Development of the Reflux Finding Score for Infants and Its Observer Agreement

Rachel J. van der Pol, MD[1], Maartje M. J. Singendonk, MsC[1], Astrid M. König, MD, PhD[2], Hans Hoeve, MD, PhD[3], Quinten Kammeijer, MD[2], Bas Pullens, MD[3], Erik van Spronsen, MD, PhD[2], George Thomas, MD, PhD[2], Lenka Vermeeren, MD, PhD[2], Marc A. Benninga, MD, PhD[1], and Michiel P. van Wijk, MD, PhD[1]

Objective It is hypothesized that laryngeal edema is caused by laryngopharyngeal reflux (LPR) (ie, gastroesophageal reflux extending into the larynx and pharynx). The validated reflux finding score (RFS) assesses LPR disease in adults. We, therefore, aimed to develop an adapted RFS for infants (RFS-I) and assess its observer agreement.

Study design Visibility of laryngeal anatomic landmarks was assessed by determining observer agreement. The RFS-I was developed based on the RFS, the found observer agreement, and expert opinion. An educational tutorial was developed which was presented to 3 pediatric otorhinolaryngologists, 2 otorhinolaryngologists, and 2 gastroenterology fellows. They then scored videos of flexible laryngoscopy procedures of infants who were either diagnosed with or specifically without laryngeal edema.

Results In total, 52 infants were included with a median age of 19.5 (0-70) weeks, with 12 and 40 infants, respectively, for the assessment of the laryngeal anatomic landmarks and the assessment of the RFS-I. Overall interobserver agreement of the RFS-I was moderate (intraclass correlation coefficient = 0.45). Intraobserver agreement ranged from moderate to excellent agreement (intraclass correlation coefficient = 0.50-0.87).

Conclusion A standardized scoring instrument was developed for the diagnosis of LPR disease using flexible laryngoscopy. Using this tool, only moderate interobserver agreement was reached with a highly variable intraobserver agreement. Because a valid scoring system for flexible laryngoscopy is lacking up until now, the RFS-I and flexible laryngoscopy should not be used solely to clinically assess LPR related findings of the larynx, nor to guide treatment. *(J Pediatr 2014;165:479-84)*.

Laryngomalacia is the most common congenital anomaly of the larynx accounting for over 60% of all noninfectious stridors in infancy.[1] In infants with laryngomalacia, laryngeal edema is frequently seen during laryngoscopic examination. It is commonly thought to be the result of gastroesophageal reflux (GER) extending into the larynx and pharynx: laryngopharyngeal reflux (LPR).[2-6] However, evidence for this causality is lacking. It is hypothesized that a partial obstruction of the airway because of laryngomalacia causes a negative intrathoracic pressure, facilitating LPR to occur, which in turn causes laryngeal edema. This subsequently leads to a vicious circle of increased obstruction, more LPR, and edema.[7]

In adults, LPR can reliably be detected using flexible laryngoscopy and a validated reflux finding score (RFS).[8] In children, evaluation of the larynx with flexible laryngoscopy might be hampered by the smaller anatomic landmarks and smaller endoscopes with lower image resolution. Furthermore, reflux patterns in infants differ significantly from those in adults as a result of different feeding patterns and posture. In infants, common endoscopic findings suggested to be LPR-related are edema and erythema of the arytenoids, postglottic and vocal fold edema, and erythema.[9] Despite the abovementioned differences between adults and infants, no specific scoring instrument exists for this age group.

Based on medical history, physical examination, and laryngoscopic findings, proton pump inhibitor (PPI) therapy is commonly initiated when LPR-related laryngeal edema is suspected to be the cause of symptoms.[10,11] Contrary to adults, however, PPI therapy has been proven ineffective for classical GER symptoms in infants such as regurgitation and excessive crying and are, thus, not approved by the Federal Drug Administration.[12] Furthermore, no evidence is available for their use in LPR-related symptoms in infants.

The aim of this study was to develop a scoring instrument to evaluate signs of LPR seen on flexible laryngoscopy in infants, the RFS for infants (RFS-I) and to assess its inter- and intraobserver agreement.

GER	Gastroesophageal reflux
ICC	Intraclass correlation coefficient
LPR	Laryngopharyngeal reflux
PPI	Proton pump inhibitor
RFS	Reflux finding score
RFS-I	RFS for infants

From the [1]Department of Pediatric Gastroenterology and Nutrition, Emma Children's Hospital, [2]Department of Otorhinolaryngology, Academic Medical Center, Amsterdam, The Netherlands; and [3]Department of Otorhinolaryngology-Head and Neck Surgery, Erasmus Medical Center, Rotterdam, The Netherlands

The authors declare no conflicts of interest.

http://dx.doi.org/10.1016/j.jpeds.2014.05.022

Methods

For the development of the RFS-I and the observer agreement of the RFS-I, we randomly selected 52 infants under the age of 18 months from a database of laryngoscopic procedures performed in the otorhinolaryngology department of the Academic Medical Center Amsterdam between 2006 and 2010. For the development of the established RFS-I, 12 infants without evidence of LPR-related laryngeal findings during clinical flexible laryngoscopy were selected. To evaluate inter- and intraobserver agreement of the RFS-I, 20 additional infants with reported flexible laryngoscopy evidence of LPR and 20 additional infants without any reported pathologic findings were selected. Patients were excluded if 1 of the following conditions were present: history of malformation of the esophagus and/or history of surgery of the gastrointestinal or pulmonary tract. The study was approved by the medical ethics committee of the Academic Medical Center, Amsterdam.

Development of the RFS-I

First, we selected anatomic landmarks (Figure 1) that could potentially be included in the RFS-I. In patients without any reported abnormalities during the original flexible laryngoscopy, we evaluated the viability of the laryngoscopic detection of these landmarks by assessing inter- and intraobserver agreement. Two experienced pediatric otorhinolaryngologists and 1 otorhinolaryngology resident reviewed the anonymized video clips. Reviewers performed their analysis blinded from symptom presentation and independently from each other. Both visibility and aspect of the selected landmarks were assessed. To determine intraobserver agreement, all reviewers performed a second evaluation of the video clips with a minimum interval of at least 2 days after the first analysis. For intraobserver agreement, the mean of the intraobserver scores of the 3 observers was used. The visibility of a specific landmark was arbitrarily scored as 'good' if at least 1 of the inter- or intraobserver scores, ranged between 0.61 and 1.00, and the other accompanying inter- or intraobserver score was at least 0.35. Landmarks were considered to be 'fair to moderate' when both inter- or intraobserver score ranged between 0.35 and 0.60. Values lower than those mentioned above were considered invalid. Second, based on expert opinion on pediatric laryngoscopy and based on items that were included in the original RFS, items for the RFS-I were selected.

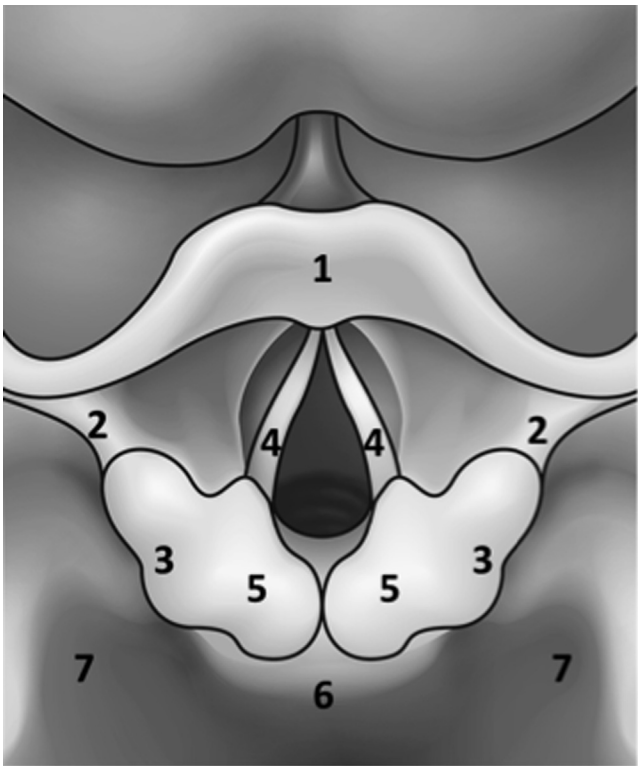

Figure 1. Anatomic landmarks (1 = epiglottis, 2 = aryepiglottic fold, 3 = cuneiform cartilages, 4 = vocal cords, 5 = arytenoids, 6 = postcricoid region, 7 = piriform sinus).

Inter- and Intraobserver Agreement of the RFS-I

The 40 selected laryngoscopy video clips were scored by 3 pediatric otorhinolaryngologists, 2 general otorhinolaryngologists, and 2 gastroenterology fellows (group 1, 2, and 3, respectively) from 2 different centers. The participants were first presented an educational tutorial that explained the scoring items of the RFS-I and showed static images of all different scoring options per scoring item. Next, reviewers were presented the video clips in a randomized order. All reviewers evaluated these video clips independently, blinded for the patient's clinical profile and findings during the clinical flexible laryngoscopy. In order to evaluate intraobserver agreement, all reviewers were asked to perform a second rating of the laryngoscopic video clips at least 2 days after their first assessment. Finally, observers where asked for comments on video quality and any other comments they might have after completing their review.

Statistical Analyses

Data were analyzed using IBM SPSS Statistics 19 (SPSS Inc, Armonk, New York). For the analysis of the development of the RFS-I, inter- and intraobserver agreement was determined for all selected landmarks using Fleiss kappa (kappa further annotated as κ). For the RFS-I, inter- and intraobserver agreement was determined per scoring item of the RFS-I and for all scoring items combined. For categorical data, inter- and intraobserver agreement was calculated using Cohen κ (2 observers) and Fleiss κ (>2 observers). For ordinal data, weighted κ and the intraclass correlation coefficient (ICC) were used. Fleiss κ was calculated by using a premade syntax for SPSS (available from corresponding author). For the assessment of the inter- and intraobserver agreement of the RFS-I, we applied the arbitrary but common scale for κ and ICC values: 0.00 = no agreement, 0.01-0.20 = slight agreement, 0.21-0.40 = fair agreement, 0.41-0.60 = moderate