

Laryngopharyngeal Reflux: Inter-rater Reliability of Reflux Finding Score in Clinical Practice

Venkatesha Belur Keshavamurthy,¹ Yogeeshha Beesanahalli Sadashivappa,² Nithin K Harsham²

ABSTRACT

Introduction

Laryngopharyngeal reflux (LPR) is a clinical entity, caused due to retrograde flow of gastric content into the larynx. It imposes great diagnostic challenge to clinicians because of lack of specific symptoms, signs and validated tool. Reflux Finding Score (RFS) has been proposed for clinical diagnosis of LPR. Inter observer reliability of RFS tool was found out using two independent observers.

Materials and Methods

In this Cross sectional study, ninety LPR subjects were subjected to 90° rigid endoscopic examination of larynx and RFS scores were calculated by two observers. Inter observer reliability of RFS was calculated using Kappa value.

Results

Most commonly observed laryngeal findings were erythema/hyperemia, posterior commissure hypertrophy, and vocal fold edema. Subglottic edema, thick endolaryngeal mucus and erythema had fair agreement.

Conclusion

There was a poor agreement between observers for total RFS score. The variables like subglottic edema, erythema/hyperemia, and thick endolaryngeal mucus showed fair agreement between the raters. The findings of diffuse laryngeal edema and posterior commissure hypertrophy showed poor inter rater reliability.

Keywords

Laryngopharyngeal Reflux; Reflux Finding Score; Endoscopy

Laryngopharyngeal reflux (LPR) is a recently described clinical entity which is caused due to the retrograde flow of gastric contents into the throat, i.e. into laryngopharynx.¹ The importance of LPR has been increasing in clinical practice as it is implicated as the etiological agent in a variety of conditions like reflux laryngitis to laryngeal carcinoma. It has been estimated that up to 10% of patients presenting to an otolaryngologist's office is LPR and 50% of all patients suffering from hoarseness and voice

disorder may have significant LPR.² Patients with LPR usually present with non specific symptoms like globus sensation, vocal fatigue, hoarseness, frequent throat clearing, dysphagia and chronic cough. LPR associated laryngoscopic findings include inter arytenoid erythema, infraglottic edema, ventricular obliteration, posterior commissure hypertrophy, granuloma / granulation and thick endolaryngeal mucus. As the symptoms and clinical signs attributed to LPR are non-specific and there are no validated diagnostic tools for LPR, the diagnosis of this condition imposes a great challenge to otorhinolaryngologist.

Belafsky developed a diagnostic tool called Reflux Finding Score (RFS) based on the laryngoscopic findings in patients with suspected LPR to ease the clinical diagnosis.³ It is an 8 item clinical severity score. Score ranges from 0 (no abnormal findings) to maximum of 26 (worst score) depending up on the presence and severity

1 - Department of ENT, S.S. Institute of Medical Sciences & Research Centre, Davangere, Karnataka

2 - S.S. Institute of Medical Sciences & Research Centre, Davangere, Karnataka

Corresponding author:

Dr Venkatesha Belur Keshavamurthy
email: bk.venkatesha@gmail.com

of the laryngoscopic findings. As the laryngoscopic findings are subjective in interpretation, the RFS score can vary from evaluator to evaluator for a single patient. So, this study makes an attempt to find out the inter-rater reliability of RFS tool in diagnosing LPR using RFS tool.

Materials and Methods

A cross sectional observational study was conducted in a tertiary hospital in central Karnataka, for a period of two years from November 2016 to October 2018. Institutional ethics committee approval was obtained for the study. A total of 90 subjects were recruited who presented with symptoms suggestive of LPR after obtaining a valid informed written consent satisfying following inclusion and exclusion criteria.

All patients aged between 16 and 45 years, clinically diagnosed with LPR, were included in the study. People with asthma / COPD/ organic laryngeal disorders not associated with LPR and persons with the history of previous radiotherapy or head and neck surgeries or psychiatric illness were excluded from this study.

After initial clinical evaluation, all the 90 patients underwent laryngoscopic examination using a 90° rigid pharyngolaryngoscope. Laryngoscopic findings were recorded and the RFS score was rated independently for each patient by two raters by reviewing the recorded video of the laryngoscopy at different points of time (with more than 10 years of Otorhinolaryngological practice)

The RFS is an 8 item scoring system based on the presence and the severity of the laryngoscopic findings. If there was no abnormal findings the score will be zero and the maximum score of 26 was given to the severe findings. A score of more than 7 was diagnostic of LPR. (Table I)

Inter rater reliability among the raters were calculated using kappa statistics for the individual variables of RFS as well as the total RFS score. Kappa is a statistical method which measures inter-rater agreement for categorical items. As it takes into account the possibility of agreement occurring by chance, it is thought to be more robust measure than simple percent agreement

calculation.

The kappa value ranges from -1 to +1. When the observed agreement is perfect, kappa will be +1. If the observed agreement equals the chance expected agreement, kappa will be 0. If the observed agreement is less than the chance expected agreement, kappa will become negative. (Table II)

Results

Age of the study group ranged from 20 to 45 years (Mean age 36.1). Majority of the subjects were between 40-45 years. There were 49 males (54.4%) and 41 females (45.6%) in the study group. Most common laryngoscopic findings (Fig. 1) in the study group were erythema/ hyperemia and posterior commissure hypertrophy 89 (98.9%), vocal fold edema 67(74.4%), ventricular obliteration 21(23.3%), thick endolaryngeal mucus 21(23.3%), diffuse laryngeal edema 17(18.9%), subglottic edema 13(14.4%), granuloma/granulation 11(12.2%). (Fig. 2)

The RFS score ranged from 2.5 to 14 (mean = 6.01 ± 2.14). The mean RFS score by the first rater was 6.31 and the mean RFS score by the second rater was 5.72.

Inter-rater reliability was assessed by using kappa value for each variable of RFS as well as for total RFS score. The kappa value for total RFS score was 0.152, indicating only a slight agreement that exists between the two raters. There was a fair agreement between the raters for the variables of subglottic edema, erythema /hyperemia (Fig. 3) and thick endolaryngeal mucus. (kappa value 0.371, 0.230, 0.323 respectively). Findings of diffuse laryngeal edema and posterior commissure hypertrophy showed poor agreement between raters. (kappa value -0.024 and -0.032 respectively). None of the variables in RFS showed perfect agreement between the two raters. (Table III)

There were 14 patients in the study group whose RFS score showed difference of 5 and more between the raters. (Table IV) Among the individual variables of RFS in these patients, ventricular obliteration, vocal fold edema (Fig. 4), diffuse laryngeal edema and posterior commissure hypertrophy showed maximum variability between the raters.

Table I. Showing Reflux Finding Score (RFS)

Subglottic edema (pseudo sulcus)	0 = absent
	2 = present
Ventricular obliteration	0 = absent
	2 = partial
	4 = complete
Erythema /hyperemia	0 = absent
	2 = only in the arytenoids
	4 = diffuse
Vocal fold edema	0 = absent
	1 = mild
	2 = moderate
	3 = diffuse
	4 = polypoidal
Diffuse laryngeal edema	0 = absent
	1 = mild
	2 = moderate
	3 = severe
	4 = obstruction
Posterior commissure hypertrophy	0 = absent
	1 = mild
	2 = moderate
	3 = severe
	4 = obstruction
Granuloma /granulation tissue	0 = absent
	2 = present
Thick endolaryngeal mucus	0 = absent
	2 = present
Total	

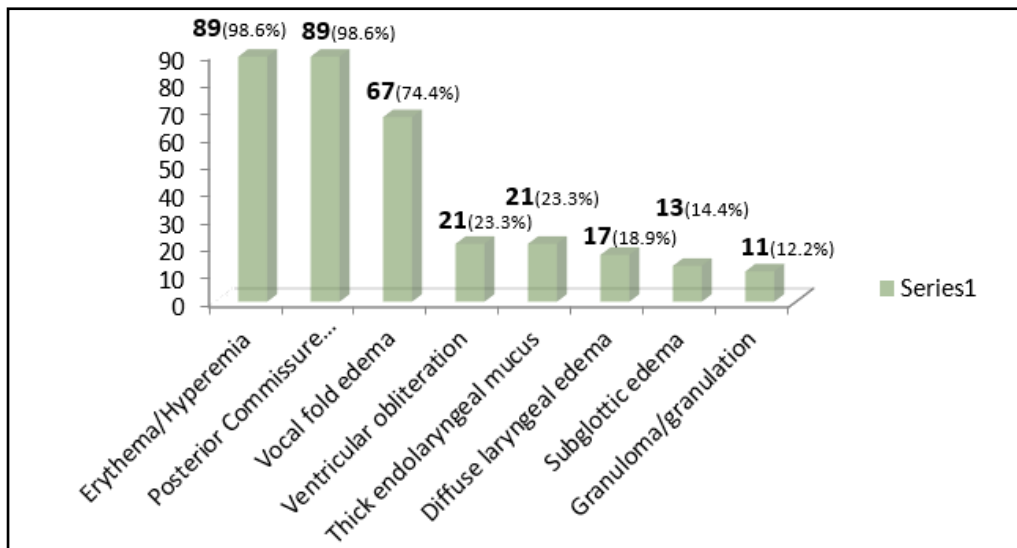


Fig. 1. Distribution of laryngeal signs

Discussion

Various laryngeal signs attributed to LPR are erythema or edema of the posterior one-third of the glottis, hyperemia of the posterior larynx, cobble stoning, and “heaping up” or thickening of the inter arytenoid mucosa (pachydermialaryngis), but none of these signs are specific and pathognomonic to this condition and hence there is a diagnostic difficulty. Reflux Finding Score (RFS) was developed by Belafsky in order to facilitate

the clinical diagnosis of LPR based on the scoring of certain common laryngeal findings according to their presence and severity on endoscopy. In their study they had established validity and high inter and intra rater reliability of this tool. Ever since this tool is being clinically utilised for the diagnosis of LPR controversies exist on the reliability of RFS variables. Most of the variables in RFS tool are subjective in interpretation.

In the Study conducted by Mesallam et al, RFS score was rated by four raters demonstrated high inter



Fig. 2. Laryngoscopic image in which there is high agreement among the raters for the findings of posterior commissure hypertrophy and granulation tissue

Table II. Showing interpretation of kappa value⁴

KAPPA VALUE	INTERPRETATION
<0	Poor agreement
0.0 – 0.2	Slight agreement
0.21- 0.4	Fair agreement
0.4 – 0.6	Moderate agreement
0.61- 0.8	Substantial agreement
0.81- 1.0	Perfect agreement

Table III: showing inter-rater reliability of RFS

RFS PARAMETERS	KAPPA VALUE
Sub glottis edema	0.371
Ventricular obliteration	0.098
Erythema / hyperemia	0.230
Vocal fold edema	0.112
Diffuse laryngeal edema	-0.024
Posterior commissure hypertrophy	-0.032
Granuloma / granulation tissue	0.145
Thick endolaryngeal mucus	0.323
Average kappa (\pm SD)	0.152 \pm 0.147

rater and intra rater reliability.⁵ Karakaya et al also demonstrated high intra rater and inter rater reliability for RFS score; findings of Vocal fold edema and the thick endolaryngeal mucus showed highest intra rater and inter rater agreement.⁶

Study by Branski et al noted relatively poor inter and intra rater agreement for the variables of the laryngeal findings.⁷ Only finding of edema of the musculo membranous fold exhibited fair reliability. Eren et al could not find inter observer agreement among the RFS variables.⁸ Highest inter-rater agreement was noted for the finding of thick endolaryngeal mucus. Other findings like pseudosulcus, ventricular obliteration, diffuse laryngeal edema and posterior commissure hypertrophy had low inter rater agreement.

In the present study we could not find high reliability of RFS among the raters. There was only slight agreement between the raters for total RFS score. (kappa value 0.152; SD \pm 0.147). Among the individual variables of RFS, none of the variables had perfect agreement. The variables like subglottic edema, erythema/ hyperemia, and thick endolaryngeal mucus showed fair agreement



Fig. 3. Laryngoscopic image in which there is high agreement among the raters for the findings of arytenoids erythema.

between the raters. The findings of diffuse laryngeal edema and posterior commissure hypertrophy showed poor inter-rater reliability.

Lack of prior sensitization of the observers/raters, intrinsic observer bias could have led to the poor consistency in rating. The degree and severity of findings vary from observer to observer and it depends on many factors like types of endoscopes used (rigid vs. flexible), lighting set up of the instrument, recordings

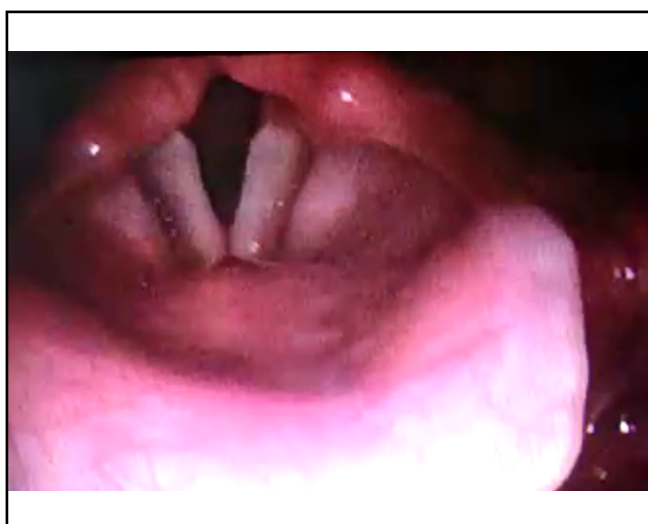


Fig. 4. Laryngoscopic imaging in which there is poor agreement between the raters for vocal fold edema.

Table IV: Showing differences in scoring among the RFS variables

SL NO.	RFS SCORE		DIFFERENCE IN THE SCORE BETWEEN THE OBSERVERS FOR THE VARIABLES									
	O1	O2	SUBGLOTTIC EDEMA	VENTRICULAR OBLITERATION	ERYTHEMA	VOCAL FOLD EDEMA	DIFFUSE LARYNGEAL EDEMA	POSTERIOR COMMISSURE HYPERTROPHY	GRANULOMA	ENDOLARYNGEAL MUCUS		
1	11	3	0	2	0	2	1	1	0	2		
2	14	5	0	4	2	1	2	0	0	0		
3	9	15	2	0	0	0	2	0	0	2		
4	4	12	0	0	2	1	2	1	2	0		
5	11	17	2	0	0	0	1	1	0	2		
6	6	17	0	2	0	2	2	1	2	2		
7	14	5	2	2	2	2	1	0	0	0		
8	8	3	0	2	0	1	0	2	0	0		
9	9	4	0	2	0	2	1	0	0	0		
10	11	4	0	4	0	2	1	0	0	0		
11	10	5	0	0	0	2	1	0	0	2		
12	10	5	0	0	2	1	1	1	0	0		
13	11	6	0	0	0	1	1	1	2	0		
14	11	4	0	0	2	1	1	1	0	2		

O1- Observer 1, O2- Observer2.

and the experience of the raters. So RFS as a clinical tool to be cautiously interpreted in the background context for inters rater reliability, and there needs to be prior sensitization of raters with a set of standardized examinations, to foster consistency and to reduce inter rater variations of laryngeal findings.

Conclusion

RFS as a clinical tool for diagnosis of LPR should be viewed with limited utility as it exhibits low inter-rater reliability. Variables in RFS are highly subjective in interpretation. However, considering few findings like thick endolaryngeal mucus which consistently show higher inter observer reliability can be given more scores in RFS scale while diagnosing LPR.

References

1. 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 Otorhinolaryngology-Head and Neck Surgery. *Otolaryngol Head Neck Surg.* 2002; 127(1):32-5
2. Datta R, Datta K, Venkatesh MD. Laryngopharyngeal reflux: Larynx on fire. *Medical Journal Armed Forces India* 2010; 66:245-8
3. Belafsky PC, Postma GN, Koufman JA. Validity and reliability of the reflux finding score (RFS). *Laryngoscope* 2001; 111: 1313-7
4. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159-74
5. Mesallam TA, Stemple JC, Sobeih TM, Elluru RG. Reflux symptom index versus reflux finding score. *Ann Otol Rhinol Laryngol.* 2007; 116: 436-40
6. Karakaya NE, Akbulut S, Altıntaş H, Demir MG et al. The reflux finding score: Reliability and correlation to the reflux symptom index. *Journal of Academic Research in Medicine* 2015; 5: 68-74
7. Branski RC, Bhattacharyya N, Shapiro J. The reliability of the assessment of endoscopic laryngeal findings associated with Laryngopharyngeal reflux disease. *Laryngoscope* 2002; 112: 1019-24
8. Eren E, Arslanoğlu S, Aktaş A, Kopar A et al. Factors confusing the diagnosis of Laryngopharyngeal reflux: the role of allergic rhinitis and inter-rater variability of laryngeal findings. *Eur Arch Otorhinolaryngol* 2014; 271: 743-7.