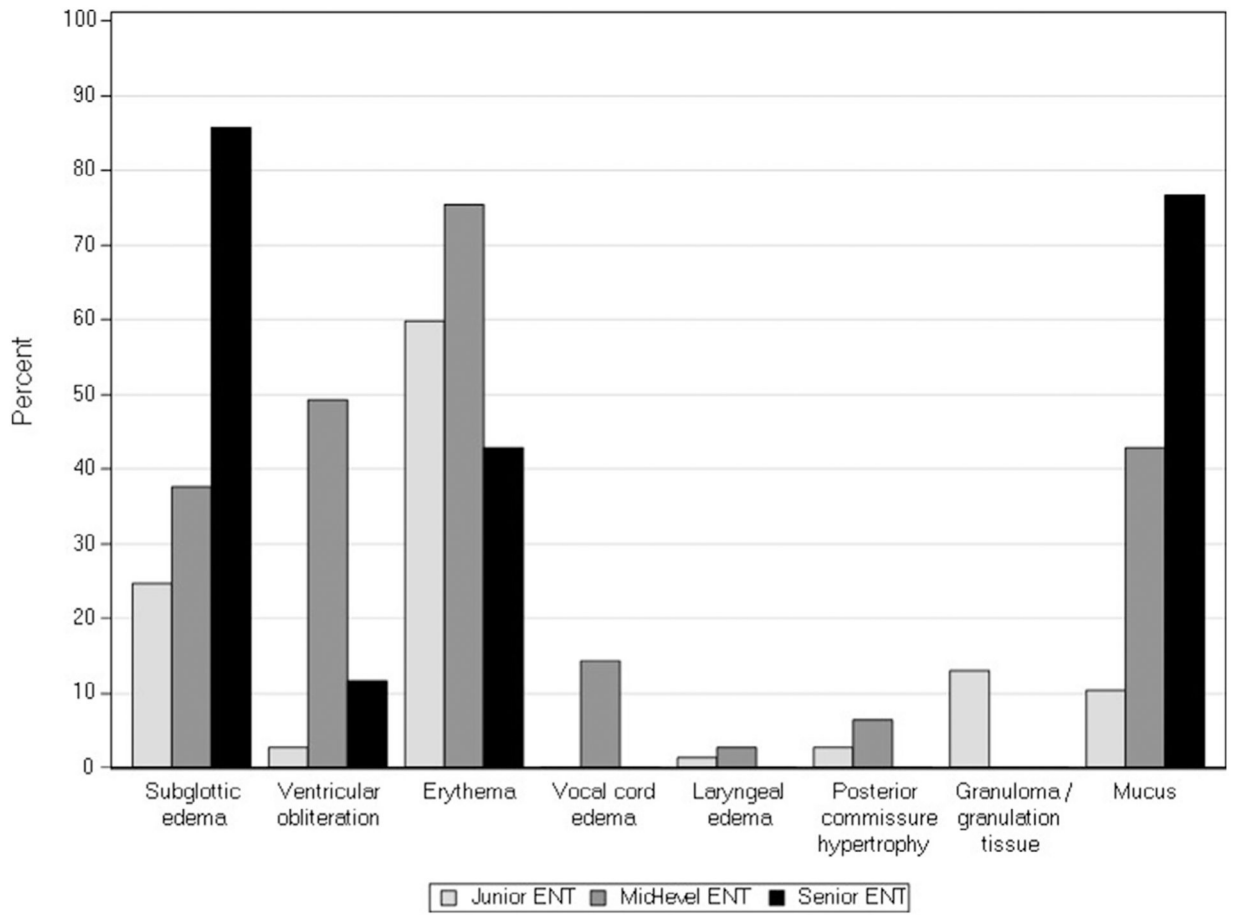
Figure 1.
Percentages of airways scored at the highest severity levels for each of the RFS subgroups per otolaryngologist (ENT).

Junior: r = 0.01
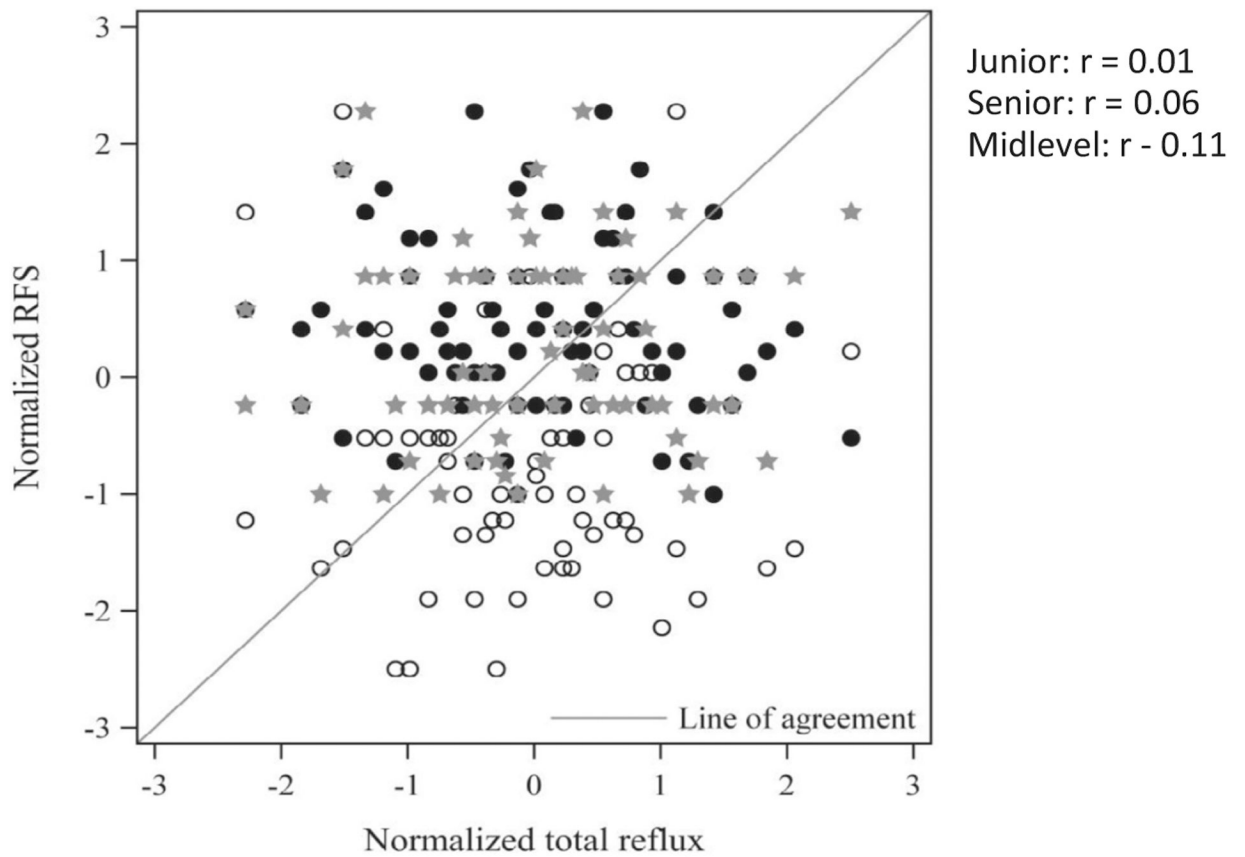Senior: r = 0.06
Midlevel: r - 0.11

Figure 2.
Lack of correlation between RFS and total number of reflux events (*open circle*: junior otolaryngologist; *solid circle*: midlevel otolaryngologist; *star*: senior otolaryngologist).
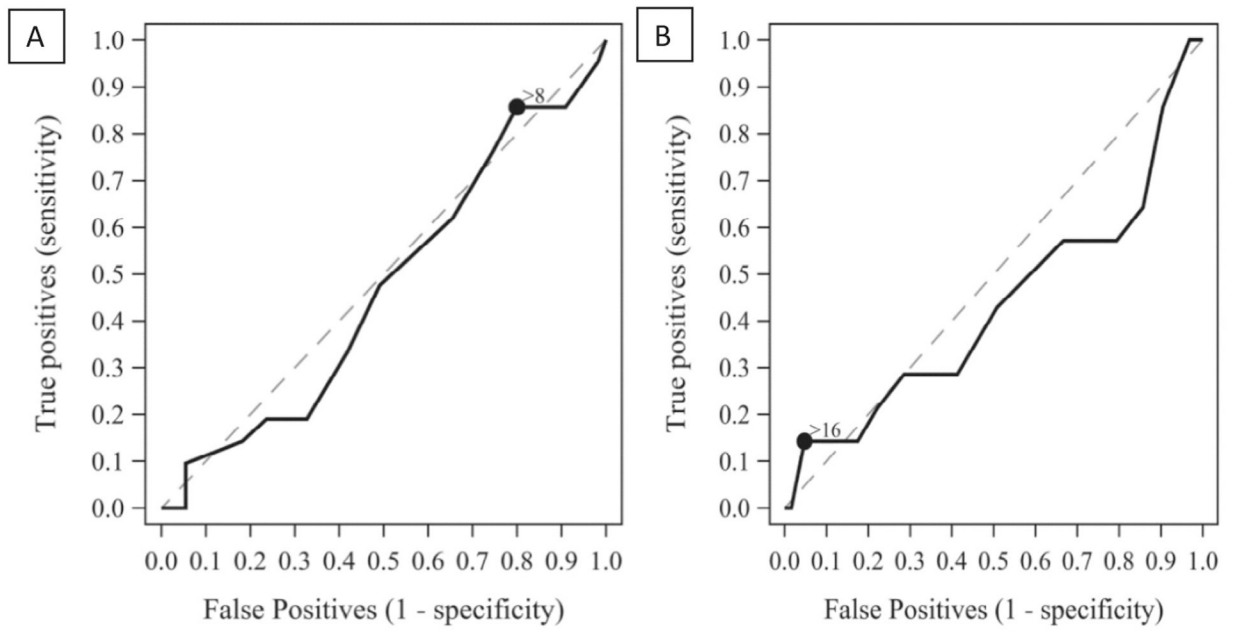
Figure 3.
Receiver operating characteristic (ROC) curves for RFS compared with the **A,** gold standard pH probe or **B,** MII.

Table.

The relationship between mean RFS by ENT and abnormal pH-MII testing

	Junior ENT Mean RFS (SD)	Midlevel ENT Mean RFS (SD)	Senior ENT Mean RFS (SD)
Abnormal impedance	9.2 (6.3)	12.3(3.0)	12.4(3.6)
Normal impedance	8.4 (4.6)	14.4(3.2)	13.2(3.7)
Abnormal pH probe	8.5 (5.1)	13.9(2.5)	12.8(3.3)
Normal pH probe	8.6 (4.8)	14.1 (3.5)	13.1 (3.8)

ENT, otolaryngologist.

Shown are mean (SD) for each ENT by abnormal and normal pH-MII testing.