Published in final edited form as:

Otolaryngol Head Neck Surg. 2016 December ; 155(6): 923–935. doi:10.1177/0194599816664330.

Patient-Reported Outcome Measures Related to Laryngopharyngeal Reflux: A Systematic Review of Instrument Development and Validation

David O. Francis, MD, MS^{1,2,3}, Dhyanes A. Patel, MD^{4,5}, Rohit Sharda, MD^{4,5}, Kristen Hovis, BA⁵, Nila Sathe, MA, MLIS^{3,7}, David F. Penson, MD, MPH^{2,5,7,8,9}, Irene Feurer, PhD^{2,10}, Melissa L. McPheeters, PhD, MPH^{3,7,11}, and Michael F. Vaezi, MD, PhD, MS(epi)^{4,5}

¹Department of Otolaryngology, Bill Wilkerson Center

²Center for Surgical Quality and Outcomes Research

³Vanderbilt Evidence-based Practice Center, Vanderbilt University Medical Center

⁴Division of Gastroenterology and Hepatology

⁵Department of Medicine

⁶Vanderbilt University Medical Center; Vanderbilt University School of Medicine

⁷Department of Health Policy, Vanderbilt University Medical Center

⁸Department of Urology, Vanderbilt University Medical Center

⁹Geriatric Research and Education Coordination Center, Veteran's Administration Tennessee Valley Health System

¹⁰Departments of Surgery and Biostatistics, Vanderbilt University Medical Center

¹¹Center for Population Health Sciences, Vanderbilt University Medical Center

Abstract

Objectives—Patient-reported outcome (PRO) measures are often used to diagnose LPR and monitor treatment outcomes in clinical and research settings. The present systematic review was designed to identify currently available LPR-related PRO measures and to evaluate each measure's instrument development, validation, and applicability.

Data Sources—MEDLINE via PubMed interface, Cumulative Index of Nursing and Allied Health Literature, and Health and Psychosocial Instrument databases were searched using relevant vocabulary and key terms related to PRO measures and LPR.

Review Methods—Three investigators independently performed abstract review, full-text review, and applied a previously developed checklist to critically assess measurement properties of each study meeting inclusion criteria.

Corresponding Author: David O. Francis, MD MS, Department of Otolaryngology, Vanderbilt Bill Wilkerson Center, Medical Center East, South Tower 7302, 1215 21st Avenue South, Nashville, TN 37232, Phone: 615.343.7464, david.o.francis@vanderbilt.edu.

Disclosures and Competing Interest: None by any author

Results—Of 4,947 studies reviewed, seven LPR-related PRO measures (publication year 1991 – 2010) met criteria for extraction and analysis. Two focused on globus and throat symptoms. Remaining measures were designed to assess LPR symptoms and monitor treatment outcomes in patients. None met all checklist criteria. Only 2/7 used patient input to devise item content and 2/7 assessed responsiveness to change. Thematic deficiencies in current LPR-related measures are inadequately demonstrated: content validity, construct validity, plan for interpretation, and literacy level assessment.

Conclusion—Laryngopharyngeal reflux is often diagnosed based on symptoms. Currently available LPR-related PRO measures used to symptomatically identify suspected LPR patients have disparate developmental rigor and important methodological deficiencies. Care should be exercised to understand the measurement characteristics and contextual relevance before applying these PRO measures for clinical, research, or quality initiatives.

Keywords

Laryngopharyngeal reflux; extraesophageal reflux; measurement; patient-reported outcome measures; developmental properties; validation; systematic review

INTRODUCTION

Gastroesophageal reflux disease (GERD) affects an estimated 40% of the US population.¹ Increased attention to GERD has shifted focus to whether reflux influences other physiologic processes beyond the esophagus. Anecdotes dating back to 1960s have hypothesized connections between reflux and hoarseness,^{2,3} cancer,⁴ and airway pathologies.^{5,6} In 1991, Koufman operationalized laryngopharyngeal reflux (LPR), finding that it had a greater consequence on laryngeal function than previously considered.⁷ A distinction was made positing that LPR patients do not necessarily have classic GERD symptoms; rather a large proportion have “silent reflux” that manifests in sundry upper airway symptoms.⁸

A sensitive and specific gold standard objective diagnostic test to consistently identify patients affected by LPR remains elusive despite rapid technological advancements and vast expenditure on the subject.⁹ Efforts to correlate specific endoscopic findings and the presence of reflux have thus far met with little success.^{10–12} Despite the lack of pathognomonic signs, LPR has become a primary diagnosis and resulted in patients experiencing a barrage of medications, specialty physician visits, diagnostic tests, and surgeries.⁹

In the absence of a definitive diagnostic test, patient symptoms have become a primary method to identify those with LPR. Symptomatic differentiation of GERD and LPR should be possible if they indeed have discrete phenotypes. Therefore, patient-reported outcome (PRO) measures have become a principal means to diagnose LPR and monitor treatment outcomes. However, if PRO measures are to be used to make patient-centered, symptom-based diagnosis and treatment decisions, they must be designed with appropriate methodological rigor. Use of poorly developed measures or those intended for a different

application can have significant implications and lead to distorted, inaccurate, or equivocal findings.^{13,14}

Few clinicians or researchers have expertise in the technical methods used to develop and validate these instruments. Hence, many may presume that published PRO measures have comparably strong and appropriate measurement properties, precision, and applicability. This is problematic since nearly all published instruments purport some degree of these attributes, most often as forms of reliability or validity.¹⁵ The aim of the present study was to perform a comprehensive systematic review of the literature on LPR-related PRO measures and to rigorously evaluate each measure's developmental properties, validation, and applicability.

METHODS

This study did not involve data collection from or about human subjects and was therefore exempt from IRB approval. The systematic review methodology used herein strictly adhered to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)¹⁶

Search Strategy

An expert librarian searched MEDLINE via the PubMed interface, the Cumulative Index of Nursing and Allied Health Literature (CINAHL), and the Health and Psychosocial Instrument (HaPI) database without publication date restrictions, using relevant vocabulary terms and key terms related to PRO measures and swallowing disorders including laryngopharyngeal and extraesophageal reflux (Table 1). The search was performed in July 2015 and limited to English language publications. Reference lists of included articles and reviews related to measurement of LPR were hand-searched to identify potentially pertinent articles.

Study Selection

We developed inclusion and exclusion criteria in consultation with an expert panel that included a statistician with expertise in measurement theory, a research librarian, systematic review methodologists, and researchers and clinicians who study and treat LPR. Three investigators (DOF, RS, DAP) independently reviewed abstracts for all identified studies, and those meeting predetermined criteria (Table 2) were advanced to full text review. Abstracts that did not provide adequate information to determine eligibility were also advanced to the full text review phase. We contacted authors of articles for supporting information when necessary.

Patient Reported Outcome Measure Assessment

Three investigators independently assessed each study's methodology using a criteria checklist developed *a priori* (Table 3).¹⁷ In brief, the checklist was designed to help systematic reviewers identify components deemed important to the development of PRO measures including: 1) conceptual model, 2) content validity, 3) reliability, 4) construct validity, 5) scoring and interpretation, and 6) respondent burden and presentation. Definitions of these concepts are provided in Table 4.

Each reviewer was trained and calibrated on appropriate application of the checklist using a methodology described separately.¹⁷ They were then independently tasked with evaluating all identified PRO measures. Upon completion, reviewers met to discuss and come to consensus on scoring discrepancies. A senior investigator and psychometrician (IDF) adjudicated any remaining discrepancies.

Data Extraction

Components of PRO measure development were entered into evidence tables. These included name and acronym, authors, year(s) published, objective and intended construct, setting of development (e.g., tertiary care, community) and country, targeted population, type of scale used (e.g., visual analog scale), number of items (i.e., questions) included; the presence of proposed subscales and, when present, what each was designed to specifically measure.

Data Synthesis

Data from unique PRO measures demonstrated heterogeneity in constructs, methodology, and intended purpose and therefore were not appropriate for aggregation or meta-analysis. Instead, individual PRO characteristics, measurement properties, and functionality were summarized independently and in detail.

RESULTS

A PRISMA diagram describing study flow and inclusions is shown in Figure 1. Of 4,947 studies reviewed, seven provided initial developmental data on LPR-related PRO measures. Most excluded articles were not relevant to LPR or instrument development. The majority of measures were symptom indices related to LPR^{18–23}, while one was intended to measure health-related quality of life among those with LPR.²⁴ The earliest instruments focused on globus and throat symptoms^{18,23}, and the remainder on supra-esophageal reflux (SER)²² and LPR^{19,21,23,24} (Table 5).

All related PRO measures were developed at tertiary care academic centers, although the LPR-HRQL did also recruit patients from one regional medical center and one multi-specialty group practice (Table 6).²⁴ Country of study origin included the United States (3), Great Britain (3), and Sweden (1) with years of publication ranging from 1991 to 2010 (Table 6).^{20,23} Sample sizes used in development were extracted. These included targeted groups (e.g., LPR-attributed symptoms) and the overall populations, which incorporated both symptomatic subjects and non-affected controls. Involved participants varied from (targeted/overall) 25/50 to 279/985 patients in the Reflux Symptom Index (RSI)²¹ and Supraesophageal Reflux Questionnaire (SERQ)²², respectively. Mean or median age of patients across developmental samples ranged from 44.5 year in the Glasgow Edinburgh Throat Scale (GETS)¹⁸ to 59 years in the Pharyngeal Reflux Symptom Questionnaire (PRSQ).²³ Females were the majority in five of six remaining study samples (range 44–76%, Table 6). The study describing validation of the SERQ did not describe its cohort's age or gender distribution.²²

Question Content

Compilation of questions from the various PRO measures revealed similarities. Figure 2 shows categorized overall and measure-specific distribution of questions. Proportion of questions within each category differed between instruments. An overall mean was used to equally weight each measure as their number of items widely differed (range 9 – 43). Overall percentages for question categories in these PRO measures are shown in the left column of Figure 2. In order of frequency, questions were related to mucus/throat sensation (26%), throat clearing/cough (20%), swallowing (15%), voice (13%), reflux (13%), breathing (6%), other symptoms (4%) and general quality of life (3%). Instrument-specific distribution of questions by category is shown in the remainder of Figure 2 columns.

Some measures directly incorporated questions from others. For example, the GETS included six *verbatim* questions from Throat Questionnaire (TQ).²⁰ The 34-item Laryngopharyngeal Reflux questionnaire (LPR-34)¹⁹ combined questions from the Reflux Symptom Index (RSI)²¹ and Gastroesophageal Symptom Assessment Scale (GSAS; GERD-related questions).²⁵ It did “unbundle” thereby simplifying multi-barreled items found in the RSI into unique questions. A paraphrased example of a multi-barreled question from the RSI that contains four different processes into a single question is: “*within the last month, how did heartburn, chest pain, indigestion, or stomach acid coming up, affect you?*”

Developmental Characteristics

The developmental process and demonstrated measurement properties varied among identified PRO measures. Four instruments met at least one criterion from each of the seven domains (RSI, Pharyngeal Reflux Symptom Questionnaire [PRSQ], GETS, and Supraesophageal Reflux Questionnaire [SERQ]).^{18,21–23} Of available measures, the PRSQ met the most criteria (16/18) followed by the RSI (13/18) and LPR-HRQL (13/18). None met all assessed criteria. Analyses based on domain are outlined below and summarized in Figure 3.

Conceptual model—Each identified PRO measure was judged to have defined both the construct it intended to measure and its respective target population. Six of seven pre-specified their expected dimensionality (e.g., subscales) within the intended conceptual framework (Figure 3).

Content validity—Two PRO measures based question content on direct patient experience and symptoms derived from prospectively collected interviews or focus groups (Laryngopharyngeal Reflux-Health Related Quality of Life questionnaire [LPR-HRQL], PRSQ).^{23,24} Question content for the remainder was based solely on the opinion of content experts (e.g., otolaryngologists, laryngologists, gastroenterologists, speech-language pathologists) who care for and study this population. In fact, content experts were ubiquitously involved in questionnaire content development. Four of seven measures provided some description of the origin and rationale for selection of the final set of questions (LPR-HRQL, LPR-34, PRSQ, SERQ).^{19,22–24}

Reliability—Six LPR-related PRO measures both tested and demonstrated adequate reliability (i.e., correlation coefficient $r \geq 0.70$ or justified). Different types of reliability were evaluated. The RSI, SERQ, and Throat Questionnaire (TQ) assessed test-retest reliability, the PRSQ and GETS computed internal consistency reliability and LPR-HRQL appraised both properties. The LPR-34 questionnaire did not evaluate reliability.

Construct Validity—Most LPR-related PRO measures assessed some aspect of construct validity (GETS, RSI, LPR-HRQL, PRSQ, SERQ).^{18,21–24} Three statistically justified their dimensionality; that is, whether multiple subscales or a single scale (i.e., common factor) existed.^{18,23,24} For example, PRSQ performed exploratory factor analysis to identify and confirm its four domains structure related to LPR: cough, voice, dysphagia, and reflux. Four PRO measures established convergent validity by showing expected associations with other questionnaires focused on a similar construct (GETS, RSI, LPR-HRQL, PRSQ).^{18,21,23,24} None tested associations with clinical correlates.

All instruments had a stated goal tracking outcomes longitudinally, but only RSI²¹ and LPR-HRQL²⁴ met the criterion of showing responsiveness to change. The RSI evaluated responsiveness by treating the 25 included subjects with LPR with a 6-month course of proton pump inhibitors. Each subject was re-administered the RSI and scores were shown to significantly improve (RSI score 20.9 to 12.8) from pre- to post-treatment. Responsiveness was calculated for the LPR-HRQL by assessing the change from baseline to 4- and 6-month PPI treatment endpoints. An *a priori* defined minimally important improvement in HRQL was achieved for all LPR-HRQL domains at both the time-points.

The RSI and PRSQ demonstrated known-group validity. For example, the RSI compared scores of 25 patients with symptoms attributed to LPR to 25 age- and gender-matched patients without this condition and found a significant difference in scores. A variant of criterion validity was assessed by SERQ. Authors designated the treating physician's overall impression of SER as the "gold standard," and patients were thus categorized by physician impression into three groups: 1) symptoms likely caused by SER, 2) symptoms not caused by SER, and 3) uncertain role of SER. Receiver operating characteristic curves adjusted for chronic sinusitis and over-the-counter medication, yielding an area under the curve (AUC) of 0.72. This AUC implies that SERQ has "fair" ability to discriminate patients that physicians believed to have SER from those they did not.

Interpretation & Scoring—Scoring approaches differed among PRO measures. Some used a simple summation with higher total scores indicating higher degrees of the construct being measured.^{18,19,23} However, most measures were not designed to yield an overall total score, but rather to provide discrete, domain-specific scoring (TQ, SERQ, PRSQ, LPR-HRQL).^{20,22–24}

Regarding missing data, only PRSQ described a plan for managing incomplete questionnaires. In cases of PRSQ domains with missing items, non-missing items within that given domain were rescaled to generate a value comparable to subjects responding to all items. If more than 50% of items within the domain were absent, the domain score was also set as missing.

Three LPR-related measures provided some description of score scaling.^{21,23,24} A minimum clinically meaningful change was calculated for all domains within the LPR-HRQL. Interestingly, this clinically important change was defined as that change in each domain that corresponded with a significant (1-point) decrease in the physician-reported symptom severity score. Thus, the minimally important change was determined using physician perception of severity not necessarily by what mattered to the affected patients (e.g., patient important difference). In contrast, both the RSI and PRSQ used the same cut-off (RSI=13) for what an abnormal score represents, thus dichotomizing normal versus abnormal. In neither case, was scaling further delineated (e.g., anchors, clinically important change, minimally important change).

Burden & Presentation—Five PRO measures were considered to have reasonable degree of burden to the patient and administratively (TQ, GETS, RSI, PRSQ, SERQ).^{18,20,21,23} Two instruments were considered overly burdensome based on number of questions (LPR-HRQL, LPR-34).^{19,24} No LPR-related measure offered an estimation of the literacy level needed for its comprehension or completion. Among identified instruments, three did not provide a readily accessible method of viewing the complete sets of included questions (LPR-HRQL, LPR-34, SERQ).^{19,22,24}

DISCUSSION

With growing emphasis on patient-centered outcomes and comparative effectiveness research, PRO measures have become a predominant method to systematically collect patient-centered data, monitor treatment outcomes, and direct clinical decision-making. These instruments can be designed to quantify phenomena that lack a clear criterion. This is relevant for LPR where lack of a definitive diagnostic test has led to diagnostic uncertainty and controversy. Practitioners and researchers have proposed various approaches to diagnose LPR objectively, but each has been found inconsistent with suboptimal diagnostic sensitivity and specificity.^{26–28} Consequently, most practicing clinicians rely on patient history and symptomatology to diagnose LPR and monitor treatment outcomes.²⁶ This is often quantified in both clinical and research settings using PRO measures.

The appropriateness of using PRO measures for these purposes is predicated on their intent and developmental measurement properties. The present study systematically reviewed the literature on LPR-related PRO measure in order to assess their developmental characteristics, validation, and applicability. All seven measures identified were designed to measure throat-related symptoms attributed to LPR, but had disparate developmental rigor. The range of target individuals involved in development and/or validation of these measures varied significantly from 25 to over 273.^{21,22} It is generally recommended that variable and subject sampling are optimized for factor/principal components analysis-based methods and/or that there be greater than 100 participants involved in validation.^{29,30} Four studies achieved this standard (TQ, SERQ, PRSQ, LPR-HRQL).^{18,22–24} Adequacy, applicability, and generalizability of measures that include few individuals from the target population in development should be questioned.

The development of LPR-related PRO measures has been hindered by several factors including 1) non-specificity of LPR-symptoms, 2) lack of a sensitive and specific diagnostic test for LPR, and 3) lack of laryngeal finding specificity. Non-specificity of LPR-symptoms has resulted, in part, in diversity of question categories among instruments, which parallels the contemporary clinical picture of patients presumed to have LPR (e.g., mucus/throat sensation, throat clearing/cough, dysphagia, dysphonia, reflux, dyspnea, general quality of life). This highlights the lack of a single, clear, pathognomonic symptom(s) of this condition. Specificity of these PRO measures has been challenged. Recent studies have shown significant overlap between RSI scores suggestive of LPR and other non-reflux related throat conditions.^{31,32} One found that patients with glottic insufficiency had pathologically elevated RSI scores, which normalized after its surgical correction with injection augmentation.³¹

Lack of sensitivity and specificity of objective diagnostic tests for LPR has also affected how current measures defined their target populations. Most studies enrolled participants presented with throat or LPR-associated symptoms to an outpatient clinic^{18–20,22} or for pH monitoring,²³ without physiologic testing. In contrast, the RSI used ambulatory, 24-hour, double-probe pH monitoring to confirm its LPR diagnoses.^{7,33} While such physiologic testing is widely used to assess the presence of acid in the esophagus or hypopharynx of patients with LPR-related symptoms, its role in causally associating patients' symptoms to reflux remains controversial, with 70% and 50% sensitivities of distal and proximal probes, respectively.^{27,34}

One PRO measure, the PRSQ, used laryngeal findings via the Reflux Finding Score (RFS) to define the target population.³⁵ The specificity and reliability of this confirmatory test and laryngeal findings of LPR in general, have been scrutinized and challenged by several studies.^{12,36–41} Since its validation,³⁵ documented inter-rater agreement has ranged from poor to fair.^{12,37,39,41,42} Moreover, the majority of asymptomatic controls have signs considered consistent with LPR.^{10,12,43} Therefore, these signs may represent a tissue continuum that can be confounded by other processes (e.g., allergic rhinitis^{41,44}, type of scope^{10,45}, prior knowledge of patient's symptoms³⁷) rather than distinct pathology.^{38,43}

Our analysis focused on identifying and evaluating PRO measure developmental characteristics and functionality. All published measures purported reliability and/or validity. This simple statement is often considered sufficient legitimization of a PRO measure's quality by end-users. It is important to recognize that reliability and validity are not discrete concepts and exist on a spectrum.⁴⁶ Only the RSI met at least one criterion in each developmental domain.^{18,21–23} However, none met all criteria. Reasons for deficiencies are multifactorial.

Overall, patient-centeredness was lacking in LPR-related PRO measure development. Only the LPR-HRQL and PRSQ directly engaged patients in developing question content, despite all measures claiming to be patient-centric. The foundation for PRO measures is the target population perspective and experience. Thus, omitting patients at this stage compromises the content validity and fidelity of scores and creates a condition in which patients answer

questions designed by and based on the experience and opinions of content experts who do not live with their particular condition.

Strength of construct validity varied among instruments. Less than half of measures empirically justified their proposed subscales. This is a critical omission particularly when claiming that subscales are measuring discrete aspects of the overarching construct (e.g., LPR). Such domains within LPR may at face value appear to be discrete, and many instruments instruct that subscales be scored separately. However, statistical justification is lacking to ensure their independence and that they are not measuring overlapping aspects of the construct.

While the ability to monitor change in LPR-symptoms or quality of life is an attribute most measures espouse, few instruments adequately demonstrated responsiveness to change. In this analysis, only two instruments met this criterion (RSI, LPR-HRQL).^{21,24} It is possible that other may demonstrate responsiveness to change in future studies. Nonetheless, based on initial development, most PRO measures may not be appropriate for and could give spurious results in clinical trials and other comparative effectiveness studies.

A minority of measures provided information on how to interpret the score.^{21,23,24} Interpreting the meaning of scores is a common problem faced by end-users and hinders the clinical utility of these instruments. The RSI and PRSQ provided a cut-off for abnormal, thus allowing dichotomization, but not gradation of scores for scaling. The clinically important change calculated for each domain within the LPR-HRQL is a valuable feature for end-users who otherwise would have to evaluate for statistical differences in scores, which may lack clinical significance.⁴⁷ However, even this important change was based on physician perception of severity rather than what mattered to the patient. Several strategies to determine a clinically important change exist.⁴⁷ Omitting this feature of interpretability represents a weakness in most LPR-related PRO measures and limits their usefulness in clinical and research applications.

Incomplete questionnaires are also common occurrences in both clinical practice and research applications. Implications of missing data can be significant, particularly if missing systematically, thus introducing bias. Many techniques for dealing with incomplete data exist, but only the PRSQ provided a framework. Additionally, no instrument described its literacy level. Respondents who do not understand the questions are less likely to complete them, which can result in bias based on patient educational level.

At present, a sensitive and specific criterion objective diagnostic test for LPR is lacking, therefore PRO measures are often used in the diagnostic process for these patients. Available PRO measures have important developmental and psychometric limitations. Care should be exercised to understand the developmental characteristics of PRO measures before selecting and advocating their use in research or clinical applications. Measures vary in their developmental strengths and some might be helpful adjuncts to diagnostic testing in outcome assessment. Importantly, of the seven measures identified, only the RSI and the LPR-HRQL measured aspects of responsiveness to change and could therefore be appropriately used to track clinical outcomes over time.

There are limitations to this review process. Despite the careful design, the search may not have captured all available literature, as it is poorly indexed. Hand-searches were used to mitigate this limitation. We also limited our search to English publications only. Applicable PRO measures may have been published in other languages that this review did not capture. There is also the risk of subjectivity in scoring PRO measure characteristics. Every effort was made to minimize this risk by using three independent reviewers for each instrument considered.

CONCLUSIONS

Patient-reported outcome measures are currently a principle method of diagnosing LPR and monitoring effectiveness of targeted therapies. Despite their prominence, available PRO measures were constructed with disparate developmental rigor. Important thematic deficiencies among instruments include a lack of patient involvement in the item development process, variable construct validity (e.g., responsiveness), and poor interpretability and scaling properties. However, several had important strengths and could be considered as adjunctive outcomes in the assessment of patients with symptoms attributed to or who have been diagnosed with LPR. Care must be taken to understand the developmental characteristics of PRO measures before selecting, advocating for, and using them in research and clinical applications.

Acknowledgments

Funding provided by the National Institute for Deafness and Communication Disorders K23DC013559 (DOF) and L30DC012687 (DOF). This work was presented at Digestive Disease Week® 2016, May 21, 2016; San Diego, CA.

References

1. Locke GR 3rd, Talley NJ, Fett SL, Zinsmeister AR, Melton LJ 3rd. Prevalence and clinical spectrum of gastroesophageal reflux: a population-based study in Olmsted County, Minnesota. *Gastroenterology*. 1997; 112:1448–1456. [PubMed: 9136821]
2. Hallewell JD, Cole TB. Isolated head and neck symptoms due to hiatus hernia. *Arch Otolaryngol*. 1970; 92:499–501. [PubMed: 5506061]
3. Chodosh PL. Gastro-esophago-pharyngeal reflux. *Laryngoscope*. 1977; 87(9 Pt 1):1418–1427. [PubMed: 895305]
4. Jacobs A. Post-Cricoid Carcinoma: Regional Incidence in England and Wales. *Br Med J*. 1963; 2:1373–1375. [PubMed: 14063028]
5. Delahunty JE, Cherry J. Experimentally produced vocal cord granulomas. *Laryngoscope*. 1968; 78:1941–1947. [PubMed: 5722897]
6. Cherry J, Margulies SI. Contact ulcer of the larynx. *Laryngoscope*. 1968; 78:1937–1940. [PubMed: 5722896]
7. 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.
8. Koufman JA. Laryngopharyngeal reflux is different from classic gastroesophageal reflux disease. *Ear Nose Throat J*. 2002; 81(9 Suppl 2):7–9. [PubMed: 12353431]
9. Francis DO, Rymer JA, Slaughter JC, et al. High economic burden of caring for patients with suspected extraesophageal reflux. *Am J Gastroenterol*. 2013; 108:905–911. [PubMed: 23545710]

10. 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. [PubMed: 16369176]
11. Qadeer MA, Swoger J, Milstein C, et al. Correlation between symptoms and laryngeal signs in laryngopharyngeal reflux. *Laryngoscope*. 2005; 115:1947–1952. [PubMed: 16319603]
12. Powell J, Cocks HC. Mucosal changes in laryngopharyngeal reflux--prevalence, sensitivity, specificity and assessment. *Laryngoscope*. 2013; 123:985–991. [PubMed: 23208751]
13. Patrick DL, Burke LB, Powers JH, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value Health*. 2007; 10(Suppl 2):S125–137. [PubMed: 17995471]
14. Regnault A, Hamel JF, Patrick DL. Pooling of cross-cultural PRO data in multinational clinical trials: how much can poor measurement affect statistical power? *Quality Life Res*. 2015; 24:273–277.
15. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality Life Res*. 2012; 21:651–657.
16. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ*. 2015; 349:g7647. [PubMed: 25555855]
17. Francis DO, McPheeters ML, Noud M, Penson DF, Feurer ID. Checklist to Operationalize Measurement Characteristics of Patient-Reported Outcome Measures. *Systematic Reviews*. In Press.
18. Deary IJ, Wilson JA, Harris MB, MacDougall G. Globus pharyngis: development of a symptom assessment scale. *J Psychosom Res*. 1995; 39:203–213. [PubMed: 7595878]
19. Papakonstantinou L, Leslie P, Gray J, Chadwick T, Hudson M, Wilson JA. Laryngopharyngeal reflux: a prospective analysis of a 34 item symptom questionnaire. *Clin Otolaryngol*. 2009; 34:455–459. [PubMed: 19793278]
20. Wilson JA, Deary IJ, Maran AG. The persistence of symptoms in patients with globus pharyngis. *Clinical otolaryngology and allied sciences*. 1991; 16:202–205. [PubMed: 2070540]
21. Belafsky PC, Postma GN, Koufman JA. Validity and reliability of the reflux symptom index (RSI). *J Voice*. 2002; 16:274–277. [PubMed: 12150380]
22. Dauer E, Thompson D, Zinsmeister AR, et al. Supraesophageal reflux: validation of a symptom questionnaire. *Otolaryngol Head Neck Surg*. 2006; 134:73–80. [PubMed: 16399184]
23. Andersson O, Ryden A, Ruth M, Moller RY, Finizia C. Development and validation of a laryngopharyngeal reflux questionnaire, the Pharyngeal Reflux Symptom Questionnaire. *Scand Journal Gastroenterol*. 2010; 45:147–159.
24. Carrau RL, Khidr A, Gold KF, et al. Validation of a quality-of-life instrument for laryngopharyngeal reflux. *Arch Otolaryngol Head Neck Surg*. 2005; 131:315–320. [PubMed: 15837899]
25. Rothman M, Farup C, Stewart W, Helbers L, Zeldis J. Symptoms associated with gastroesophageal reflux disease: development of a questionnaire for use in clinical trials. *Dig Dis Sci*. 2001; 46:1540–1549. [PubMed: 11478508]
26. Vaezi MF, Hicks DM, Abelson TI, Richter JE. Laryngeal signs and symptoms and gastroesophageal reflux disease (GERD): a critical assessment of cause and effect association. *Clin Gastroenterol Hepatol*. 2003; 1:333–344. [PubMed: 15017651]
27. Vaezi MF, Schroeder PL, Richter JE. Reproducibility of proximal probe pH parameters in 24-hour ambulatory esophageal pH monitoring. *Am J Gastroenterol*. 1997; 92:825–829. [PubMed: 9149194]
28. Shaker R, Bardan E, Gu C, Kern M, Torrico L, Toohill R. Intraparyngeal distribution of gastric acid refluxate. *Laryngoscope*. 2003; 113:1182–1191. [PubMed: 12838017]
29. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007; 60:34–42. [PubMed: 17161752]
30. Velicer WF, Fava JL. Effects of variable and subject sampling on factor pattern recovery. *Psychol Methods*. 1998; 3:231–251.

31. Patel AK, Mildenhall NR, Kim W, Carroll TL. Symptom overlap between laryngopharyngeal reflux and glottic insufficiency in vocal fold atrophy patients. *Ann Otol Rhinol Laryngol*. 2014; 123:265–270. [PubMed: 24671482]
32. Rafii B, Talierecio S, Achlatis S, Ruiz R, Amin MR, Branski RC. Incidence of underlying laryngeal pathology in patients initially diagnosed with laryngopharyngeal reflux. *Laryngoscope*. 2014; 124:1420–1424. [PubMed: 24155128]
33. Postma GN. Ambulatory pH monitoring methodology. *Ann Otol Rhin Laryngol Suppl*. 2000; 184:10–14.
34. Vaezi MF. Review article: the role of pH monitoring in extraoesophageal gastro-oesophageal reflux disease. *Aliment Pharmacol Ther*. 2006; 23(Suppl 1):40–49. [PubMed: 16483269]
35. Belafsky PC, Postma GN, Koufman JA. The validity and reliability of the reflux finding score (RFS). *Laryngoscope*. 2001; 111:1313–1317. [PubMed: 11568561]
36. Oelschlager BK, Eubanks TR, Maronian N, et al. Laryngoscopy and pharyngeal pH are complementary in the diagnosis of gastroesophageal-laryngeal reflux. *J Gastrointest Surgery*. 2002; 6:189–194.
37. Chang BA, MacNeil SD, Morrison MD, Lee PK. The Reliability of the Reflux Finding Score Among General Otolaryngologists. *J Voice*. 2015; 29:572–577. [PubMed: 26118936]
38. 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. [PubMed: 12160267]
39. Kelchner LN, Horne J, Lee L, et al. Reliability of speech-language pathologist and otolaryngologist ratings of laryngeal signs of reflux in an asymptomatic population using the reflux finding score. *J Voice*. 2007; 21:92–100. [PubMed: 16546351]
40. Musser J, Kelchner L, Neils-Strunjas J, Montrose M. A comparison of rating scales used in the diagnosis of extraesophageal reflux. *J Voice*. 2011; 25:293–300. [PubMed: 20202786]
41. Eren E, Arslanoglu S, Aktas 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; 271:743–747. [PubMed: 23999594]
42. Mesallam TA, Stemple JC, Sobeih TM, Elluru RG. Reflux symptom index versus reflux finding score. *Ann Otol Rhinol Laryngol*. 2007; 116:436–440. [PubMed: 17672246]
43. 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. [PubMed: 12512644]
44. Randhawa PS, Mansuri S, Rubin JS. Is dysphonia due to allergic laryngitis being misdiagnosed as laryngopharyngeal reflux? *Logopedics, phoniatrics, vocology*. 2010; 35:1–5.
45. Eller R, Ginsburg M, Lurie D, Heman-Ackah Y, Lyons K, Sataloff R. Flexible laryngoscopy: a comparison of fiber optic and distal chip technologies-part 2: laryngopharyngeal reflux. *J Voice*. 2009; 23:389–395. [PubMed: 19185459]
46. Newton PE, Shaw SD. Standards for Talking and Thinking About Validity. *Psychol Methods*. 2013; 18:301–319. [PubMed: 23834419]
47. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Clinical Significance Consensus Meeting G. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc*. 2002; 77:371–383. [PubMed: 11936935]

Figure 1.
Disposition of studies identified for this review

*Numbers do not tally as studies could be excluded for multiple reasons

Figure 2.
Overall and patient-reported outcome measure symptom- and domain-specific distribution of questions

PRO measure	Conceptual Model			Content Validity			Reliability		Construct Validity				Interpretation & Scoring			Burden & Presentation		
	Construct defined	Target population defined	Expected subscales described	Patient devised items	Content experts involved	Description of item development	Reliability tested	coefficient adequate	Justification of subscales	Longitudinal validity	Convergent validity	Known-group validity	Plan for scoring measure	Plan for missing data	Scaling described	Length reasonable	Literacy level	Items viewable
TQ																		
GETS																		
RSI																		
LPR-HRQL																		
LPR-34																		
PRSQ																		
SERQ																		

TQ: Throat Questionnaire; GETS: Glasgow Edinburgh Throat Scale; RSI: Reflux Symptom Index; LPR-HRQL: Laryngopharyngeal Reflux-Health-Related Quality of Life questionnaire; LPR-34: 34-item Laryngopharyngeal Reflux questionnaire; PRSQ: Pharyngeal Reflux Symptom Questionnaire; SERQ: Supraesophageal Reflux Questionnaire

Figure 3.

Summary comparison of measurement properties among identified patient-reported outcome measures (blue = criterion met)

Table 1

Keywords used in Systematic Review

Search terms	
#1	Swallowing disorder(s), disordered swallowing, swallowing difficulty, difficulty swallowing, swallowing dysfunction(s), dysfunctional swallowing, swallow function, swallowing function, swallowing preservation, swallowing therapy, deglutition disorder(s), deglutition dysfunction(s), dysphagia, dysphagic, laryngeal penetration, inefficient swallowing, ineffective swallowing, incomplete swallowing, globus, aspiration, pharyngeal diseases, pharyngeal dysfunction, esophageal diseases, esophageal motility disorders, esophageal dysfunction, regurgitation, reflux, gastroesophageal reflux, gastroesophageal reflux disease, GERD, post-nasal drainage, sinus drainage, cough, feeding difficulties, extraesophageal reflux, EER, laryngopharyngeal reflux, LPR, reflux laryngitis, swallow, swallowing, dysphagia, deglutition
#2	Psychometric(s), scale(s), score(s), inventory(s), questionnaire(s), index/indices, instrument(s), clinical assessment tools, patient satisfaction, treatment outcome(s), outcome assessment(s), outcomes, health outcomes, outcome measure(s), measurement disability, handicap, evaluation, survey(s), reproducibility, validity, validation, reliability, qualitative studies, quality of life, severity of illness, validation studies, measurement properties, developmental properties

Table 2

Inclusion Criteria for Abstract Review

Original Research (includes systematic reviews and meta-analysis but not narrative reviews)?
Research is on human subjects?
Study addresses reflux or swallowing-related concept?
Study addresses a patient-reported outcome, instrument, questionnaire, or survey?
Study addresses development, validity testing, and/or reliability testing of a patient reported outcome measure, instrument, survey, or questionnaire?
Study performed in adult population (>= 18 years of age)?

Table 3

“Checklist” of key characteristics to consider when evaluating a patient-reported outcome (PRO) measure. Instructions: Please indicate, in the “score” column, whether or not the information provided in the citation/source document meets each of the criteria (0 = criterion not met, 1 = criterion met).

CONCEPTUAL MODEL	SCORE
1. Has the PRO construct to be measured been specifically defined?	
2. Has the intended respondent population been described?	
3. Does the conceptual model address whether a single construct/scale or multiple subscales are expected?	
CONTENT VALIDITY	
4. Is there evidence that members of the intended respondent population were involved in the PRO measure's development?	
5. Is there evidence that content experts were involved in the PRO measure's development?	
6. Is there a description of the methodology by which items/questions were determined (e.g., focus groups, interviews)?	
RELIABILITY	
7. Is there evidence that the PRO measure's reliability was tested (e.g., test-retest, internal consistency)?	
8. Are reported indices of reliability adequate (e.g., ideal: $r = 0.80$; adequate: $r = 0.70$; or otherwise justified)?	
CONSTRUCT VALIDITY	
9. Is there reported quantitative justification that single scale or multiple subscales exist in the PRO measure (e.g., factor analysis, item response theory)?	
10. Is the PRO measure intended to measure change over time? If YES , is there evidence of both test-retest reliability AND responsiveness to change? Otherwise, award 1 point if there is an explicit statement that the PRO measure is NOT intended to measure change over time.	
11. Are there findings supporting expected associations with existing PRO measures or with other relevant data?	
12. Are there findings supporting expected differences in scores between relevant known groups?	
SCORING & INTERPRETATION	
13. Is there documentation how to score the PRO measure (e.g. scoring method such as summing or an algorithm)?	
14. Has a plan for managing and/or interpreting missing responses been described (i.e., how to score incomplete surveys)?	
15. Is information provided about how to interpret the PRO measure scores [e.g. scaling/anchors, (what high and low scores represent), normative data, and/or a definition of severity (mild → severe)]?	
RESPONDENT BURDEN & PRESENTATION	
16. Is the time to complete reported and reasonable? OR , if it is <u>NOT</u> reported, is the number of questions appropriate for the intended application?	
17. Is there a description of the literacy level of the PRO measure?	
18. Is the entire PRO measure available for public viewing (e.g., published with the citation, or information provided about how to access a copy)?	

Francis et al. Checklist to Operationalize Measurement Characteristics of Patient-Reported Outcome Measures. *Systematic Reviews*. In Press

Table 4**Glossary of Measurement Properties of Patient-Reported Outcome Measures**

Domain	Explanation
Conceptual Model	A conceptual model provides a rationale for and description of the concepts and target population that a measure is intended to assess.
Content Validity	Content validity refers to evidence that a PRO measure's domain(s) are appropriate for its intended use. Items and conceptual domains should be relevant to the target population's concerns. The PRO measure's development should include direct input from patients and also from content experts. There should be a clear description of the process by which included questions were derived.
Reliability	Reliability is the degree to which scores are free from random (measurement) error. Internal consistency reliability, the degree to which segments of a test (e.g., individual items) are associated with one another, reflects precision at a single time point. Test-retest reliability refers to the reproducibility of scores over two administrations, typically in close temporal proximity, among respondents who are assumed not to have changed on the relevant domains. Traditionally cited minimum levels for reliability coefficients are 0.70 for group-level comparisons and 0.90 to 0.95 for individual comparisons. Reliability estimates lower than these conventions should be justified in the context of the proposed PRO measure's intended application.
Construct Validity	Construct validity refers to whether a test measures intended theoretic constructs or traits and directly affects the appropriateness of the measurement-based inferences. Several different forms exist and are outlined below. Dimensionality: Empirical demonstration of dimensionality (e.g., factor analysis) provides evidence of whether a single scale or multiple subscales exist in the PRO measure. Responsiveness to change (longitudinal validity) is the extent to which a PRO measure detects meaningful change over time when it is known to have occurred. It is predicated on demonstration of both test-retest-reliability (stability when no change is expected) and clinically meaningful change when it is expected. Convergent validity is the degree to which a PRO measure's scores correlate with other instruments that measure the same construct or with related clinical indicators (e.g., diagnostic test). <i>A priori</i> hypotheses about expected associations between a PRO measure and similar or dissimilar measures should be documented. Known-groups validity is the degree to which a PRO measure is able to differentiate between groups that empiric evidence has shown to be different (e.g., cases and controls).
Interpretability & Scoring	Interpretability is the degree to which the meaning of the scores can be easily understood. Scoring refers to the "rules" for computing total scores or scales, if relevant. A description of how to score the measure should be provided (e.g., summation, algorithm). Missing responses are a common occurrence in both clinical and research settings and can affect an end-user's ability to interpret results. A pre-specified plan for managing missing responses can mitigate the risk of bias resulting from the necessity to exclude cases with missing data. Scaling is the process of distributing the full range of respondents' possible scores with respect to the measured attribute. A relative score then represents a subject's location in relation to others on a common scale. It allows cross-sectional and longitudinal quantification of the magnitude of the attribute that is reported and its change over time. Both cross-sectional and longitudinal changes in scores need to be contextualized to allow interpretation of their meaning. Ideally scaling should be based on an understanding of what represents a clinically important or patient-important change in the construct being measured.
Burden & Presentation	Burden refers to the time, effort, or other demands placed on respondents or those administering the instrument. This includes number and complexity of items. The literacy level needed to understand and complete the measure is another important aspect of burden. While most experts recommend literacy be at the 6 th grade reading level or lower, this criterion should be contextualized to the intended target population. Presentation refers to a questionnaire's appearance in light of its intended mode of administration. It is important that prospective users be able to preview a measure in its entirety (e.g., items and response options) to ensure its appropriateness for the intended application.

Table 5

Measurement Aims, Target Population, and Item Characteristics for Patient-Reported Outcome Measures Related to Laryngopharyngeal Reflux

Instrument	Year	Measurement Aim	Target Population	Language	Number of items/Subscales	Response options	Subscales
TQ	1991	To assess persistence of pharyngeal symptoms in globus patients	Patients with globus pharyngis	English	12 items	5-point scale (0–4)	NA
GETS	1995	To determine validity of GETS to assess and monitor the severity of common throat symptoms	Patients with globus pharyngis	English	10 items in 3 subscales	8-point scale (0–7)	<ol style="list-style-type: none"> 1 Dysphagia 2 Globus 3 Pain/Swelling
RSI	2002	To develop a reliable and validated, self-administered nine-item reflux symptom index for assessment of symptoms in patients with LPR	Patients with LPR	English	9 items	6-point scale (0–5)	NA
LPR-HRQL	2005	To evaluate reliability, validity, and responsiveness of a new LPR-specific health related quality of life instrument, designed to be self-administered by patients	Patients diagnosed with LPR with symptoms	English	43 items in 5 subscales	7- or 10-point scales (0–6 or 1–10)	<ol style="list-style-type: none"> 1 Voice/Hoarse 2 Cough 3 Clear Throat 4 Swallow 5 Overall Impact of Reflux
SERQ	2006	To develop and validate a comprehensive supraglottic reflux symptoms questionnaire that will help provide uniformity in clinical research and patient care	Patients with a primary pharyngolaryngeal complaint	English	9 items	5-point scale (1–5) for onset, frequency, and severity	NA
LPR-34	2009	To perform a preliminary evaluation as to whether any different clusters of symptoms map to specific syndromes such as globus, postnasal drip, or chronic hoarseness attributed to reflux and to assess prevalence of reflux symptoms in voice clinic patients	Patients with symptoms attributed to LPR	English	34 items	6-point scale (0–5)	NA
PRSQ	2010	To develop and validate a comprehensive LPR disease-specific, self-administered questionnaire, which includes both frequency and severity dimensions	Patients with upper airway symptoms commonly associated with LPR	English	17 items in 4 subscales	6-point scale for both frequency and severity	<ol style="list-style-type: none"> 1 Cough 2 Voice/Hoarse 3 Dysphagia 4 Reflux

TQ: Throat Questionnaire; GETS: Glasgow Endinburgh Throat Scale; RSI: Reflux Symptom Index; LPR-HRQL: Laryngopharyngeal Reflux Health-Related Quality of Life; SERQ: Supraglottic Reflux Questionnaire; LPR-34: 34-Item Symptom Questionnaire; PRSQ: Pharyngeal Reflux Symptom Questionnaire; LPR: Laryngopharyngeal Reflux

Table 6

Study Population Characteristics and Setting Involved in Development of Patient-Reported Outcome Measures Related to Laryngopharyngeal Reflux

Article	Instrument	Study Population	Setting	N	Distribution of Pathology	Mean age (SD), Range	Male (%)	Country
Wilson et al. 1991	TQ	Patients with globus sensation	Special clinic in Department of Otolaryngology, Royal Infirmary, Edinburgh	N=89	Globus pharyngis=89	48 (13)	24%	UK
Deary et al. 1995	GETS	Patients at ENT outpatient clinic with diagnosis of globus pharyngis and no structural pathology	Outpatients in Departments of Otolaryngology at Royal Infirmary, Glasgow and Royal Infirmary	N=105	Globus pharyngis=105	44.5 (10.0), 23 – 64	30.5%	UK
Belafsky et al. 2002	RSI	Patients with LPR who diagnosis was confirmed by ambulatory, 24-hr double-probe pH monitoring before and after 6 months of treatment with twice-daily PPIs and gender- & age-matched controls without symptoms or other evidence of LPR	Center for Voice Disorder of Wake Forest University, Department of Otolaryngology, Wake Forest University School of Medicine	Cases=25 Controls=25	LPR=25	Cases: 57 (17) Controls: 57 (17)	Cases: 56% Controls: 56%	US
Carrau et al. 2005	LPR-HRQL	Patients newly diagnosed with LPR (diagnosed <1 month before enrollment) or patients with relapse and not under current treatment	Six sites in 4 states in Southern, Mid-Atlantic, and Northeastern regions of the US including 4 academic centers, 1 regional medical center, and 1 multispecialty group practice	N=117	LPR=117	48.4 (14.6)	33.3%	US
Dauer et al. 2006	SERQ	Patients with a primary pharyngolaryngeal complaint (hoarseness, excessive throat clearing, globus sensation, sore throat, difficulty swallowing, dry or hacking cough, or nocturnal cough) based on chart review	Outpatient tertiary care clinics in Dept. of Otolaryngology, Division of Gastroenterology and Hepatology, and Division of Pulmonary and Critical Care Medicine, Mayo Clinic	Feasibility=48 Concurrent validity=71 Reproducibility=119 Predictive validity=224	Throat symptoms attributable to SER=237	NR	NR	US

Article	Instrument	Study Population	Setting	N	Distribution of Pathology	Mean age (SD), Range	Male (%)	Country
Papakonstantinou et al. 2009	LPR-34	Patients reporting typical throat symptoms including hoarseness, globus sensation, cervical dysphagia, non-specific cervical discomfort, and unexplained throat level cough spasms	Ear Nose and Throat Department of Freeman Hospital Newcastle upon Tyne	N=62	Throat symptoms attributed to LPR=62	Median 59, 23–96	39%	UK
Andersson et al. 2010	PRSQ	Patients with at least 1 symptom commonly associated with laryngopharyngeal reflux (e.g. hoarseness, chronic cough, globus, or chronic throat clearing) for greater than 1 month referred for pH monitoring	Department of Otolaryngology Karolinska University Hospital, Huddinge or Sahlgrenska University Hospital, Goteborg	N=228	Throat symptoms attributed to LPR=228	59 (13)	44%	Sweden

TQ: Throat Questionnaire; GETS: Glasgow Edinburgh Throat Scale; RSI: Reflux Symptom Index; LPR-HRQL: Laryngopharyngeal Reflux Health-Related Quality of Life; SERQ: Supraesophageal Reflux Questionnaire; LPR-34: 34-Item Symptom Questionnaire; PRSQ: Pharyngeal Reflux Symptom Questionnaire; ENT: Ear, Nose, and Throat; PPI: proton pump inhibitor; SER: supraesophageal reflux