VALUE IN HEALTH 18 (2015) 396–403

Available online at www.sciencedirect.com

ScienceDirect

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

Patient-Reported Outcomes

Responder Definition of a Patient-Reported Outcome Instrument for Laryngopharyngeal Reflux Based on the US FDA Guidance

Han-Chung Lien, MD, MPH[1,2,3], *Chen-Chi Wang, MD*[4,5,6], *Shou-Wu Lee, MD*[1,7], *Jeng-Yuan Hsu, MD*[8], *Hong-Zen Yeh, MD*[1,2], *Chung-Wang Ko, MD*[1,2], *Chi-Sen Chang, MD, PhD*[7,9], *Wen-Miin Liang, PhD*[10,*]

[1]*Division of Gastroenterology, Taichung Veterans General Hospital, Taichung, Taiwan;* [2]*Department of Internal Medicine, National Yang-Ming University, Taipei, Taiwan;* [3]*Department of Public Health, China Medical University, Taichung, Taiwan;* [4]*Department of Otolaryngology, Taichung Veterans General Hospital, Taichung, Taiwan;* [5]*School of Medicine, National Yang-Ming University, Taipei, Taiwan;* [6]*School of Speech Language Pathology & Audiology, Chung Shan Medical University, Taichung, Taiwan;* [7]*Department of Internal Medicine, Chung Shan Medical University, Taichung, Taiwan;* [8]*Division of Chest Medicine, Taichung Veterans General Hospital, Taichung, Taiwan;* [9]*Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan;* [10]*Graduate Institute of Biostatistics, China Medical University, Taichung, Taiwan*

ABSTRACT

Background: Different end-point measures may contribute to inconsistent therapeutic responses in relief of laryngopharyngeal reflux (LPR) symptoms. **Objectives:** We aimed to determine an a priori responder definition for a patient-reported outcome instrument, the Reflux Symptom Index (RSI), using an anchor-based method, to interpret individual treatment benefit in patients with LPR, on the basis of the US Food and Drug Administration guidance. **Methods:** Patients with chronic laryngeal symptoms suggestive of LPR underwent twice-daily 40 mg esomeprazole treatment for 12 weeks. We used a 50% or more reduction in the primary laryngeal symptom at week 12, an empirical criterion, as an anchor to dichotomize the participants into two groups, and to establish a responder definition of the RSI score change. The optimal cutoff point of the RSI score change was determined on the basis of the maximal Youden index of the receiver operating characteristic analysis. **Results:** The mean reduction in the RSI score was significantly greater in subjects with a 50% or more reduction in the primary laryngeal symptom than in those without (-11.0 ± 7.8 vs. -3.1 ± 8.3, $P < 0.0001$). A reduction of six points or more in the RSI score at week 12 was considered to be the responder definition with a sensitivity of 0.79 and a specificity of 0.70. **Conclusions:** We propose an a priori responder definition derived from an empirical criterion according to the Food and Drug Administration guidance: a reduction of six points or more in the RSI score at week 12. This preliminary estimate provides a clinically meaningful change at an individual level, although additional studies and validations across various languages are required.

Keywords: anchor-based method, laryngopharyngeal reflux, patient-reported outcome, responder definition.

Introduction

Laryngopharyngeal reflux (LPR) or reflux laryngitis is an established extraesophageal manifestation of gastroesophageal reflux disease [1,2]. Management of LPR, however, is controversial because although treatment with proton pump inhibitors (PPIs) is often recommended [3,4], the treatment efficacy remains inconsistent among controlled trials [5–7]. Inappropriate instruments and/or inconsistent end-point measures of patient-reported outcomes (PROs) may explain, at least in part, the failure to demonstrate any treatment benefit [7]. For example, none of the eight randomized placebo-controlled trials in a meta-analysis used validated disease-specific PRO instruments for demonstrating treatment benefit of PPIs [5].

The Reflux Symptom Index (RSI) is a disease-specific self-administered questionnaire used for evaluation of LPR symptom severity. It was developed and psychometrically validated in a sample of patients with LPR [8], has been used in centers worldwide for many years, and has been translated into several languages [9–12]. In addition, two recent randomized controlled trials used the RSI and found that PPIs were more effective than the placebo [6,7].

Although a disease-specific instrument can be used to measure PRO, a statistically significant score change may not be clinically relevant [13]. Subsequently, *minimally important differences*, defined as the smallest meaningful difference derived from point estimates of mean differences among groups, may mask important changes for individuals [14]. Recently, the US Food and Drug

Note. H.-C. Lien and C.-C. Wang contributed equally to this work as first authors.

* *Address correspondence to:* Wen-Miin Liang, PhD, Graduate Institute of Biostatistics, China Medical University, 91, Hsueh-Shih Road, Taichung, Taiwan 40402.

E-mail: wmliang@mail.cmu.edu.tw.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2015.01.001

Administration (FDA) released a PRO guidance [15] that recommended determining an a priori responder definition for PRO instruments, that is, the change in an individual patient's PRO score over a predetermined time period that can be considered to constitute a treatment benefit. In the guidance, anchor-based methods using an empirical criterion such as patient rating of change on exit from a clinical trial are recommended to establish a responder definition, whereas distribution-based methods are considered to play a supportive role [15]. Moreover, a proportion of treatment responders can be calculated to compare group differences in addition to assessing the conventional mean differences between groups. Therefore, this approach is advantageous in both clinical practice and clinical trials for the interpretation of individual and group treatment benefits, respectively.

Previously, de Vet et al. [16] proposed an anchor-based minimally important change distribution method to assess minimally important changes. They dichotomized the patients into "important improvement" and "no important change" according to a predefined anchor, and then determined a cutoff value on the score change of the PRO instrument on the basis of the maximal Youden index of the receiver operative characteristic (ROC) curve or the 95% limit. Their method takes advantage of both an external criterion (anchor-based approach) and a measure of variability (distribution-based approach) to determine the responder definition at the individual level, and, therefore, can be applied as an end-point measure to interpret treatment benefits.

In this article, we tried to determine an a priori responder definition of the Chinese-version RSI for patients with LPR in accordance with the FDA guidance. We first tested the reliability and responsiveness of the RSI in Taiwanese patients with LPR and assessed whether the target concept of the RSI correlates with an empirical criterion, that is, 50% or more reduction in the primary laryngeal symptom [5,17–19]. Furthermore, using this empirical criterion as the anchor, we attempted to determine a responder definition on the basis of the change in the RSI score during the PPI treatment using the anchor-based method along with the maximal Youden index approach for identifying the threshold.

Methods

This study was a prospective, open-label therapeutic trial conducted at the Voice & Laryngeal Pathology Laboratory and the Gastrointestinal Physiology & Motility Laboratory at Taichung Veterans General Hospital, Taiwan. The protocol was approved by Taichung Veterans General Hospital's Institutional Review Board (no. C06254) on January 29, 2007. All patients signed an informed consent form before the study.

Patient Selection

Patients (aged >20 years) with chronic laryngeal signs and symptoms suspected to be reflux-related and referred from the Department of Otolaryngology clinic between January 30, 2007, and December 31, 2012, were assessed for study eligibility.

Currently, LPR is frequently suspected on the basis of laryngeal symptoms and signs after excluding common etiologies other than reflux [20]. Because of the nonspecificity of laryngeal symptoms and signs, objective measures have been recommended to provide supportive evidence for diagnosis [4]. The inclusion criteria in the study were 1) presence of one or more laryngeal symptoms as the predominant symptom(s), including globus sensation, throat pain, hoarseness, cough, or throat clearing for three or more consecutive months before screening and 2) presence of at least one positive finding of the following objective tests: a) laryngoscopic signs of LPR with a Reflux Finding Score (RFS) of more than 7 [21], which is highly recommended for the diagnosis of LPR [4]; b) presence of reflux esophagitis diagnosed by upper gastrointestinal endoscopy; and c) presence of excessive acid reflux demonstrated by 24-hour ambulatory esophagopharyngeal pH monitoring [1].

Patients were excluded for any of the following conditions: 1) respiratory or gastrointestinal malignancy; 2) radiation therapy or surgery for the head, neck, lung, or gastrointestinal tract; 3) trauma or surgery near the larynx; 4) currently smoking or history of heavy smoking and substance or alcohol abuse history; 5) infectious laryngitis in the previous 3 months; 6) exposure to environmental irritants in the past 3 months; 7) vocal cord papilloma, enlarged lingual or palatine tonsils, or goiter; 8) excessive voice use; 9) bronchial asthma; 10) chronic cough attributable to angiotensin-converting enzyme inhibitor, or known chronic pulmonary or tracheobronchial etiologies, such as eosinophilic bronchitis, bronchiectasis, positive methacholine provocation test result, or response to inhaled or systemic steroid; 11) pharyngeal (Zenker's) diverticulum or esophageal stasis syndrome, such as achalasia; anxiety or depression with response to at least 1 month of treatment with an anxiolytic or an antidepressant [22]; 12) chronic or allergic rhinosinusitis, nasal polyposis, or postnasal drip with response to at least 1 month of medical therapy with antihistamine, topical steroid spray, or defined by nasal endoscopy or computed tomography scan; 13) participation in another investigational drug study in the previous month; 14) acid suppressive therapy within 4 weeks before recruitment; 15) need for continuous therapy with theophylline, iron supplements, warfarin, antifungal drugs, and digitalis; 16) women were required to be nonpregnant and nonlactating and to maintain effective contraception if of child-bearing potential; and 17) had a serious illness that would interfere with study participation, or refused to participate.

Screening Period

Patients who met the eligibility criteria were enrolled in a 2- to 4-week run-in period to screen out ineligible or noncompliant participants and to confirm that the severity of the primary laryngeal symptom was unchanged before treatment. We used a self-administered four-point Likert scale (0 = none, 1 = mild, 2 = moderate, and 3 = severe) to assess five laryngeal symptoms (globus sensation, throat pain, hoarseness, cough, and throat clearing) [19]. Participants identified the single most bothersome symptom among the five symptoms as the primary laryngeal symptom before treatment. The Likert scale was used to assess the symptom severity in two assessments 7 to 14 days apart. The severity of the primary laryngeal symptom was required to be at least moderate (or ≥2 points) in both assessments.

Study Design

Participants took an oral esomeprazole 40 mg tablet (Nexium; AstraZeneca Pharmaceuticals, Södertälje, Sweden) 30 minutes before breakfast and 30 minutes before dinner because a twice-daily dose was previously shown to be superior to a regular dose in a subset of patients [17]. Before esomeprazole treatment, the RSI instrument was administered at baseline. During the treatment period, patients' adherence to treatment, adverse events, and concomitant medication were evaluated and documented by filling out the questionnaires with the assistance of the study nurse at 4-, 8-, and 12-week follow-up visits before an interview with the investigator (H.-C. Lien). At week 12 after completion of esomeprazole treatment, each participant filled out the self-administered questionnaires including a 10-cm visual analogue scale to assess the improvement in the primary laryngeal symptom and the RSI score for outcome measures.

Laryngoscopy and Upper Gastrointestinal Endoscopy

Laryngoscopy was performed using a flexible nasolaryngoscope (VNL-1171K; Pentax, Tokyo, Japan) to exclude malignancies of the

upper airway and to document LPR signs on the basis of the RFS at enrollment. The RFS is an eight-item scoring scale for assessing clinical severity, ranging from 0 (no abnormal finding) to 26 (worst score possible) [21]. The same laryngologist (C.-C. Wang) performed all the RFS scoring examinations.

The presence of reflux esophagitis was evaluated using upper gastrointestinal endoscopy (GIF XQ-240; Olympus, Tokyo, Japan) on the basis of Los Angeles classification before treatment.

Twenty-Four–Hour Ambulatory Esophagopharyngeal pH Monitoring

An ambulatory 24-hour pH catheter incorporating three antimony sensors into a bifurcated probe with a single connector was used (Sandhill Scientific, Highlands Ranch, Colo). A detailed description of the technique has been described previously [23]. Briefly, manometry was used to position the proximal sensor 1 cm above the upper esophageal sphincter, the middle sensor at 10 cm distal to the proximal one, and the distal sensor at 5 cm above the lower esophageal sphincter. The participants consumed their usual diet but excluded citrus fruit, acidic beverages, carbonated beverages, caffeinated beverages, and any antireflux medications including PPIs. An *excessive esophagopharyngeal acid reflux* was defined as the presence of 1) one or more episode of pharyngeal acid reflux [23] and/or 2) excessive distal esophageal acid reflux, that is, 4.6% of total acid exposure time with pH $<$ 4 at 5 cm above the upper margin of the lower esophageal sphincter [24].

The RSI Questionnaire

The RSI is a self-administered nine-item symptom questionnaire and can be completed in less than 1 minute. The scale for each item ranges from 0 (no problem) to 5 (severe problem), with a maximum total score of 45 for the assessment of symptom severity in patients with LPR [8].

Linguistic Translation

Linguistic translation of the RSI from English to Chinese was performed by two bilingual translators (S.P. Lien and H.-C. Lien) who independently translated the RSI into Chinese. The two versions were reconciled in a joint discussion by both translators to detect errors and to obtain a unified final version. A back translation was then performed by a native English speaker (A. Lee) who has lived in Taiwan for 20 years and speaks fluent Chinese [25]. The back-translated English version was compared with the original English-version RSI by a committee comprising a methodologist (W.M. Liang), a bilingual gastroenterologist (C.S. Chang), a language professional (S. Brenda), and three translators (H.-C. Lien, S.P. Lien, and A. Lee). The committee reviewed all the translations including forward and back translations and reached a consensus on any discrepancy to produce a prefinal version. Subsequently, a pretest of the prefinal version was conducted using 25 randomly selected patients with LPR with education levels ranging from elementary school to university. Each participant was asked whether he or she had difficulty understanding the Chinese version of the questionnaire. Subsequent modifications were made according to their suggestions. The final version was approved by the committee with no semantic differences and no conceptual differences detected compared with the original English version (see Appendix Fig. 1 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2015.01.001.

Outcome Measures

Two outcome measures were used to evaluate the esomeprazole treatment response. One was 50% or more reduction in the primary laryngeal symptom (globus sensation, throat pain, hoarseness, cough, and throat clearing), which is an empirical responder criterion to differentiate responders from nonresponders commonly used in previous LPR clinical trials [5,17–19]. We used a 10-cm visual analogue scale by asking, "Compared to the baseline status (before treatment), what is the percentage of improvement in your primary laryngeal symptom?" (0 cm, no improvement or deterioration; 10 cm, 100% improvement) at week 12. Participants were reminded of the single most bothersome symptom they identified as the primary laryngeal symptom among five laryngeal symptoms before treatment. The other outcome measure was the change in the total RSI score measured from baseline to week 12 during the treatment period.

Statistical Analyses

Descriptive statistics were used to summarize demographic data. To examine overall score distributions, the proportions of respondents with the lowest (0) and highest possible RSI scores (45) were calculated for the presence of floor and ceiling effects.

Validation of the Chinese-version RSI

The Chinese-version RSI was validated by the evaluation of reliability and responsiveness. The reliability of the RSI was examined by internal consistency and test-retest reliability. The internal consistency was assessed by Cronbach α, and the acceptable overall value of Cronbach α was between 0.7 and 0.9 [26]. The test-retest reliability was calculated by comparing the RSI scores between the two visits 7 to 14 days apart before treatment in a randomly selected subgroup of 43 subjects. The sample size adequacy for test-retest reliability was examined according to the guidelines set by Bland and Altman [27], in which the standard error of the within-subject SD (S_w) is shown to be dependent on the number of subjects (n) and the number of observations per study subject (m). Our sample of 43 subjects with two repetitions (n = 43; m = 2) indicated that we can be confident that the estimate of S_w that we obtained was within 22% of its true (population) value [27]. The test-retest reliability was expressed by intraclass correlation coefficient. Intraclass correlation coefficient values of more than 0.7 are considered acceptable [28]. The responsiveness to change during treatment was evaluated by effect size, which was calculated by dividing the mean difference between the baseline and week 12 by the SD at baseline. The effect size was determined to assess the relative size of change: an effect size of 0.2 was considered to be small, 0.5 to be medium, and 0.8 or greater to be large [29,30].

Conceptual Association between the Anchor and the RSI

The relationship between the targeted concept of the RSI and the concept measured by the anchor was evaluated by determining the association between the mean change in RSI scores from baseline and a 50% or more reduction in the primary laryngeal symptom at week 12 using *t* test and effect size.

Anchor-Based Method to Determine responder Definition on the RSI

We used the anchored-based method according to the US FDA recommendation [15]. First, we used an empirical criterion, that is, 50% or more reduction in the primary laryngeal symptom at week 12, as the anchor to dichotomize patients into two groups [16]. A 50% or more reduction in the primary (or global) laryngeal symptom has been used as an end-point measure in previous clinical trials [5,17–19]. Second, we plotted the cumulative distribution of the RSI score change for each group (Fig. 2). We then defined the sensitivity as cumulative percentage for patients with 50% or more reduction in the primary laryngeal symptom, and the specificity as 1 – cumulative percentage for patients with less

Table 1 – Demographics variables and clinical baseline characteristics of 96 subjects with suspected laryngopharyngeal reflux.

Variable	Total (N = 96)	Completed* at week 12 (N = 84)	Not completed at week 12 (N = 12)
Age (y), mean ± SD	49.5 ± 12.6	49.3 ± 12.6	50.9 ± 13.4
BMI (kg/m^2), mean ± SD	23.7 ± 3.8	23.6 ± 3.8	23.7 ± 3.6
Sex: male, n (%)	54 (56.3)	50 (59.5)	4 (33.3)
Primary laryngeal symptom, n (%)			
Globus sensation	28 (29.2)	25 (29.8)	3 (25.0)
Throat pain	15 (15.6)	14 (16.7)	1 (8.3)
Hoarseness	29 (30.2)	25 (29.8)	4 (33.3)
Cough	15 (15.6)	11 (13.1)	4 (33.3)
Throat clearing	9 (9.4)	9 (10.7)	0 (0.0)
Typical reflux syndrome, n (%)	59 (61.5)	52 (61.9)	7 (58.3)
Erosive esophagitis, n (%)	60 (63.2)	55 (66.3)	5 (41.7)
Excessive esophagopharyngeal acid reflux, n (%)	49 (51.0)	44 (52.4)	5 (41.7)
Reflux Finding Score, median (IQR)	6 (4–7)	6 (4–7)	6 (5–8)
RSI score, median (IQR)	17 (12–22)	17 (12–22)	17.5 (10–26)

BMI, body mass index; IQR, interquartile range; RSI, Reflux Symptom Index; typical reflux syndrome, heartburn or acid regurgitation.
* Completed: Patients with no missing data at week 12.

than 50% reduction in the primary laryngeal symptom. Third, we plotted the ROC curve on the basis of sensitivity and specificity calculated for each point of the RSI score change. Finally, we used the maximal Youden index from the ROC curve to determine an optimal cutoff point of the RSI score change that served as the definition of responder. The sensitivity and the specificity were also calculated in subgroups with different baseline scores, that is, less than 12, 12 to 27, and 28 or more. A baseline score of less than 12 was considered to be a low baseline based on the 95% upper limits of the norm in the previous studies [8,11,12]. A baseline score of 28 or more was chosen as the high baseline from the top 10% of the subjects.

Results

Flow of Patients

A total of 228 subjects were assessed for eligibility. The intent-to-treat population consisted of 96 subjects. Twelve subjects dropped out, so 84 subjects were included in the per-protocol analysis (see Appendix Fig. 2 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2015.01.001). Esomeprazole was generally well tolerated. There was no serious adverse event requiring emergency care or hospitalization. The most commonly reported adverse events were abdominal fullness, constipation, diarrhea, headache, and dyspepsia. There were no differences in the baseline characteristics between participants and nonparticipants (Table 1). Among 84 subjects who completed the study, 38 reported 50% or more reduction in the primary laryngeal symptom at week 12. The remaining 46 reported less than 50% reduction in the primary laryngeal symptom at week 12, including 23 subjects with no improvement or deterioration on the basis of the visual analogue scale (see Appendix Fig. 2 in Supplemental Materials).

Validation of the Chinese-Version RSI

At baseline, there was neither floor effect nor ceiling effect for total RSI scores (Table 2). Mean RSI scores changed from 17.5 ± 7.5 at baseline to 10.8 ± 7.2 at week 12 ($P < 0.0001$). The Chinese-version RSI showed a good internal consistency, good test-retest reliability, and a good responsiveness (Table 2).

The Conceptual Association between the Anchor and the RSI

The mean RSI score reduction in 38 subjects with a 50% or more reduction in the primary laryngeal symptom was significantly greater than that in 46 subjects with a less than 50% reduction in the primary laryngeal symptom at week 12 (−11.0 ± 7.8 vs. −3.1 ± 8.3, $P < 0.0001$; effect size 0.99; Fig. 1). Similarly, the mean RSI score reduction was also greater for individual RSI items in subjects with a 50% or more reduction in the primary laryngeal symptom than in those without (see Appendix Fig. 3A in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2015.01.001). The phenomenon existed for most items in each subgroup

Table 2 – Floor and ceiling effects, internal consistency, test-retest reliability, and responsiveness for the Chinese-version RSI.

Measure	Criterion	Range	Ideal	Result
Floor effect (%)		0–100		0.0
Ceiling effect (%)		0–100		0.0
Internal consistency	Cronbach α	0–1	>0.7	0.74
Test-retest reliability	Intraclass correlation coefficient	0–1	>0.7	0.79
Responsiveness	Effect size*		>0.8	0.92
	Paired t test (P value)	0–1	<0.05	<0.0001

RSI, Reflux Symptom Index.
* Effect size, based on Cohen's definition $d = \frac{\bar{x}_1 - \bar{x}_2}{s}$; $s = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2}}$.

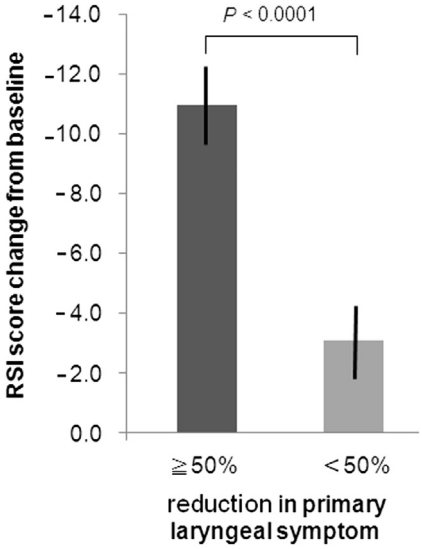

Fig. 1 – Mean changes in RSI scores from baseline between patients with 50% or more and less than 50% reduction in the primary laryngeal symptom at week 12. (The "I" bars represent standard error.) RSI, Reflux Symptom Index.

with different primary laryngeal symptoms despite the small sample size (see Appendix Fig. 3B-F in Supplemental Materials).

The Anchor-Based Method to Determine the Responder Definition for the RSI

The cumulative percentage of patients at various cutoff points of the RSI score reduction showed distinct distributions between patients with and without 50% or more reduction in the primary laryngeal symptom at week 12 (Fig. 2). The ROC curve based on the sensitivity and specificity calculated for each point of the RSI score was plotted, and the area under the ROC curve was 0.77. A reduction of at least six points in the RSI score from the baseline was determined to be the definition of responder on the basis of the maximal Youden index. The sensitivity was 0.79 and the specificity was 0.70 for the diagnosis of a 50% or more reduction in the primary laryngeal symptom at week 12 (Table 3). The sensitivity and the specificity of a six-point reduction in the RSI score varied at different baseline RSI scores (Table 3), but remained stable across different subgroups of each primary laryngeal symptom (see Appendix Table 1 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2015.01.001).

Discussion

In this study, we applied a disease-specific PRO instrument, the RSI, in the assessment of patients with LPR to determine an a priori responder definition for outcome measure using the anchor-based method in accordance with the final version of the PRO guidance released by the US FDA in 2009.

There are three major findings of this study. First, the Chinese-version RSI was shown to be reliable and responsive to change during the treatment in Taiwanese patients with LPR. Second, the empirical criterion, a 50% or more reduction in the primary laryngeal symptom, was highly correlated with the change in RSI scores in response to the PPI treatment. Third, a responder definition of a reduction of six points or more in the RSI score at week 12 (possible range −45 to 45) using the anchor-based method was proposed to measure the treatment benefit.

The FDA guidance emphasizes the importance of taking the patient's perspective into account. To capture optimal informa-

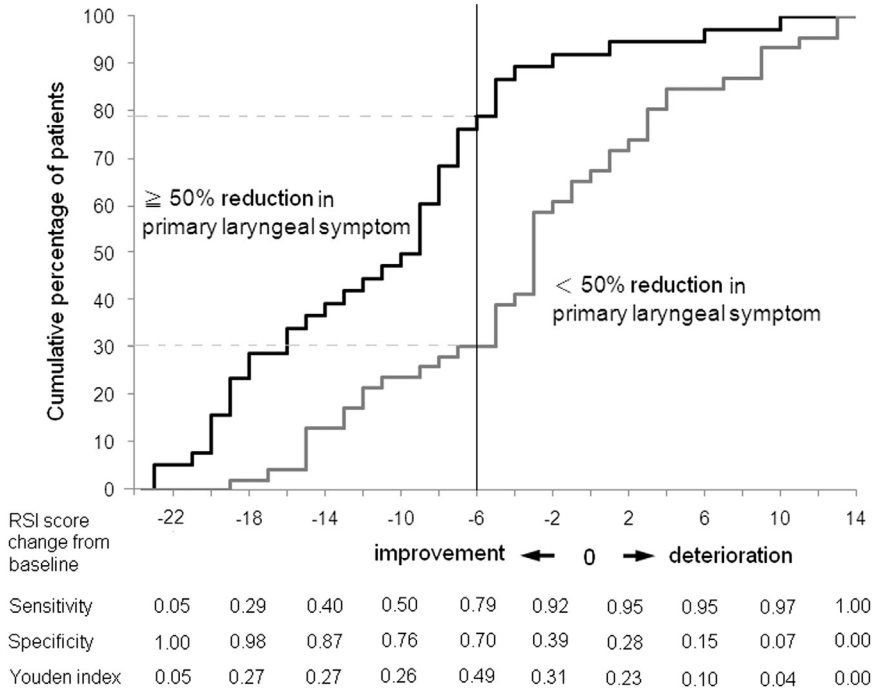

Fig. 2 – Illustrative cumulative distribution function shows a distinct difference in cumulative percentage at the RSI score reduction of six points from baseline between patients with 50% or more and less than 50% reduction in the primary laryngeal symptom at week 12. (X-axis, the RSI score change from baseline; Y-axis, cumulative percentage of patients; Sensitivity, cumulative percentage for patients with ≥50% reduction in the primary laryngeal symptom; Specificity, 1 − cumulative percentage for patients with <50% reduction in the primary laryngeal symptom). RSI, Reflux Symptom Index.

Table 3 – Sensitivity and specificity of the RSI score reduction of at least six points as responder definition for diagnosing an empirical criterion, i.e., ≥50% reduction in the primary laryngeal symptom, among all subjects and among subgroups with different baselines.

Diagnostic performance	All Subjects (n = 84)	Subgroups with different baselines		
		<12 (n = 19)	12–27 (n = 56)	≥28 (n = 9)
Sensitivity	0.79 (30/38)	0.17 (1/6)	0.90 (26/29)	1.00 (3/3)
Specificity	0.70 (32/46)	1.00 (13/13)	0.67 (18/27)	0.17 (1/6)

RSI, Reflux Symptom Index.

tion from the patient's experience, disease-specific PRO instruments should be developed on the basis of a review of the literature and input from physicians and targeted patients. Currently, only three LPR-specific symptom questionnaires are available: the Supraesophageal Reflux Questionnaire [31], the Pharyngeal Reflux Symptom Questionnaire [32], and the RSI. Both the Supraesophageal Reflux Questionnaire and the Pharyngeal Reflux Symptom Questionnaire, however, have not been validated with respect to responsiveness to change during the treatment. In contrast, the RSI instrument was developed and validated in a sample of pH-test–proven patients with LPR by Belafsky et al. [8], including good construct validity demonstrated by patients with a five-point or better improvement in the RSI score being 11 times more likely to have a five-point improvement in the Voice Handicap Index score, and known group validity demonstrated by a significantly higher mean RSI score (21.2) in patients with LPR before treatment than in asymptomatic individuals (11.6). In addition, a greater reduction in RSI scores after the PPI treatment was also found to be associated with an abnormal pH test result in Chinese patients with suspected LPR [18].

In the FDA guidance, the anchor-based method using an empirical criterion is recommended to determine the responder definition for the PRO instrument at an individual patient level. [15]. To be useful, the anchors chosen should have intuitive meaning, be easier to interpret than the PRO measure itself, and should correlate well with the target concept of the PRO instrument. In this study, we chose a 50% or more reduction in the primary laryngeal symptom as the anchor because it was the primary or the most bothersome laryngeal symptom identified by the patient and has been used as an end-point measure in previous clinical trials [5,17–19]. In addition, previous focus group discussions have revealed that LPR-related symptoms such as voice problems, chronic cough, throat clearing, and swallowing difficulties are the key concerns of patients with LPR and negatively affect health-related quality of life [33]. Our finding of a high correlation between the mean RSI score reduction from baseline and a 50% or more reduction in the primary laryngeal symptom at treatment end may therefore justify the application of the anchor chosen to determine the responder definition for the RSI instrument. Alternatively, a laryngoscopic signs scoring system, such as the RFS [21], or a 50% or more reduction in the global laryngeal symptom [5] may be considered as possible anchor candidates. The former, however, may require a longer duration to observe any change [34], and may not correlate well with symptoms [35], whereas the latter averages a complex of changes in symptoms over a long period of time and is subject to recall bias.

Recently, Vakil et al. [36] reviewed current PRO instruments in gastroesophageal reflux disease and found that only five have been used as end points in clinical trials. Among them, either no responder was defined or responder was defined as freedom from symptoms. Alternatively, the Reflux Questionnaire-GI subscale uses a score below 1.73 derived from the 95th percentile of healthy controls to define symptomatic response [37]. The use of such distribution-based methods as the sole basis for determining a responder definition, however, is considered supportive but not appropriate in the FDA guidance [15]. Traditionally, the disadvantage of the distribution-based approach is that it does not provide information about the clinical importance, whereas the disadvantage of the anchor-based method is its inability to take into account the variability of the instrument and/or the sample. In this study, we combined the anchor-based and distribution-based methods to take advantage of both an external criterion and a measure of variability [16]. With the anchor-based method, the patients can be dichotomized into with and without marked improvement. With these two sample distributions, the optimal cutoff value of the score change in the PRO instrument can be determined using an ROC curve with the maximal Youden index. Using these approaches, we found that the cutoff value was a reduction of six points or more in the RSI score from baseline, which had a sensitivity of 79% and a specificity of 70% for the prediction of the anchor (Fig. 2). In addition, the value of a six-point change remains quite stable across the five primary laryngeal symptoms (see Appendix Table 1 in Supplemental Materials), suggesting its suitability for use as a responder definition in future clinical trials.

The responder definition of the RSI instrument in our study may be advantageous in both clinical practice and clinical trials, because it enables the interpretation of the treatment benefit at the patient level, thereby facilitating the patient-doctor communication with respect to the treatment efficacy. In addition, it assesses the current state in a complex of LPR symptoms with no or little recall bias, and therefore can appropriately reflect the health status of individuals in real time. Furthermore, this method is also likely to be superior to the use of the conventional minimally important difference [14], which was derived from point estimates of intrapatient mean group change, possibly masking individual important change, and therefore was not included in the revised 2009 FDA guidance.

There may, however, be some limitations. First, the responder definition of the RSI based on a score reduction of six points or more may not be invariant across various baseline values using our approach. Patients with a low baseline RSI score (<12) have a low sensitivity, whereas patients with a high baseline RSI score (≥28) have a low specificity in the prediction of the anchor (Table 3). Not surprisingly, patients with a baseline RSI score of less than 12, which is within the range of normal controls in previous studies [8,11,12], may have either fewer or less severe symptoms, and thus are less likely to have a reduction of six points despite adequate or complete relief of the primary laryngeal symptom. In contrast, a high baseline RSI score of 28 or a larger arbitrarily chosen value on the basis of the top 10% of our study subjects may indicate an outlier value. Such patients may have either more symptom items or more severe symptom. Because of the phenomenon of "regression to the mean," they are likely to have a reduction of six points or more despite inadequate relief of the primary laryngeal symptom. This may be an inherent limitation of using the PRO instrument for the determination of a responder definition. In such patients with

high baseline RSI score, a different cutoff point may be required to define another responder definition [16]. In addition, the value of a high baseline RSI score may vary from study to study. Future large-scale studies may need to clarify these issues. Second, the precision of a six-point estimate may be limited because data were obtained from a relatively small sample size in a single referral center. This limitation, however, may be attenuated in part by applying a reliable and responsive PRO instrument in a cohort receiving medications with a long follow-up period. Nevertheless, future research is warranted to confirm our findings. Third, the RSI lacks content concerning frequency and duration of symptoms [32] and the focus group discussion was not mentioned in the development study [8]. Despite its imperfections, however, it was shown to be superior to other validated LPR-specific symptom questionnaires in terms of ease in practice, widespread use, and containing most relevant symptoms. Finally, the RSI instrument consisted of eight throat-symptom items and one esophageal-symptom item, and should only be used in patients with major complaint of reflux-related throat symptoms if a responder definition is to be used as the end point of PRO measures.

Conclusions

We found that a reduction of six points or more in the RSI score from baseline was sufficiently sensitive and specific for use as the responder definition to interpret the treatment benefit in patients with LPR. This responder definition meets the US FDA guidance for industry to evaluate PROs in medical product development to support labeling claims, and may be used as the end-point measure in clinical practice and clinical trials. A reduction of six points or more in the RSI score, however, is a preliminary estimate, and additional studies across other language versions are required.

Acknowledgments

We thank Dr. Belafsky for his generosity to agree to do the translation of the Reflux Symptom Index into Chinese language, Dr. El-Wui Loh from Kaohsiung Municipal Kai-Syuan Psychiatric Hospital of Taiwan for his valuable comments, and Ms Kareen Chong for her secretarial work. Parts of information given in this article have been used in a poster presentation at the Third Biennial Conference of the Asian Neurogastroenterology and Motility Association, Penang, Malaysia, March 1, 2013.

Source of financial support: Taiwan's Ministry of Science and Technology supported this study (grant no. NSC102-2314-B-075A-004-MY2).

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2015.01.001 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

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 Otolaryngology - Head and Neck Surgery. Otolaryngol Head Neck Surg 2002;127:32–5.

[2] Vakil N, van Zanten SV, Kahrilas P, et al. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. Am J Gastroenterol 2006;101:1900–20.

[3] Kahrilas PJ, Shaheen NJ, Vaezi MF, et al. American Gastroenterological Association Medical Position Statement on the management of gastroesophageal reflux disease. Gastroenterology 2008;135:1383–91.

[4] Ford CN. Evaluation and management of laryngopharyngeal reflux. JAMA 2005;294:1534–40.

[5] Qadeer MA, Phillips CO, Lopez AR, et al. Proton pump inhibitor therapy for suspected GERD-related chronic laryngitis: a meta-analysis of randomized controlled trials. Am J Gastroenterol 2006;101:2646–54.

[6] Reichel O, Dressel H, Wiederänders K, Double-blind Issing WJ. placebo-controlled trial with esomeprazole for symptoms and signs associated with laryngopharyngeal reflux. Otolaryngol Head Neck Surg 2008;139:414–20.

[7] Lam PK, Ng ML, Cheung TK, et al. Rabeprazole is effective in treating laryngopharyngeal reflux in a randomized placebo-controlled trial. Clin Gastroenterol Hepatol 2010;8:770–6.

[8] Belafsky PC, Postma GN, Koufman JA. Validity and reliability of reflux symptom index (RSI). J Voice 2002;16:274–7.

[9] Schindler A, Mozzanica F, Ginocchio D, et al. Reliability and clinical validity of the Italian Reflux Symptom Index. J Voice 2010;24:354–8.

[10] Printza A, Kyrgidis A, Oikonomidou E, Triaridis S. Assessing laryngopharyngeal reflux symptoms with the Reflux Symptoms Index: validation and prevalence in Greek population. Otolaryngol Head Neck Surg 2011;145:974–80.

[11] Farahat M, Malki KH, Mesallam TA. Development of the Arabic version of Reflux Symptom Index. J Voice 2012;26(814):e15–9.

[12] Cohen JT, Gil Z, Fliss DM. The reflux symptom index–a clinical tool for the diagnosis of laryngopharyngeal reflux. Harefuah 2005;144:826–9:(912).

[13] Wright JG. The minimal important difference: who's to say what is important? J Clin Epidemiol 1996;49:1221–2.

[14] Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. COPD 2005;2:63–7.

[15] U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for industry 2009; Patient-reported outcome measures: use in medical product development to support labeling claims. Available from: http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf. [Accessed March 23, 2014].

[16] de Vet HC, Ostelo RW, Terwee CB, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. Qual Life Res 2007;16:131–42.

[17] 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;115:1230–8.

[18] Lien HC, Wang CC, Liang WM, et al. Composite pH predicts esomeprazole response in laryngopharyngeal reflux without typical reflux syndrome. Laryngoscope 2013;123:1483–9.

[19] Vaezi MF, Richter JE, Stasney CR, et al. Treatment of chronic posterior laryngitis with esomeprazole. Laryngoscope 2006;116:254–60.

[20] Ahmed TF, Khandwala F, Abelson TI, et al. Chronic laryngitis associated with gastroesophageal reflux: prospective assessment of differences in practice patterns between gastroenterologists and ENT physicians. Am J Gastroenterol 2006;101:470–8.

[21] Belafsky PC, Postma GN, Koufman JA. The validity and reliability of the reflux finding score (RFS). Laryngoscope 2001;111:1313–7.

[22] Oyer SL, Anderson LC, Halum SL. Influence of anxiety and depression on the predictive value of the reflux symptom index. Ann Otol Rhinol Laryngol 2009;118:687–92.

[23] Lien HC, Wang CC, Hsu JY, et al. Classical reflux symptoms, hiatus hernia and overweight independently predict pharyngeal acid exposure in patients with suspected reflux laryngitis. Aliment Pharmacol Ther 2011;33:89–98.

[24] Hu WH, Wong NY, Lai KC, et al. Normal 24-hour ambulatory proximal and distal gastroesophageal reflux parameters in Chinese. Hong Kong Med J 2002;8:168–71.

[25] Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 1993;46:1417–32.

[26] DeVellis RF. Scale Development. (2nd ed.). Thousand Oaks, Calif: Sage, 2003.

[27] Bland JM, Altman DG. Statistics notes: measurement error. BMJ 1996;313:744.

[28] Nunnally JC, Bernstein IH. Psychometric Theory. New York: McGraw-Hill, 1994.

[29] Cohen J. Statistical Power Analysis for the Behavioural Sciences. New York: Academic Press, 1977.

[30] Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27(3, Suppl.):S178–89.
[31] Dauer E, Thompson D, Zinsmeister AR, et al. Supraesophageal reflux: validation of a symptom questionnaire. Otolaryngol Head Neck Surg 2006;134:73–80.
[32] Andersson O, Rydén A, Ruth M, et al. Development and validation of a laryngopharyngeal reflux questionnaire, the Pharyngeal Reflux Symptom Questionnaire. Scand J Gastroenterol 2010;45: 147–59.
[33] Lenderking WR, Hillson E, Crawley JA, et al. The clinical characteristics and impact of laryngopharyngeal reflux disease on health-related quality of life. Value Health 2003;6:560–5.
[34] Belafsky PC, Postma GN, Koufman JA. Laryngopharyngeal reflux symptoms improve before changes in physical findings. Laryngoscope 2001;111:979–81.
[35] Belafsky PC, Postma GN, Amin MR, Koufman JA. Symptoms and findings of laryngopharyngeal reflux. Ear Nose Throat J 2002;81(9, Suppl. 2):10–3.
[36] Vakil NB, Halling K, Becher A, Rydén A. Systematic review of patient-reported outcome instruments for gastroesophageal reflux disease symptoms. Eur J Gastroenterol Hepatol 2013;25:2–14.
[37] Stanghellini V, Armstrong D, Mönnikes H, et al. Determination of ReQuest-based symptom thresholds to define symptom relief in GERD clinical studies. Digestion 2005;71:145–51.