Article

Validity and Reliability of the Reflux Sign Assessment

Annals of Otology, Rhinology & Laryngology
1–13

DOI: 10.1177/0003489419888947
journals.sagepub.com/home/aor
SAGE

Jérôme R. Lechien, MD, PhD, MS[1,2,3,4], Alexandra Rodriguez Ruiz, MD[1,3], Didier Dequanter, MD, PhD[1,3], Francois Bobin, MD[1,5], Francois Mouawad, MD, PhD, MSc[6], Vinciane Muls, MD[1,7], Kathy Huet, PhD[1,4], Bernard Harmegnies, PhD[1,4], Sarah Remacle, MD[8], Camille Finck, MD, PhD[1,8*], and Sven Saussez, MD, PhD[1,2,3*]

Abstract

Objective: To develop and validate the Reflux Sign Assessment (RSA), a clinical instrument evaluating the physical findings of laryngopharyngeal reflux (LPR).

Methods: A total of 106 patients completed a 3-month treatment based on the association of diet, pantoprazole, alginate, or magaldrate with the LPR characteristics (acid, nonacid, mixed). Forty-two asymptomatic individuals completed the study (control group). The RSA results and reflux finding score (RFS) were documented for the LPR patients at baseline and after treatment. Intrarater reliability was assessed through a test-retest blinded evaluation of signs (7-day intervals). Interrater reliability was assessed by comparing the RSA evaluations of three blinded otolaryngologists through Kendall's W. Responsiveness to change was evaluated through a comparison of the baseline and 3-month posttreatment findings. The RSA cutoffs for determining the presence and absence of LPR were examined by receiver operating characteristic (ROC) analysis.

Results: A total of 102 LPR patients completed the study (68 females). The mean age was 53 years. The mean RSA at baseline was 25.95 ± 9.58; it significantly improved to 18.96 ± 7.58 after 3 months of therapy ($P < .001$). RSA exhibited good intra- ($r = 0.813$) and interrater (Kendall's W = 0.663) reliabilities (N = 56). There was no significant association between the RSA, gastrointestinal endoscopy findings, and the types of reflux (acid, nonacid, or mixed) according to impedance-pH monitoring. An RSA >14 may be suggestive of LPR.

Conclusion: The RSA is a complete clinical instrument evaluating both laryngeal and extralaryngeal findings associated with LPR. The RSA demonstrated high intra- and interrater reliabilities and responsiveness to change.

Keywords

laryngopharyngeal, reflux, laryngitis, tool, outcome, finding, sign

Introduction

Laryngopharyngeal reflux (LPR) may be defined as an inflammatory condition of the upper aerodigestive tract tissues related to the direct and indirect effect of gastric or duodenal content reflux, inducing morphological changes in the upper aerodigestive tract.[1] LPR symptoms or findings are found in 10 to 15% of outpatients visiting otolaryngology-head & neck surgery departments,[2] and up to 50% of patients in laryngology specific practices.[3]

The majority of symptoms and signs are nonspecific, yielding the diagnostic difficult.[4] In 2001, Belafsky et al developed reflux symptom index (RSI) and reflux finding score (RFS), which are two clinical instruments that characterize the severity of symptoms and findings of LPR patients.[5,6] These two instruments are well-used around the world but are incomplete, omitting many usual symptoms (ie, throat pain, halitosis, odynophagia, regurgitations) and findings (ie, pillar anterior erythema, pharyngeal inflammation, coated tongue, tongue tonsil hypertrophy, pharyngeal sticky mucus).[4,7-9] Nowadays, there is no clinical instrument assessing both laryngeal and extralaryngeal findings associated with LPR, which is validated on a large number of patients with a clear diagnostic of LPR regarding multichannel intraluminal impedance-pH monitoring (MII-pH).[8] In the context of the nonspecificity of reflux-associated signs, the lack of consideration of some signs may undoubtedly alter the clinical evaluation of LPR patients at baseline and throughout treatment, yielding the assessment of treatment efficacy inaccurate.

The members of the LPR Study Group of Young Otolaryngologists of the International Federation of Oto-Rhino-Laryngological Societies (YO-IFOS)[10] have developed Reflux Sign Assessment (RSA) for the evaluation of laryngeal and extralaryngeal findings of LPR patients. The aim of this study is to evaluate the reliability and the validity of the RSA.

Materials and methods

The local ethics committee approved the study protocol (n°BE076201837630).

Development of RSA

The development of the RSA started after the World Ear, Nose, and Throat (ENT) Congress of IFOS (Paris, 2017), in which international experts decided to develop a new instrument evaluating the clinical findings of LPR. The RSA was developed for the diagnosis and the follow-up of suspected LPR patients or those with a diagnosis confirmed by pH studies.

The RSA's content (signs, scoring system, and presentation) was based on expert opinions[11] and two systematic reviews describing findings attributed to LPR in the current literature.[1,8] Because some previous clinical instruments reported low interrater reliabilities,[8] great importance was attached to a scoring system for items with a low possibility of interrater disparity. In that way, a scoring system that classified signs as "mild," "moderate," or "severe" was avoided in favor of scoring systems that are as descriptive as possible. The weight assignment of each finding was based on the prevalence of the related finding in an initial cohort of 101 LPR patients; the methodology being described in Appendix 1. The RSA is presented in Figure 1. Examples of positive clinical findings used in the RSA are illustrated in Figures 2 and 3. The score is subdivided into three parts based on the localization of the signs (oral cavity, pharynx, and larynx). The total score is the sum of each item score. The maximal total score is 72.

To rigorously study the validity and reliability of the RSA, the current study was conducted according to a checklist of recommendations designed to obtain valid and reliable clinical instruments (Table 1).[8]

Subjects and Setting

A total of 106 adult patients with LPR were recruited from January 2017 to March 2019 from the ENT Departments of CHU Saint-Pierre (Brussels, Belgium) and the Elsan Polyclinique of Poitiers (Poitiers, France). The LPR diagnosis was made with a 24-hour MII-pH. As recommended,[12] gastrointestinal (GI) endoscopy was performed in patients with gastroesophageal reflux disease (GERD) symptoms based on the Reflux Symptom Score (RSS).[13] In our previous work, we have reported a reduction of GERD symptom perception in older people with LPR,[14] GI endoscopy was systematically proposed for patients $\geq$60 years old. The exclusion criteria were similar to those described in a previous publication[13] and included smoker, alcohol dependence, upper respiratory tract infection within the last month, current use of anti-reflux treatment (ie, PPIs, antihistamine, alginate, and magaldrate), pregnancy, neurological or psychiatric illness, previous history of neck surgery or trauma, malignancy, history of head and neck radiotherapy, and active seasonal allergies or asthma.

To determine normative data of the RSA, 42 sex- and age-matched asymptomatic individuals (24 females; 19-73 years of age) without any evidence of LPR were included. They fulfilled the requirement of an RSS $<$13 and underwent videolaryngoscopy to assess the RSA and RFS. They were excluded if one or more previously described exclusion criteria were met.

Multichannel Intraluminal Impedance-pH Monitoring

The MII-pH protocol has been described in previous publications.[13,15] A senior gastroenterologist or a senior otolaryngologist analyzed the MII tracings using a standardized method.[16]

[1]Research Committee of the Young-Otolaryngologists of the International Federations of Oto-rhino-laryngological Societies (YO-IFOS)
[2]Laboratory of Anatomy and Cell Biology, Faculty of Medicine, UMONS Research Institute for Health Sciences and Technology, University of Mons (UMons), Mons, Belgium
[3]Department of Otorhinolaryngology and Head and Neck Surgery, CHU Saint-Pierre, Brussels, Belgium
[4]Laboratory of Phonetics, Faculty of Psychology, Research Institute for Language sciences and Technology, University of Mons (UMons), Mons, Belgium
[5]Department of Otorhinolaryngology and Head and Neck Surgery, Elsan Polyclinique de Poitiers, France
[6]Department of Otorhinolaryngology and Head and Neck Surgery, CHU de Lille, Lille, France
[7]Department of Gastroenterology and Endoscopy, CHU Saint-Pierre, Université Libre de Bruxelles, Brussels, Belgium
[8]Department of Otorhinolaryngology and Head and Neck Surgery, CHU de Liège (Sart Tilman), Liège, Belgium

*Pr Finck and Pr Saussez have equally contributed to this work and should be regarded as *joint last authors.*

Corresponding Author:
Jérôme R. Lechien, MD, PhD, MS, Laboratory of Anatomy and Cell Biology, Faculty of Medicine, UMONS Research Institute for Health Sciences and Technology, University of Mons (UMONS), Avenue du Champ de mars, 6, Mons, B7000, Belgium.
Email: jerome.lechien@umons.ac.be

Reflux Sign Assessment			Scores
Oral cavity			
1. Anterior pillar erythema	absent = 0	présent = 4	
2. Uvula erythema ± edema	absent = 0	présent = 3	
3. Coated tongue	absent = 0	présent = 2	
Oral cavity subscore	…/9		
Pharyngeal cavity			
1. Nasopharyngeal wall erythema ± inflammatory granulations	absent = 0	présent = 2	
2. Posterior oro- or hypopharyngeal wall erythema	absent = 0	présent = 4	
3. Posterior oro- or hypopharyngeal wall inflammatory granulations	absent = 0	présent = 3	
4. Tongue tonsil hypertrophy:	no hypertrophy = 0 Apparent vallecula only when tongue sticked = 3 Unapparent vallecula irrespective the tongue = 4		
5. Contact between epiglotitis and tongue tonsils	absent = 0	présent = 4	
6. Pharyngeal sticky mucus	absent = 0	présent = 4	
Pharyngeal cavity subscore	…/21		
Larynx			
Subglottic edema ± erythema	absent = 0	présent = 1	
Ventricular band erythema ± edema	absent = 0	présent = 2	
Epiglottis redness ± edema	absent = 0	présent = 3	
Posterior commissure & retro-cricoid			
1. Erythema	Absent = 0 Arytenoids/inter-arytenoid only = 4 Diffuse to posterior commissure = 5		
2. Inter-arytenoid granulatory tissue	absent = 0	présent = 2	
3. Posterior commissure hypertrophy	absent = 0	présent = 5	
4. Retro-cricoid erythema	absent = 0	présent = 3	
5. Retro-cricoid edema (=contact between retro-cricoid area & hypopharyngeal posterior wall during breathing/opening glottis)	absent = 0	présent = 4	
Vocal folds			
1. Endolaryngeal sticky mucus deposit	absent = 0	présent = 3	
2. Vocal fold erythema	absent = 0	présent = 1	
3. Edema of the free-edge or the entire vocal folds	absent = 0	présent = 1	
4. Vocal fold lesions (2-point per lesion): granuloma(s), nodules, polyp(s), Reinke's edema, ulceration(s), keratosis.	…		
Laryngeal subscore	.../42		
RSA Total score:	…/72		

Figure 1. Reflux Sign Assessment.
Note. The tool is subdivided into three parts according to the sign localization: oral cavity, pharynx, and larynx. The total score is calculated by the sum of each item score. The rater adds two additional points in case of benign lesion of the vocal folds because they can be associated with LPR. The maximum score is 72.

A distal reflux event was defined as an episode in which reflux reached the two impedance sensors closest to the lower esophageal sphincter (LES). A proximal reflux event was defined as an episode in which reflux reached the two impedance sensors in the hypopharynx. An acidic event consisted of gaseous or liquid reflux with a pH $\leq$ 4.0, while a nonacidic event was a gaseous or liquid reflux with a pH $>$ 4.0. Regarding recent study, the LPR diagnosis was based on the occurrence of $\geq$1 proximal episode.[17] GERD was defined when $>$4.0% of the 24-hour recording had a pH below 4.0 or a DeMeester score $>$14.72. An acid reflux episode consisted of an episode with a pH $\leq$ 4.0. A nonacid reflux episode consisted of an episode with a pH $>$ 4.0. Because there are no guidelines for the definition of acid, nonacid and mixed LPR

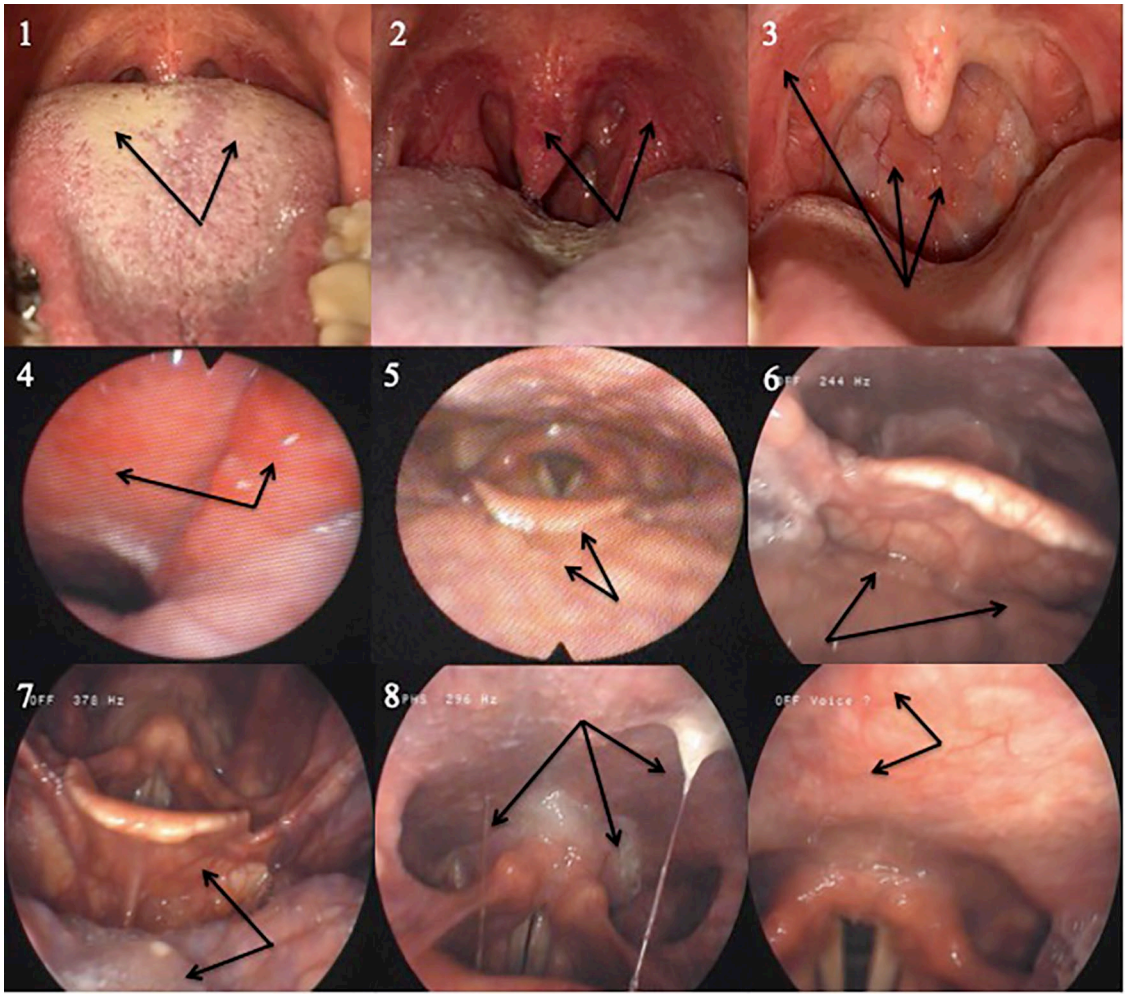


Figure 2. Laryngopharyngeal reflux signs described in Reflux Sign Assessment.
Note. Reflux Sign Assessment considers the following LPR-associated signs in the scoring system: coated tongue (1), uvula and anterior pillar erythema (2, 3), oropharyngeal erythema and inflammatory granulations (3), nasopharyngeal erythema (4), moderate-to-severe tongue tonsil hypertrophy (5, in this case, the vallecula were not apparent when the tongue was sticked leading to a score of 4), mild-to-moderate tongue tonsil hypertrophy (6 & 7, in this case, the vallecula are apparent when the tongue was sticked leading to a score of 3), pharyngeal sticky mucus (or pooling in the left sinus piriform), and erythema of the oro-hypopharyngeal posterior wall.

disease, LPR was defined as acid when the ratio of the number of acid reflux episodes to the number of nonacid reflux episodes was >2. LPR was defined as nonacid when the ratio of the number of acid reflux episodes to the number of nonacid reflux episodes was <0.5. Mixed reflux consisted of a ratio ranging from 0.51 to 2.0.

Treatments

The therapeutic algorithm was based on recent recommendations of our LPR Study Group.[4,18] The MII-pH characteristics, that is, the time/position at which episodes occurred (upright and daytime/recumbent and night-time), the reflux profile (acid, nonacid, mixed), and the occurrence of GERD, were used to propose a personalized treatment scheme[13] including diet, behavioral changes, and PPIs (pantoprazole, 20 mg once or twice daily) ± alginate (Gaviscon Advance®, Reckitt Benckiser, Slough, UK) ± magaldrate (Riopan®, Takeda, Zaventem, Belgium). Each patient received a validated grid with recommended diet and behavior changes that took into account the patient's personalized habits.[19,20]

Statistical Methods

To perform the various analyses, videolaryngostroboscopies and oral cavity photos of all patients and healthy individuals

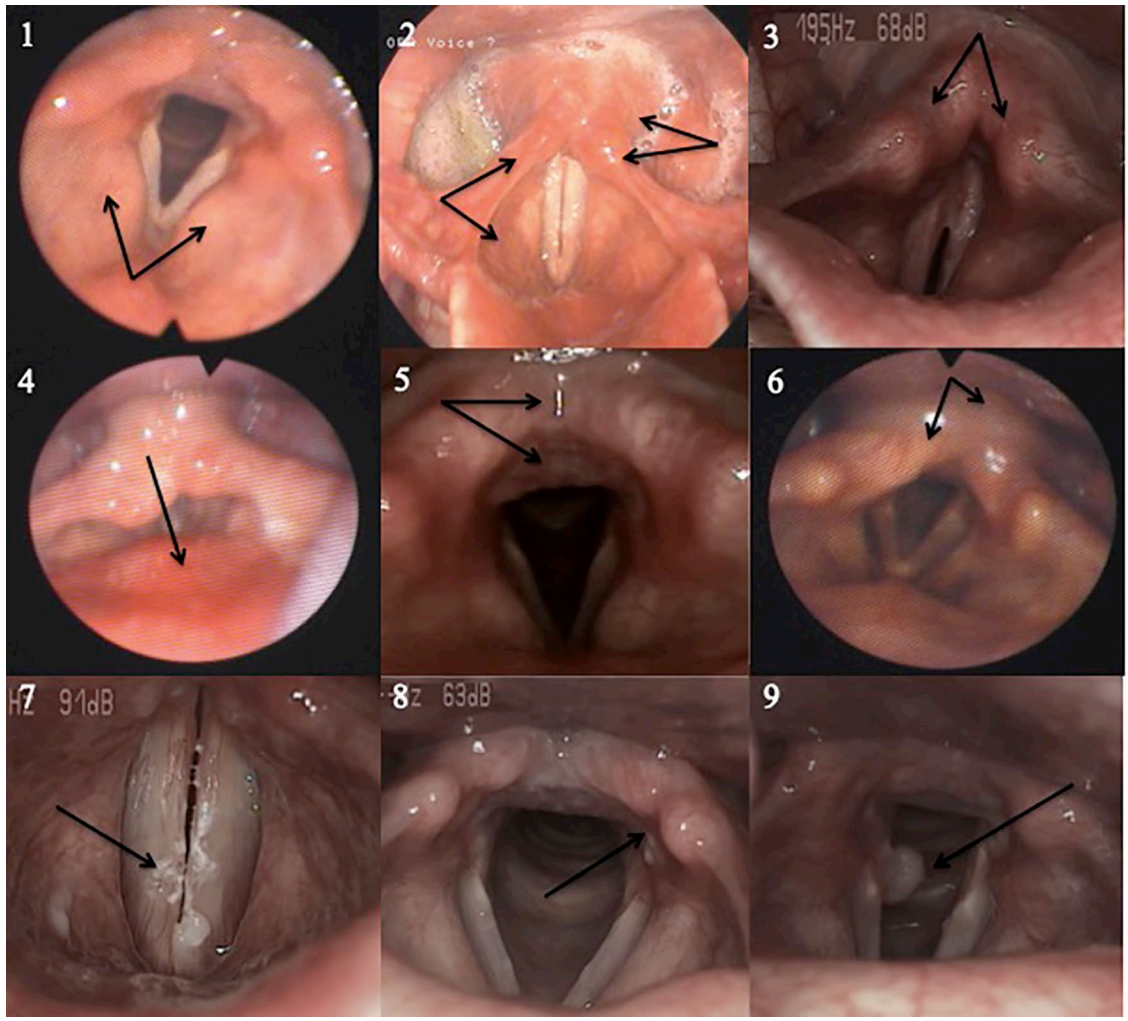


Figure 3. Laryngopharyngeal reflux signs described in Reflux Sign Assessment (2).
Note. Reflux Sign Assessment considers the following LPR-associated signs in the scoring system: hypertrophy of the ventricular bands (1), diffuse laryngeal erythema (2; ie, ventricular bands, retrocricoid area, posterior commissure), arytenoid/inter-arytenoid erythema (3), epiglottis erythema, granulation of the posterior commissure (5), hypertrophy of the posterior commissure (5, 6), retrocricoid edema (6), endolaryngeal sticky mucus (7), keratosis (8), and granuloma (9).

were stored on a database. Physicians had access to the database without knowing the time of the recording (baseline *versus* posttreatment).

Reliability. Internal consistency was measured using Cronbach's alpha for all items on the baseline RSA for patients and controls. Intrarater reliability was assessed through test-retest blinded evaluations of signs, which were performed for all included patients by a laryngologist at 7-day intervals. The test-retest analysis was made on the recordings stored in the database. Each item and the RSA total score were analyzed using Spearman's rank correlation coefficient (r_s); $r_s \geq 0.80$ was considered ideal, while $r \geq 0.70$ was considered adequate for the reliability of the total score.[8]

Concordance. The concordance analysis (interrater reliability) was assessed by comparing the RSA evaluations of three blinded otolaryngologists (specifically, one laryngologist (CF) and two general otolaryngologists (JRL, FB)) with Kendall's W (coefficient of concordance). Similar analysis was made for RFS on the same recordings. Kendall's W was used to assess the similarity between the judges' ratings of the RSA subscores; RFS and RSA total scores. Because the assessment of findings may be tiring (2-4 minute to rate both RSA and RFS on the photo and video recordings; impacting the quality of the assessment), we limited the number of patients to 56 ($N < 60$) for the analyses of interrater reliability. We selected the 56 first recruited patients.

Table 1. Definition of the Measurement Properties of Signs of Instruments Analyzed in the Study.

Domain	Definition
Conceptual model	
Construct definition	It provides a rationale for and description of the concepts and target population that a measure is intended to assess.
Target population	
Expected subscales	
Content validity	It refers to evidence that an instrument is appropriate for its intended use. Items and conceptual domains must be relevant to the targeted population.
Content expert involved	The instrument's development of signs must include direct input from experts. There should be a clear description of the process by which included signs were derived.
Description of item development	The items described in the instrument must reflect the most common signs encountered in the disease.
Reliability	The degree to which scores are free from random (measurement) error.
Internal consistency reliability	Extent to which items within each domain are interrelated.*
Test-retest reliability	Stability of scores over time when no change is expected in the concept of interest.*
Concordance	The degree of agreement among raters.
Construct validity	It refers to whether an instrument measures intended theoretic constructs or traits and directly affects the appropriateness of the measurement-based inferences.
Responsiveness to change	The extent to which an instrument detects meaningful changes over time that have occurred after baseline.***
Convergent validity	The degree to which the sign score correlates with other instruments measuring the same construct or with related clinical indicators.**
Known-groups validity	The extent to which the instrument can discriminate between groups that are known to differ on the variables being measured.***
Interpretability & scoring	The degree to which the meaning of the scores can be easily understood.
Plan for scoring measure	A description of how to score the measure should be provided (sum, algorithm).
Plan for missing data	A prespecified plan for managing missing responses can mitigate the risk of bias resulting from the necessity to exclude cases with missing data.
Scaling described	The process of distributing the full range of respondents' possible scores with respect to the measured attribute.

Note. *consistent: >0.70 for group-level comparisons; **<0.30, low correlation; 0.30 to 0.60, moderate correlation; >0.60, strong correlation (Pearson or Spearman analysis); ***large change: >0.80; moderate change: 0.50 to 0.79; small change: 0.2 to 0.49.

Validity. Convergent validity was assessed through an analysis of the correlation between the RSA and RFS of all included patients (Spearman's rank correlation coefficient). Internal validity was assessed through a statistical comparison of the RSA items and total scores of LPR patients and asymptomatic individuals using the Mann–Whitney *U* test.

Construct validity. The responsiveness to change of the RSA and RFS was assessed by comparing the baseline and 12 weeks posttreatment scores. For this analysis, we used the video recordings of all included patients. The convergent validity was assessed through a study of the similarity of the evolution of both scores. Changes in the RSA and RFS from pre- to posttreatment were evaluated using the Wilcoxon signed-rank test.

Normative data. The RSA threshold for determining the presence and absence of LPR was examined with a receiver operating characteristic (ROC) analysis based on the data of LPR patients and healthy individuals.

Statistical analyses were performed using the Statistical Package for the Social Sciences for Windows (SPSS version 22,0; IBM Corp, Armonk, NY, USA). Multiple linear regression was used to identify potentially significant relationships between patient characteristics, relevant GI findings, reflux types and clinical presentation. A level of significance of $P < .05$ was used.

Results

The characteristics of the patients are presented in Table 2. There were 42 patients with acid LPR, 27 with mixed LPR and 33 with nonacid LPR. Forty-nine patients had GERD. Esophagitis, hiatal hernia and LES insufficiency were found in 33.9%, 31.3%, and 48.8% patients, respectively. GI endoscopy findings were normal in 23.8% of patients. There were no significant associations between

Table 2. Characteristics of LPR Patients.

Characteristics		
Age	N/mean ± SD	Range
Mean ± SD	52.8 ± 16.0	21-77
Gender		
Male	34	33%
Female	68	67%
GI endoscopy (N = 80)	N	Prevalence
Normal	19	23.8%
Esophagitis (LA Grading system)		
Los Angeles Grade A	*25*	31.3%
Los Angeles Grade B	*1*	1.3%
Los Angeles Grade C	*0*	0%
Los Angeles Grade D	*1*	1.3%
Hiatal hernia	25	31.3%
LES insufficiency	39	48.8%
Gastritis	24	30.0%
Duodenitis	2	2.5%
Helicobacter pilori infection	3	3.8%
LPR profiles	N	Prevalence
Acid reflux	42	41.1%
Nonacid reflux	33	32.4%
Mixed reflux	27	26.5%
GERD	49	48.0%

Abbreviations: GERD, gastroeosophageal reflux disease; GI, gastrointestinal; LA, Los Angeles; LES, lower esophageal sphincter; SD, standard deviation.

patient characteristics, GI endoscopy findings, GERD, types of reflux (acid, nonacid, mixed), and RSA.

Reliability

Cronbach's alpha for patients and controls was 0.821, which indicates high internal consistency. The test-retest reliability was high for the total score ($r_s = 0.813$, $P < .001$) and moderate to high for all item scores (Appendix 2).

Concordance

The concordance (Kendall's W) was significant for all subscores of the RSA (oral cavity, pharyngeal, laryngeal and vocal cord lesions; $P < .001$) and for the RSA total score ($P < .001$); the concordance for RSA total score being strong (W = 0.663). The concordance for RFS total score was lower than RSA (W = 0.242; Table 3).

Validity

The RSA total score was correlated with the RFS total score of the LPR patients ($r_s = 0.668$, $P < .001$), indicating high external validity. The correlation between RFS and RSA laryngeal subscore was higher and significant ($r_s = 0.835$, $P < .001$). The mean RSA of the healthy controls was 7.52.

Table 3. Interrater Reliability Analysis (Concordance).

Reflux Sign Assessment Subscores	Kendall'W	*P*-value
Oral cavity subscore	0.480	<.001
Pharyngeal cavity subscore	0.443	<.001
Laryngeal subscore	0.425	<.001
RSA Total score	0.663	<.001
RFS Total score	0.242	.001

Note. According to recommendation,[8] Kendall's W > 0.600 consists of a strong correlation among raters. Abbreviations: RFS, reflux finding score; RSA, reflux sign assessment; W, Kendall's W.

An RSA score >14 may be suggestive of LPR according to the ROC curve (Figure 4; N = 98). This value has a sensitivity and specificity of 89.1 and 95.2, respectively. RSA had a higher discriminatory value than the RFS. The RSA showed high internal validity because all item scores for the RSA were significantly higher in the LPR patients than in the asymptomatic individuals (Table 4).

Responsiveness

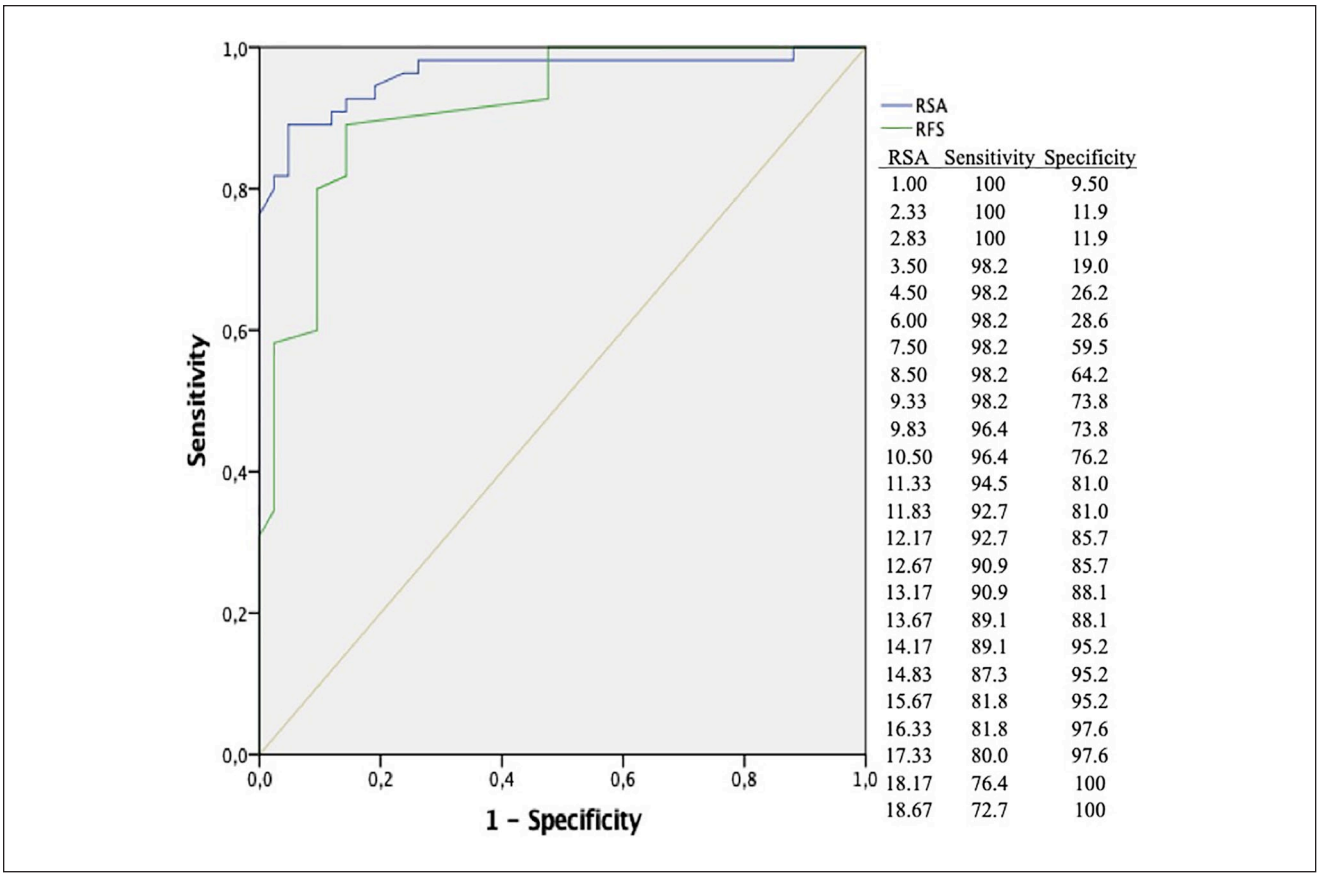
The RSA and the RFS total scores significantly improved from baseline to 3 months posttreatment (Table 5). The improvement of patients with acid, nonacid, and mixed LPR did not significantly differ.

Discussion

The objective of this study was to report the reliability and validity of the RSA, which was developed to assess laryngeal and extralaryngeal findings associated with LPR. Some clinical instruments have already been developed without validation studies or with poor properties, especially poor interrater reliability, which limits their daily use (Appendix 3).[5,21-25]

The content of the RSA was based on an initial assessment of the prevalence of signs in 101 LPR patients and considers many extralaryngeal signs associated with LPR, that is, hypo- or oropharyngeal erythema,[22,26] uvular erythema,[26] retrocricoid edema,[27,28] coated tongue,[1] and lingual tonsil hypertrophy.[29] These signs were ignored in the majority of studies conducted over the past few years and are missing from the currently available clinical instruments[1,8] despite the fact that they are found in many patients. The initial evaluation of signs in the 101 LPR patients allowed weight assignments for each item of the RSA that could improve the sensitivity of the instrument.[8]

The concurrent internal consistency and test-retest reliabilities of the RSA were both >0.800, indicating excellent reliability. Intrarater reliability has been reported for four other clinical instruments for assessing LPR,[5,21,23,24] which show a range of test-retest reliability values from 0.26 to



RSA	Sensitivity	Specificity
1.00	100	9.50
2.33	100	11.9
2.83	100	11.9
3.50	98.2	19.0
4.50	98.2	26.2
6.00	98.2	28.6
7.50	98.2	59.5
8.50	98.2	64.2
9.33	98.2	73.8
9.83	96.4	73.8
10.50	96.4	76.2
11.33	94.5	81.0
11.83	92.7	81.0
12.17	92.7	85.7
12.67	90.9	85.7
13.17	90.9	88.1
13.67	89.1	88.1
14.17	89.1	95.2
14.83	87.3	95.2
15.67	81.8	95.2
16.33	81.8	97.6
17.33	80.0	97.6
18.17	76.4	100
18.67	72.7	100

Figure 4. Receiver operating characteristic (ROC) curve of the Reflux Sign Assessment and Reflux Finding Score.
Note. A cutoff >14 may be suggestive of LPR with a sensitivity of 89.1 and a specificity of 95.2. RSA is more discriminant than RFS for the LPR diagnosis (curves).

Table 4. Comparison of RSA of LPR and Asymptomatic Individuals.

Reflux Sign Assessment Items	LPR	Healthy	*P*-value
1. Anterior pillar erythema	3.30 ± 1.44	1.71 ± 2.00	<.001
2. Uvula erythema ± edema	1.53 ± 1.40	0.01 ± 0.01	<.001
3. Coated tongue	1.29 ± 0.83	0.19 ± 0.59	<.001
Oral cavity subscore	5.79 ± 2.69	1.90 ± 2.21	<.001
1. Nasopharyngeal wall erythema ± inflammatory granulations	0.68 ± 0.91	0.14 ± 0.52	.001
2. Posterior oro- or hypopharyngeal wall erythema	2.27 ± 1.74	0.29 ± 0.96	<.001
3. Posterior oro- or hypopharyngeal wall inflammatory granulations	0.90 ± 1.23	0.29 ± 0.90	.001
4. Tongue tonsil hypertrophy	2.33 ± 1.34	1.45 ± 1.64	.005
5. Contact between epiglotitis and tongue tonsils	2.75 ± 1.76	0.86 ± 1.66	<.001
6. Pharyngeal sticky mucus	2.18 ± 1.77	0.24 ± 0.91	<.001
Pharyngeal cavity subscore	9.90 ± 4.80	3.24 ± 2.99	<.001
Sub- and supraglottic areas			
1. Subglottic edema ± erythema	0.04 ± 0.16	0.02 ± 0.15	.112
2. Ventricular band erythema ± edema	1.28 ± 0.82	0.01 ± 0.01	<.001
3. Epiglottis redness ± edema	1.11 ± 1.30	0.07 ± 0.46	<.001
Posterior commissure			
1. Commissure posterior/arytenoid erythema	3.52 ± 1.95	0.90 ± 1.76	<.001
2. Inter-arytenoid granulatory tissue	0.42 ± 0.76	0.01 ± 0.01	<.001

(continued)

Table 4. (continued)

Reflux Sign Assessment Items	LPR	Healthy	*P*-value
3. Posterior commissure hypertrophy	3.52 ± 1.95	0.29 ± 1.04	<.001
4. Retrocricoid erythema	0.88 ± 1.31	0.01 ± 0.01	<.001
5. Retrocricoid edema	1.77 ± 1.77	0.38 ± 1.19	<.001
Vocal folds			
1. Endolaryngeal sticky mucus deposit	1.29 ± 1.25	0.71 ± 1.29	.003
2. Vocal fold erythema	0.07 ± 0.22	0.02 ± 0.15	.063
3. Edema of the free-edge or the entire vocal folds	0.07 ± 0.18	0.02 ± 0.15	.003
4. Vocal fold lesions	0.07 ± 0.36	0.00 ± 0.00	.179
Laryngeal subscore	12.55 ± 6.07	2.38 ± 2.77	<.001
RSA Total score	25.95 ± 9.58	7.52 ± 4.37	<.001

Note. The statistics were made through Mann–Whitney *U* Test (102 LPR patients and 42 healthy individuals). Abbreviations: LPR, laryngopharyngeal reflux; RSA, reflux sign assessment.

Table 5. Pre- to Posttreatment Sign Evolution Regarding RSA and RFS.

Reflux Sign Assessment Items	Pretreatment	Posttreatment	*P*-value
1. Anterior pillar erythema	3.30 ± 1.44	3.06 ± 1.61	.151
2. Uvula erythema ± edema	1.53 ± 1.40	1.44 ± 1.34	.738
3. Coated tongue	1.29 ± 0.83	1.35 ± 0.82	.806
Oral cavity subscore	5.79 ± 2.69	4.79 ± 2.42	<.001
1. Nasopharyngeal wall erythema ± inflammatory granulations	0.68 ± 0.91	0.19 ± 0.52	.010
2. Posterior oro- or hypopharyngeal wall erythema	2.27 ± 1.74	1.49 ± 1.72	.002
3. Posterior oro- or hypopharyngeal wall inflammatory granulations	0.90 ± 1.23	0.60 ± 1.03	.021
4. Tongue tonsil hypertrophy	2.33 ± 1.34	1.92 ± 1.29	.036
5. Contact between epiglotitis and tongue tonsils	2.75 ± 1.76	2.24 ± 1.87	.059
6. Pharyngeal sticky mucus	2.18 ± 1.77	1.73 ± 1.73	.024
Pharyngeal cavity subscore	9.90 ± 4.80	6.41 ± 3.86	<.001
Sub- and supraglottic areas			
1. Subglottic edema ± erythema	0.04 ± 0.16	0.02 ± 0.02	.109
2. Ventricular band erythema ± edema	1.28 ± 0.82	0.97 ± 0.89	.023
3. Epiglottis redness ± edema	1.11 ± 1.30	0.54 ± 0.93	<.001
Posterior commissure			
1. Commissure posterior/arytenoid erythema	3.52 ± 1.95	2.44 ± 2.09	<.001
2. Inter-arytenoid granulatory tissue	0.42 ± 0.76	0.14 ± 0.43	.034
3. Posterior commissure hypertrophy	3.52 ± 1.95	2.44 ± 2.09	<.001
4. Retrocricoid erythema	0.88 ± 1.31	0.51 ± 0.94	.004
5. Retrocricoid edema	1.77 ± 1.77	1.37 ± 1.75	.180
Vocal folds			
1. Endolaryngeal sticky mucus deposit	1.29 ± 1.25	0.92 ± 1.18	.095
2. Vocal fold erythema	0.07 ± 0.22	0.05 ± 0.17	.058
3. Edema of the free-edge or the entire vocal folds	0.07 ± 0.18	0.08 ± 0.18	.796
4. Vocal fold lesions	0.07 ± 0.36	0.05 ± 0.31	.180
Laryngeal subscore	12.55 ± 6.07	7.83 ± 5.05	<.001
RSA Total score	25.95 ± 9.58	18.96 ± 7.58	<.001
RFS			
Subglottic edema	0.01 ± 0.01	0.01 ± 0.01	1.00
Ventricular obliteration	1.57 ± 1.16	1.03 ± 0.99	.060
Arytenoid/diffuse redness	2.14 ± 1.31	1.28 ± 1.28	.002
Vocal folds edema	0.15 ± 0.45	0.12 ± 0.38	.272
Diffuse laryngeal edema	0.57 ± 0.76	0.33 ± 0.51	.455
Posterior commissure hypertrophy	1.43 ± 0.80	1.01 ± 0.80	.033
Granuloma/Granulation	0.26 ± 0.65	0.20 ± 0.51	.157
Endolaryngeal mucous	0.92 ± 0.93	0.61 ± 0.79	.217
RFS Total score:	6.37 ± 2.61	4.04 ± 2.60	.003

Note. Statistics were made through Wilcoxon signed-rank test. The RSA and RFS assessments were blinded regarding the time of evaluation (pre *versus* posttreatment). Abbreviations: RFS, reflux finding score; RSA, reflux sign assessment.

0.95; the RFS had the highest intrarater reliability in the initial study of Belafsky et al (Appendix 3). The intrarater reliability of these four clinical instruments has never been tested in other studies.

The most important challenge of this study was to develop an instrument that exhibited high interrater reliability. Indeed, because the signs are nonspecific and may be found in healthy individuals, their assessment was historically characterized by poor interrater reliability irrespective of the clinical instrument used.[30,31] Because LPR is primarily diagnosed by general otolaryngologists and, to a lesser extent, by laryngologists or phoniatricians in many countries,[32] we deliberately chose two general otolaryngologists and one laryngologist when assessing the interrater reliability. As suggested in the study of Chang et al,[32] the rates of a finding could differ according to the subspecialty of the physician. For example, laryngologists could be more aware of some laryngeal signs of reflux, which remains more prevalent in laryngology offices than in general otolaryngological consultation. In this study, we found an interrater reliability of 0.663 for the RSA total score; which was better than those for RFS (0.242). This result supports overall good interrater reliability for the RSA. Comparing our results with those reported in the current literature, we might propose that the good interrater reliability of the RSA is attributable to the descriptive nature of our item severity grading system and the absence of items involving subjective evaluation. In their initial study, Belafsky et al found an interrater reliability of 0.90 for two laryngology-trained observers who completed the RFS based on clinical photos.[5] These results, however, were not confirmed in the study of Chang et al. which found poor concordance values for RFS total score and items, with intraclass correlation coefficients (ICC) ranging from 0.37 to 0.58.[32] Williams et al developed a descriptive clinical instrument for reflux laryngitis in 2004: the Laryngoscopic Grading Scale (LGS).[25] Like the RSA, the LGS was mainly descriptive and reported high concordance (69.4%) among three observers. The main weakness of the LGS is the lack of consideration of extralaryngeal signs and the non-validation of many properties, that is, test-retest reliability, internal consistency and convergent validity (Appendix 3). The other reported clinical instruments also showed moderate interrater reliability values ranging from 0.30 to 0.58.[21,23,24] Overall, the interrater reliability of the LPR instrument is still difficult to generalize due to differences among centers and physicians in material, education, experience, etc.

As reported in many previous studies, the analysis of interrater reliabilities of each sign is usually poor; which may decrease the interrater reliability of the total score of the clinical instrument. In that respect, future studies could combine the use of 'color technologies' with the current instrument. For example, the use of software to analyze the intensity and extent of redness of the mucosa could be coupled to the RSA to decrease the subjectivity of physicians' assessments of erythema. Another point that may be important to consider in the interrater reliability analysis is the material used to assess the clinical findings. In the majority of previous studies, the authors have used clinical photos to assess the reliability of an instrument. In practice, physicians do not usually use the photos of the examination to determine a severity score; instead, they determine the clinical state based on the entire nasofibroscopic examination, which is better represented in videos of nasofibroscopic examinations. Thus, to more closely approximate clinical practice, we only used videos of the nasofibroscopic examinations and avoided the use of photos, which may positively influence interrater reliability depending on the point at which the photo was taken.

Convergent validity was assessed through a correlation study between the RSA and RFS. The correlation coefficient was high, indicating good convergent validity. The significant correlation between the RFS, which assesses only laryngeal signs, and the RSA, which considers laryngeal and extralaryngeal signs, is partly due the high weighting of the laryngeal items in the RSA (42 points out of the 72-point score). The RSA total and item scores were significantly higher in LPR patients than in healthy individuals, indicating high convergent validity. The convergent validity of previous instruments has not been assessed, which limits our ability to compare our findings with the literature.

The significant improvements of RSA and RFS throughout treatment indicate that the RSA displays high construct validity. Similarly, other instruments have reported good responsiveness to change from pre- to posttreatment. The inclusion in the RSA of signs that were rarely studied over the past decades led to surprising results. Overall, the pharyngeal and laryngeal RSA subscores improved from pre- to posttreatment, but the pattern of sign evaluation seemed to vary according to the type of sign (erythema, edema, etc.) and the anatomical area. In other words, signs did not evolve similarly throughout the first part of the treatment. Overall, mucosal erythema of the nasopharynx, larynx, and posterior commissure and retrocricoid could be the most 'sensitive' signs of potential improvement. In the same way, edema of the posterior commissure and lingual tonsils (through contact between the epiglottis and tongue base) seems to improve throughout the first 3 months of treatment. A similar finding was observed for posterior wall inflammatory granulations. However, oral signs (ie, anterior pillar erythema, coated tongue, and uvular erythema/edema) did not change over the 3-month therapeutic course; some of these signs were among the most prevalent findings in LPR. They may require more time to disappear. The realization of additional research to study the evolution of these signs makes sense for two reasons. First, these signs are easy to observe, irrespective of the physician's specialty. Second, according to our clinical practice, they can reflect the chronic course of the disease, especially in the case of anterior pillar

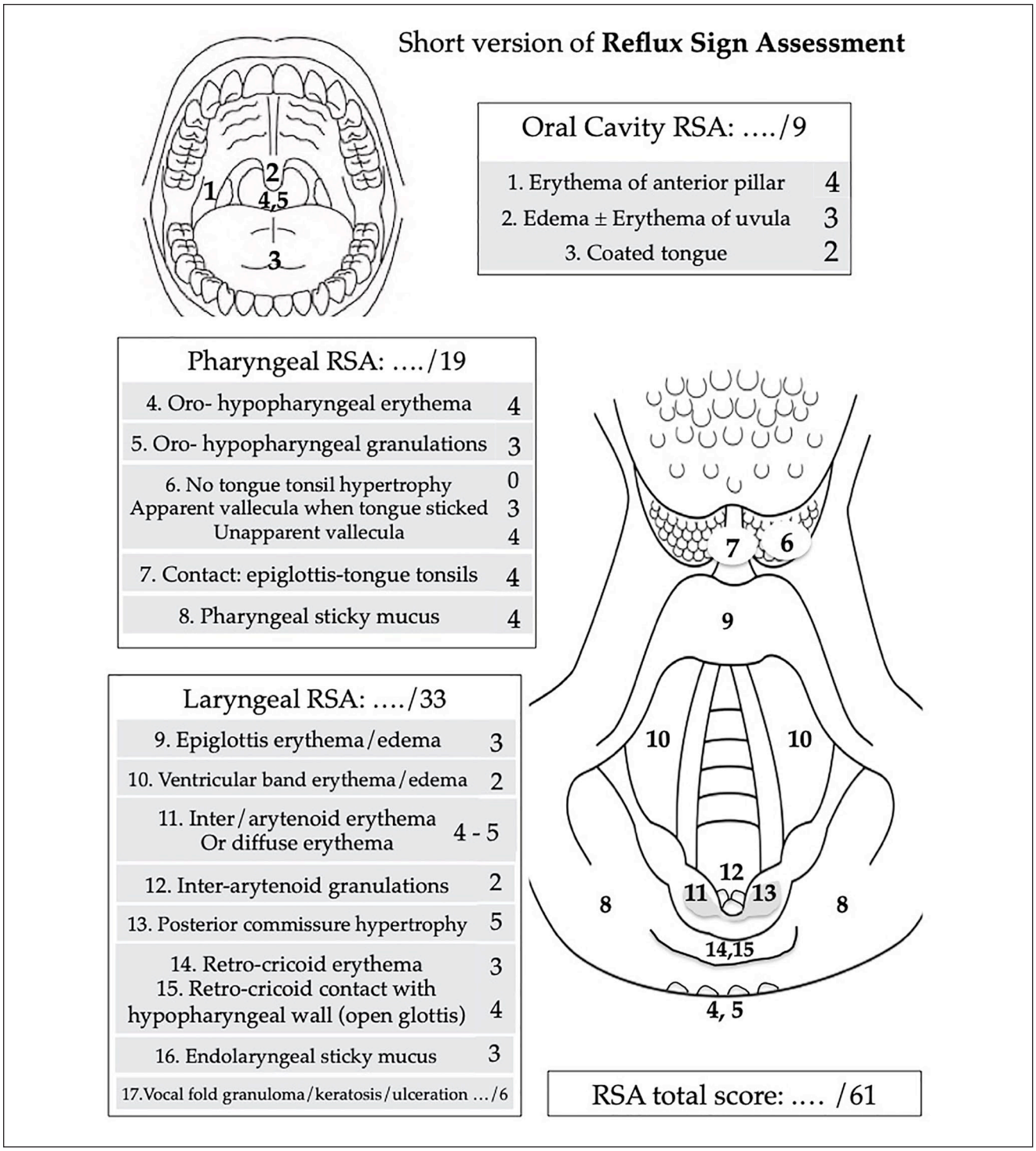


Figure 5. Short version of RSA.

Note. Items with a prevalence <20% were excluded. In addition and because many examinations are made through epipharyngoscopy, nasopharyngeal erythema was also removed. The short version of RSA consists of 16 items. The vocal fold granuloma, keratosis, or ulceration count for 2-point by lesion. This version is designed for the assessment of LPR findings in otolaryngology office.

erythema. Indeed, the three main investigators of the study (JRL, FB, & CF) often met patients in their respective reflux clinics with a history of chronic reflux and persistent erythema of the anterior pillar after the "remission" period. A certain degree of posterior commissure hypertrophy could also be persistent for the long term. Naturally, these

clinical observations are still empirical and need to be investigated through prospective long-term studies of LPR-related findings. The identification of very sensitive signs in oral or oropharyngeal areas could be helpful for family practitioners who could easily suspect LPR. The blinded assessment of signs is the main strength of this study because the knowledge of the patient's symptoms (or clinical state: 'cured' *versus* ' not cured') significantly impacts the finding assessment in LPR.[32] In that way, future studies interesting to the evolution of signs throughout treatment should consider a blinded assessment in order to avoid biased conclusion. Regarding the tiring aspect of the task and with regard to previous study,[32] we limited the interrater reliability analysis to the 56 first patients. We do not consider this point as a weakness regarding the time and the concentration required to analysis 112 laryngopharyngeal videos and 112 photos of the mouth with two clinical scores.

The improvement of signs was similar in patients with acid, nonacid and mixed LPR, which supports the results of our previous study about the efficacy of personalized treatment regarding the MII-pH results.[13]

The main weakness of the RSA is probably the length of time needed for the physician to score each item (1 minute), but given the nonspecificity of LPR-related signs, in a first step, it appeared important to consider all of these items. As experienced in this study, adequate training of the otolaryngologist could improve the use of the RSA and overcome this weakness. The development of a shorter version of the RSA could address this problem. Based on our data, the prevalence and the relevance of the studied signs, we propose a simplified version considering the most relevant findings (Figure 5). Regarding the cost and the discomfort, we did not realize MII-pH in healthy individuals to be sure that they had no LPR; representing the second weakness of this study. In addition, we did not consider the presence of obstructive apnea syndrome in LPR patients; some findings (eg, tongue tonsil hypertrophy) should be associated with apnea syndrome.

Conclusion

To better track the changes in LPR-associated signs throughout treatment, some instruments for assessing laryngeal signs have been developed over the past few years. However, these instruments do not consider the majority of extralaryngeal signs, and their validation processes were incomplete; thus, they report controversial findings. For these reasons, the members of the LPR Study Group of YO-IFOS have developed the RSA, which is a complete clinical instrument for evaluating laryngeal and extralaryngeal findings associated with LPR. The RSA demonstrated high intra- and interrater reliabilities and responsiveness to change. An RSA score >14 could be considered abnormal and may be suggestive of LPR.

Acknowledgments

Dr. Petros D. Karkos, Dr. Lise Crevier-Buchmann, Dr. Antonio Schindler, Dr. Abdul Latif Hamdan, and all members of the Expert committee where the content of the Reflux Symptom Score had been developed. American Journal Expert for the proofreading. Vesale Foundation (CHU Saint-Pierre, ULB, Brussels, Belgium) & IRIS-Recherche (Fondation Roi Baudouin) for the grants.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Acknowledgment to IRIS-Recherche (Fondation Roi Baudouin) and Vesale Grant (CHU Saint-Pierre).

ORCID iDs

Jérôme R. Lechien https://orcid.org/0000-0002-0845-0845

Kathy Huet https://orcid.org/0000-0002-7160-7489

Supplemental Material

Supplemental material for this article is available online.

References

1. Lechien JR, Saussez S, Schindler A, et al. Symptoms and signs outcomes of laryngopharyngeal reflux treatment: a critical systematic review and meta-analysis. *Laryngoscope*. 2019;129(5):1174-1187.
2. Koufman JA. The otolaryngologic manifestations of gastroesophageal reflux disease (GERD): a clinical investigation of 225 patients using ambulatory 24-hour pH monitoring and an experimental investigation of the role of acid and pepsin in the development of laryngeal injury. *Laryngoscope*. 1991;101(4 Pt 2 Suppl 53):1-78.
3. Koufman JA, Aviv JE, Casiano RR, Shaw GY. Laryngopharyngeal reflux: position statement of the committee on speech, voice, and swallowing disorders of the American Academy of Otolaryngology-Head and Neck Surgery. *Otolaryngol Head Neck Surg*. 2002;127(1):32-35.
4. Lechien JR, Akst LM, Hamdan AL, et al. Evaluation and management of laryngopharyngeal reflux disease: state of the art review. *Otolaryngol Head Neck Surg*. 2019;160(5):762-782.
5. Belafsky PC, Postma GN, Koufman JA. The validity and reliability of the reflux finding score (RFS). *Laryngoscope*. 2001;111(8):1313-1317.
6. Belafsky PC, Postma GN, Koufman JA. Validity and reliability of the reflux symptom index (RSI). *J Voice*. 2002;16(2):274-277.
7. Karkos PD, Wilson JA. Empiric treatment of laryngopharyngeal reflux with proton pump inhibitors: a systematic review. *Laryngoscope*. 2006;116(1):144-148.
8. Lechien JR, Schindler A, De Marrez LG, et al. Instruments evaluating the clinical findings of laryngopharyngeal reflux: a systematic review. *Laryngoscope*. 2019;129(3):720-736.

9. Francis DO, Patel DA, Sharda R, et al. Patient-reported outcome measures related to laryngopharyngeal reflux: a systematic review of instrument development and validation. *Otolaryngol Head Neck Surg*. 2016;155(6):923-935.
10. Ayad T, Stephenson K, A L Smit D, et al. Young Otolaryngologists of International Federation of Oto-rhino-laryngological Societies (YO-IFOS) committees. *Eur Ann Otorhinolaryngol Head Neck Dis*. 2018;135(5S):S61-S65.
11. Lechien JR, Benninger MS, Crevier-Buchman L, et al.; IFOS. Laryngopharyngeal reflux and voice disorders. Paper presented at: World Ear, Nose, and Throat Congress; June 2017; Paris, France.
12. Reichel O, Issing WJ. Should patients with pH-documented laryngopharyngeal reflux routinely undergo oesophagogastroduodenoscopy? A retrospective analysis. *J Laryngol Otol*. 2007;121(12):1165-1169.
13. Lechien JR, Bobin F, Muls V, et al. Validity and reliability of the reflux symptom score. *Laryngoscope*. 2019. doi:10.1002/lary.28017
14. Lechien JR, Finck C, Huet K, et al. Impact of age on laryngopharyngeal reflux disease presentation: a multi-center prospective study. *Eur Arch Otorhinolaryngol*. 2017;274(10):3687-3696.
15. Bobin F, Journe F, Lechien JR. Saliva pepsin level of laryngopharyngeal reflux patients is not correlated with reflux episodes. *Laryngoscope*. 2019. doi:10.1002/lary.28260
16. Sifrim D, Castell D, Dent J, Kahrilas PJ. Gastro-oesophageal reflux monitoring: review and consensus report on detection and definitions of acid, non-acid, and gas reflux. *Gut*. 2004;53(7):1024-1031.
17. Hoppo T, Sanz AF, Nason KS, et al. How much pharyngeal exposure is "normal"? Normative data for laryngopharyngeal reflux events using hypopharyngeal multichannel intraluminal impedance (HMII). *J Gastrointest Surg*. 2012;16(1):16-24; discussion 24-25.
18. Lechien JR, Saussez S, Karkos PD. Laryngopharyngeal reflux disease: clinical presentation, diagnosis and therapeutic challenges in 2018. *Curr Opin Otolaryngol Head Neck Surg*. 2018;26(6):392-402.
19. Lechien JR, Huet K, Khalife M, et al. Alkaline, protein, low-fat and low-acid diet in laryngopharyngeal reflux disease: our experience on 65 patients. *Clin Otolaryngol*. 2019;44(3):379-384.
20. Lechien JR, Bobin F, Mouawad F, et al. Development of scores assessing the refluxogenic potential of diet of patients with laryngopharyngeal reflux [published online September 12, 2019]. *Eur Arch Otorhinolaryngol*. doi:10.1007/s00405-019-05631-1
21. Steward DL, Wilson KM, Kelly DH, et al. Proton pump inhibitor therapy for chronic laryngo-pharyngitis: a randomized placebo-control trial. *Otolaryngol Head Neck Surg*. 2004;131:342-350.
22. Vaezi MF, Richter JE, Stasney CR, et al. Treatment of chronic posterior laryngitis with esomeprazole. *Laryngoscope*. 2006;116:254-260.
23. Hicks DM, Ours TM, Abelson TI, Vaezi MF, Richter JE. The prevalence of hypopharynx findings associated with gastroesophageal reflux in normal volunteers. *J Voice*. 2002;16:564-579.
24. Beaver ME, Stasney CR, Weitzel E, et al. Diagnosis of laryngopharyngeal reflux disease with digital imaging. *Otolaryngol Head Neck Surg*. 2003;128:103–108.
25. Williams RB, Szczesniak MM, Maclean JC, Brake HM, Cole IE, Cook IJ. Predictors of outcome in an open label, therapeutic trial of high-dose omeprazole in laryngitis. *Am J Gastroenterol*. 2004;99:777–785.
26. Ozturan O, Dogan R, Yenigun A, Veyseller B, Yildirim YS. Photographic objective alterations for laryngopharyngeal reflux diagnosis. *J Voice*. 2017;31(1):78-85.
27. Park W, Hicks DM, Khandwala F, et al. Laryngopharyngeal reflux: prospective cohort study evaluating optimal dose of proton-pump inhibitor therapy and pretherapy predictors of response. *Laryngoscope*. 2005;115(7):1230-1238.
28. Swoger J, Ponsky J, Hicks DM, et al. Surgical fundoplication in laryngopharyngeal reflux unresponsive to aggressive acid suppression: a controlled study. *Clin Gastroenterol Hepatol*. 2006;4(4):433-441.
29. DelGaudio JM, Naseri I, Wise JC. Proximal pharyngeal reflux correlates with increasing severity of lingual tonsil hypertrophy. *Otolaryngol Head Neck Surg*. 2008;138(4):473-478.
30. Branski RC, Bhattacharyya N, Shapiro J. The reliability of the assessment of endoscopic laryngeal findings associated with laryngopharyngeal reflux disease. *Laryngoscope*. 2002;112:1019–1024.
31. Milstein CF, Charbel S, Hicks DM, Abelson TI, Richter JE, Vaezi MF. Prevalence of laryngeal irritation signs associated with reflux in asymptomatic volunteers: impact of endoscopic technique (rigid vs. flexible laryngoscope). *Laryngoscope*. 2005;115:2256–2261.
32. Chang BA, MacNeil SD, Morrison MD, Lee PK. The reliability of the reflux finding score among general otolaryngologists. *J Voice*. 2015;29(5):572-577.