Correlation of Reflux Finding Scores on Laryngoscopy to Clinical Outcomes

Samuel R. Patton, MD, Naren V. Venkatesan, MD, Michael P. Underbrink, MD, PhD, MBA
The University of Texas Medical Branch at Galveston

ABSTRACT

BACKGROUND

The Reflux Finding Score (RFS) is a numerical scoring system created by Belafsky et al. to standardize evaluation of patients with laryngopharyngeal reflux disease (LPRD). A score of 8 or greater is considered positive. To date, it has not been validated on a large scale. The study aims to evaluate the validity of the RFS as a clinical tool to diagnose LPRD by testing its reproducibility and accuracy.

OBJECTIVE

Methods

Three scorers blinded independently and blindly reviewed 300 videostroboscopes of patients that were performed between 2005-2009. Scores were given independently. A chart review was then performed to determine the diagnosis given at the time of the videostroboscopy, whether the patient was treated with anti-reflux medication, and any abnormalities noted on recent esophagogastroduodenoscopy. The reflux finding scores of the three scorers were statistically evaluated for agreement with each other and to the findings of the chart review.

RESULTS

Of the 330 videostroboscopy performed, 168 were included in the final analysis. The inter-rater agreement across all three scorers had a kappa of 0.24 and a weighted kappa of 0.51 indicating fair agreement. Pairwise agreement was consistent amongst all three pairs. Scores #1 and #2 had the highest agreement with a weighted kappa of 0.31. Scores #1 and #2 had a similar agreement with a weighted kappa of 0.30. Scores #1 and #3 had a weighted kappa of 0.26 (Table 1). All pairwise weighted scores represented fair agreement ratings. The c-statistic scores for correlation of the scorers’ RFS’s to the chart diagnosis of LPRD were 0.554, 0.609, and 0.575 for scores #1, #2, and #3 respectively (Table 2). A stratified analysis based on age, gender, and smoking status yielded c-statistic scores ranging from 0.560 to 0.665 (Table 2) which failed to reach significance. Results of ESCD could not be included in the calculation due to insufficient patient data.

CONCLUSIONS

Our study sought to evaluate the validity of the RFS in two ways: inter-rater agreement and accuracy of the RFS by comparing it to the chart diagnosis given at the time the videostroboscopy was performed. The overall kappa score of 0.24 represents only fair agreement amongst all scorers. The pairwise weighted kappa scores also yielded fair values. The c-statistic scores obtained for a chart diagnosis of LPRD were below 0.610 for all three scorers, indicating fair-to-poor correlation. Our study had several limitations. Scorers had varying experience levels and included a PGY-3 and a PGY-5 resident. Our ability to correlate RFS in isolation to chart diagnosis of LPRD retrospectively in this study was limited. It should be noted that there was no gold standard for comparison in. There were no identifiable objective criteria for chart diagnosis of LPRD. In addition, the videostroboscopy were performed at an academic medical center with residents of varying training levels assisting in the procedures and documentation. A score of 8 or greater was used as a positive score. Further analysis may be useful to identify if achieving the threshold for a positive score would improve the correlation to the chart diagnosis.

REFERENCES