International Journal of Pediatric Otorhinolaryngology 86 (2016) 37–42

Contents lists available at ScienceDirect

International Journal of Pediatric Otorhinolaryngology

journal homepage: www.elsevier.com/locate/ijporl

Reliability of the reflux finding score for infants in flexible versus rigid laryngoscopy

Maartje M.J. Singendonk [a,*], Bas Pullens [b], Jan A.A. van Heteren [a], Henriëtte H.W. de Gier [b], Hans L.J. Hoeve [b], Astrid M. König [c], Marc P. van der Schroeff [b], Carlijn E.L. Hoekstra [c], Laura L. Veder [b], Rachel J. van der Pol [a], Marc A. Benninga [a], Michiel P. van Wijk [a]

[a] Department of Pediatric Gastroenterology and Nutrition, Emma Children's Hospital, Academic Medical Center, Amsterdam, The Netherlands
[b] Department of Otorhinolaryngology and Head and Neck Surgery, Sophia Children's Hospital, Erasmus Medical Center, Rotterdam, The Netherlands
[c] Department of Otorhinolaryngology, Academic Medical Center, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 11 January 2016
Received in revised form 11 April 2016
Accepted 12 April 2016
Available online 14 April 2016

Keywords:
LPR
Laryngopharyngeal reflux
Diagnostic

ABSTRACT

Objectives: The Reflux Finding Score for Infants (RFS-I) was developed to assess signs of laryngopharyngeal reflux (LPR) in infants. With flexible laryngoscopy, moderate inter- and highly variable intraobserver reliability was found. We hypothesized that the use of rigid laryngoscopy would increase reliability and therefore evaluated the reliability of the RFS-I for flexible versus rigid laryngoscopy in infants.
Methods: We established a set of videos of consecutively performed flexible and rigid laryngoscopies in infants. The RFS-I was scored twice by 4 otorhinolaryngologists, 2 otorhinolaryngology fellows, and 2 inexperienced observers. Cohen's and Fleiss' kappas (k) were calculated for categorical data and the intraclass correlation coefficient (ICC) was calculated for ordinal data.
Results: The study set consisted of laryngoscopic videos of 30 infants (median age 7.5 (0–19.8) months). Overall interobserver reliability of the RFS-I was *moderate* for both flexible (ICC = 0.60, 95% CI 0.44–0.76) and rigid (ICC = 0.42, 95% CI 0.26–0.62) laryngoscopy. There were no significant differences in reliability of overall RFS-I scores and individual RFS-I items for flexible versus rigid laryngoscopy. Intraobserver reliability of the total RFS-I score ranged from *fair* to *excellent* for both flexible (ICC = 0.33–0.93) and rigid (ICC = 0.39–0.86) laryngoscopies. Comparing RFS-I results for flexible versus rigid laryngoscopy *per* observer, reliability ranged from *no* to *substantial* (k = –0.16–0.63, mean k = 0.22), with an observed agreement of 0.08–0.35.
Conclusion: Reliability of the RFS-I was moderate and did not differ between flexible and rigid laryngoscopies. The RFS-I is not suitable to detect signs or to guide treatment of LPR in infants, neither with flexible nor with rigid laryngoscopy.

1. Introduction

Laryngopharyngeal reflux (LPR) disease is a common problem in infancy and refers to the backflow of gastric contents into the larynx, oropharynx and/or nasopharynx [1]. LPR disease can be regarded as an extension of gastro-esophageal reflux (GER) disease; however, others postulate it as a unique disease identity due to the differences in symptoms, diagnostic findings and treatment efficacy [2,3]. Symptom generation in LPR disease is hypothesized to directly result from injury to the mucosa of the larynx and pharynx by acidity of the refluxate and exposure to reflux components such as pepsin, bile salts, and pancreatic enzymes causing tissue edema, inflammation and decreased laryngopharyngeal sensitivity [4–7]. Infants and children with LPR disease may present with a wide variety of symptoms depending on their age, including regurgitation, vomiting, growth failure, upper-airway problems, and apnea [1,8]. Additionally, relationships between LPR disease and laryngomalacia, subglottic stenosis and chronic respiratory diseases among other pathologies have been postulated [9–11]. As the typical pattern of symptoms is chronic-intermittent, reliable diagnosis of LPR disease remains challenging among pediatricians and otorhinolaryngologists [1].

Current diagnostic tools for infant LPR disease include the quantification of reflux through combined pH and impedance-monitoring (pH–MII), laryngoscopic examination for signs of inflammation of

Abbreviations: ALTE, apparent life threatening event; CI, confidence interval; GER, gastro-esophageal reflux; ICC, intraclass correlation coefficient; LPR, laryngopharyngeal reflux; NS, not significant; pH-MII, pH-impedance monitoring; PPI, proton pump inhibitor; RFS, reflux finding score; RFS-I, reflux finding score for infants

* Corresponding author at: Department of Pediatric Gastroenterology and Nutrition, Emma Children's Hospital, AMC, Room C2-312, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands. Tel.: +31 20 5662906; fax: +31 205669478.
E-mail address: m.m.j.singendonk@amc.uva.nl (M.M.J. Singendonk).

http://dx.doi.org/10.1016/j.ijporl.2016.04.017
0165-5876/

What is already known on this topic

- Laryngoscopy is currently used for the evaluation of signs of laryngopharyngeal reflux (LPR) in infants, however convincing data and established normative values are lacking.
- The Reflux Finding Score for Infants (RFS-I) was developed to assess signs of infant LPR, but its reliability with flexible laryngoscopy is moderate and highly variable.
- Rigid laryngoscopy has the advantage of higher optical resolution compared to flexible laryngoscopy and is performed in an anesthetized patient, potentially allowing better visualization.

What this study adds

- Reliability of the RFS-I was compared for flexible versus rigid laryngoscopy in 30 infants among eight observers.
- Reliability of the RFS-I was moderate and did not differ between flexible and rigid laryngoscopies.
- The RFS-I is not suitable to detect signs of LPR in infants or to guide treatment, neither with flexible, nor with rigid laryngoscopy.

the larynx, questionnaires for symptom assessment, and empirical proton pump inhibitor (PPI) therapy. However, due to the lack of convincing data and established normative values, no gold standard exists for the detection of LPR disease in infants [1]. In an earlier study we proposed a reflux finding score for infants (RFS-I), based upon a validated combination of the adult RFS, expert opinion and assessment of visibility of anatomic landmarks in the infant larynx [12,13]. However, when applied on videos of flexible laryngoscopies of 52 infants under the age of 18 months, only moderate interobserver and highly variable intraobserver reliabilities were found, independent of the level of observers' experience. Most observers considered the reported *moderate* to *poor* quality of the videos to be inherent to the use of the small flexible laryngoscope used in children [13].

The aim of this study was therefore to assess the reliability of the RFS-I in the detection of LPR-related findings on rigid versus flexible laryngoscopy performed under general anesthesia in infants referred for laryngoscopic examination.

2. Methods

To establish inter- and intraobserver reliability, we made use of a set laryngoscopic videos of infants referred for examination between February 2011 and May 2015 to the otorhinolaryngology department of the Sophia Children's Hospital, Rotterdam, The Netherlands. Rigid laryngoscopic evaluation was preceded by flexible evaluation and performed according to a fixed protocol, allowing sufficient time to view predetermined aspects of the larynx during both evaluations. Patients were placed in a supine position. All laryngoscopy procedures were done under general anesthesia; the flexible laryngoscopies were performed in the spontaneously breathing child and rigid laryngoscopy was subsequently performed after addition of a muscle relaxant with ventilation over the rigid endoscope. Sizes of the flexible and rigid scopes ranged from 2.2 to 3.8 and from 2.5 to 3.5 mm, respectively, depending on the patient's size. This study was exempted from full review by the Medical Ethics Committees of the Academic Medical Center, Amsterdam and the Erasmus Medical Center, Rotterdam as it comprised retrospective review of clinical data of patients undergoing standard clinical practice.

Table 1
Scoring items of the RFS-I.

Findings	Score
Erythema/hyperemia/edema	0 = absent 2 = postcricoid region 4 = diffuse
Vocal cord visibility	0 = visible 2 = partly/not visible
Endolaryngeal mucus	0 = absent 2 = present

2.1. The RFS-I

The RFS-I is a 3-item clinical severity scale for the evaluation of LPR-related findings in infants during laryngoscopy, ranging from 0 (no abnormal findings) to 8 (worst score possible, Table 1) [13].

2.2. Inter- and intraobserver reliability of the RFS-I

Two series of videos of flexible and rigid laryngoscopies were scored by 4 experienced otorhinolaryngologists (group 1; performing >40 laryngoscopies annually), 2 otorhinolaryngology fellows (group 2; in training), and 2 observers naïve to laryngoscopic examination with no clinical experience in the field of pediatric otorhinolaryngology (group 3) from two different centers (Academic Medical Center, Amsterdam, and Erasmus Medical Center, Rotterdam, The Netherlands). None of the selected observers were responsible for the initial laryngoscopic procedure or clinical evaluation. Four of eight observers were familiar with the RFS-I by participating in our previous study (BP, AK, HH and RvdP). All observers received an instruction sheet prior to evaluation. All reviewers evaluated these videos independently and blinded to the patient's clinical profile and findings during the clinical laryngoscopy. In order to evaluate intraobserver reliability, observers were asked to perform a second rating of the two series >7 days after their first assessment. Videos were presented in a randomized order, different between both rating-series and between observers. Observers were asked to assess the quality of the videos (good, moderate and bad quality) and to add any additional comments.

2.3. Statistical analysis

Data were analyzed using IBM SPSS Statistics 21 (SPSS Inc, Armonk, New York). Descriptive statistics were calculated on variables related to subject characteristics. Observed and inter- and intraobserver reliabilities were determined per scoring item of the RFS-I and for all scoring items combined. Patient studies were excluded pairwise from analysis when scoring items were not uniformly obtained by all observers. For categorical data, reliability was calculated using Cohen's and Fleiss' kappas (2 and >2 observers, respectively). For ordinal data the intraclass correlation coefficient (ICC, two way mixed with absolute agreement) was used. Fleiss kappa was calculated by using a pre-made syntax for SPSS (available from corresponding author). For the assessment of inter- and intraobserver reliabilities, we applied the arbitrary but common scale for kappa and ICC values: ≤0.00 = no agreement, 0.01–0.20 = slight agreement, 0.21–0.40 = fair agreement, 0.41–0.60 = moderate agreement, 0.61–0.80 = substantial agreement, 0.81–0.99 = excellent agreement, and 1.00 = perfect agreement. We considered a kappa between 0.6 and 1.0 to be valid for use in research settings and/or clinical practice. Kappa values were considered significantly different in case of non-overlapping confidence intervals [14].