Original Article

Development and Psychometric Assessment of a Self-evaluation Questionnaire for Gastroesophageal Reflux Disease

Hye-kyung Jung,[1] Myung-Gyu Choi,[2]* Myung-Ki Baek,[3] and Justin C Y Wu[4]

[1]Department of Internal Medicine, Ewha Womans University, College of Medicine, Seoul, Korea; [2]Department of Internal Medicine, College of Medicine, The Catholic University of Korea, Seoul, Korea; [3]Department of Internal Medicine, College of Medicine, International St. Mary's Hospital, Catholic Kwandong University, Incheon, Korea; and [4]Institute of Digestive Disease, The Chinese University of Hong Kong, Shatin, Hong Kong

Background/Aims

There is uncertainty about how to measure outcomes reported by patients in gastroesophageal reflux disease (GERD). This study was conducted to develop an instrument and to determine of the definition of respondent for a patient reported outcomes to assess the efficacy of a treatment used for GERD treatment.

Methods

A structural process has developed a self-evaluation questionnaire for GERD (SEQ-GERD); health-related quality of life questionnaire for GERD (GERD-QOL) was translated through cross-cultural validation. Two-week reproducibility was evaluated and construct validity was assessed by correlating the SEQ-GERD with the Patient Assessment of Gastrointestinal Disorders (PAGI-SYM), the reflux disease questionnaire (RDQ), and GERD-QOL. Changes in SEQ-GERD scores were compared to assess the discriminative validity following 4 weeks of proton pump inhibitor administration.

Results

A total of 83 Korean patients were included (mean age 46 ± 14 years, females 61.4%). The internal consistency of the 19-item SEQ-GERD was good (alpha = 0.60-0.94) and the test–retest reliability was high (intra-class correlation coefficient = 0.67-0.95). The SEQ-GERD highly correlated with the GERD domain of the PAGI-SYM (correlation coefficient $r = 0.894$, $P < 0.001$), the RDQ-GERD ($r = 0.877$, $P < 0.001$), and GERD-QOL ($r = -0.536$, $P < 0.05$). SEQ-GERD scores significantly varied according to the overall treatment effectiveness scale of drug responsiveness and significantly decreased after drug treatment (mean differences according to the overall treatment effectiveness scale, $P = 0.020$).

Conclusion

This study supports that SEQ-GERD is reliable and valid, and can be used to evaluate the treatment response in patients with GERD.
(J Neurogastroenterol Motil 2018;24:584-592)

Key Words

Gastroesophageal reflux disease; Patient reported outcome measures; Surveys and questionnaires; Validation studies

*Correspondence: Myung-Gyu Choi, MD, PhD
Department of Internal Medicine, College of Medicine, The Catholic University of Korea, 222 Banpodae-ro, Seocho-gu, Seoul 06591, Korea
Tel: +82-2-2258-6023, Fax: +82-2-2258-2055, E-mail: choim@catholic.ac.kr

Introduction

Gastroesophageal reflux disease (GERD) is a common condition that shows diverse clinical manifestations, including erosive esophagitis (EE), non-erosive reflux disease (NERD), and Barrett's esophagus. The prevalence of GERD in the Western population is reported to be 10.0-20.0%, while it ranges from 5.2-18.3% in Asia.[1,2] Among patients who have typical GERD symptoms including heartburn and/or acid regurgitation, NERD constitutes a substantial portion of GERD cases (50.0-70.0%).[3] EE can be clearly defined by endoscopy; however, NERD is diagnosed based on the presence of typical GERD symptoms in the absence of visible esophageal mucosal lesions at endoscopy. Furthermore, GERD symptoms have been demonstrated to not correlate well with the degree of esophageal injury at endoscopy or esophageal pH monitoring.[4] These observations have led to the concept of GERD as a symptom-driven disease. Moreover, because patients seek healthcare due to their symptoms, evaluating the response of therapeutic trials is mainly based on symptoms.

Recently, the value gained from understanding health outcomes from patients' perspectives has increased. Patient-reported outcomes (PROs) measure specific health conditions directly from the subject affected, without interpretation of a patient's response by a physician or other individual. It captures the patient's illness experience in a structural format and plays an important role in clinical trials and research studies.[5] The United States Food and Drug Administration (US-FDA) has released definitive guidelines on the use of PRO instruments.[6] The guiding principles recommend the use of validated instruments to assess treatment outcomes in conducting and supporting industry-sponsored clinical trials, and are largely based on scientific standards. The guidelines recommend the documentation of evidence underlying PRO instruments, which should include reliability, validity, responsiveness, and the appropriateness of the PRO measure to a specific patient group.[7,8]

Typical or predominant GERD symptoms are not sensitive to its diagnosis and there is no globally accepted term for "heartburn" in different languages; thus, this symptom terminology cannot be simply translated from one language into another.[9] Appropriate understanding of symptom description requires adequate levels of literacy, abstract thinking, and comprehension of nuances. Therefore, the assessment of PRO instruments may be biased with cultural, educational, and linguistic differences. Because many development programs are multinational, application of PRO instruments to multiple cultures or languages is common in clinical trials. The

FDA recommends a development process to translate and culturally adapt the instrument to populations that will use the PRO instruments in trials.[10-12] We aim to develop a self-reported GERD-specific PRO instrument, which is easily applicable to Korean, based on a psychometric evaluation and validation study.[7]

Materials and Methods

The study protocol was approved by the Ethical Review Committee of Seoul St. Mary's Hospital (KC14QIMI0790) and Ewha Women's University Mokdong Hospital (2015-06-039-003). Written informed consent was obtained from all participants.

Patient Selection

Participants for the validation study were enrolled from outpatient clinics in the Department of Gastroenterology of Seoul St. Mary's hospital. GERD was defined as having one or more typical GERD symptoms, including heartburn and/or acid regurgitation for the last 1 month, with a frequency greater than or equal to 1 day per week. All patients underwent upper gastrointestinal (GI) endoscopy within 6 months of starting the study and the patients having erosive reflux disease and NERD were included. Study subjects were assessed for study eligibility by determining whether the patient had minimal comprehensibility to complete the questionnaire.

Patients were excluded from the study due to any of the following conditions at the second phase: (1) treatment with GERD-related drugs, including anti-secretory agents for more than a week and within 1 month; (2) need for continuous therapy for chronic GI or hepato-biliary disease; (3) previous abdominal surgery, except appendectomy; (4) subjects who were pregnant or lactating; (5) patients with serious systemic illness that would interfere with the study, and (6) patients who refused to participate. All subjects gave informed consent and the study was approved by the institutional review board.

This study was conducted in 2 phases. During the first phase, the questionnaire was developed and a validation study was conducted, which included 20 subjects > 20 years of age with/without GERD symptoms from April 2015 to May 2015. During the second phase, the overall treatment effectiveness (OTE) scale for GERD was used after lansoprazole (30 mg) treatment for 4 weeks in patients with GERD, between the ages of 19-75 years, who first visited the hospital or re-visited the hospital within 6 months, from June to November 2015.

Development of the Self-evaluation Questionnaire for Gastroesophageal Reflux Disease

A literature review was conducted to identify published questionnaires and scripted items related to GERD, which were reviewed by experts (M.G.C. and H.K.J.). New items were identified and the candidate item categories were sorted into an initial structure. Questions concerning GERD symptoms were developed firstly in English and then translated into Korean. Based on the structured translation process, back-translation of the translated draft demonstrated that the target version was functionally equivalent to the original English version.

The draft instrument was pilot-tested in a sample of 15 subjects with GERD symptoms to assess clarity, interpretation, overall relevance, and length of the questionnaire. The draft questionnaire was produced at a 6th-grade reading level and it utilized words specifically voiced by patients.

The final and current edition of the self-evaluation questionnaire for GERD (SEQ-GERD) is composed of 19 questions, mainly addressing the severity and frequency of GERD and GI symptoms over a 1-week recall period, and is divided into a GERD domain and an upper GI symptoms domain. The GERD domain includes the 2 typical GERD symptoms of heartburn or acid regurgitation, along with retrosternal chest pain or regurgitation, 2 GERD-related sleep disturbance symptoms, and 4 extra-esophageal symptoms (Supplementary Table). The upper GI domain also included dyspeptic symptoms (epigastric pain or soreness, postprandial fullness, early satiety, belching, bloating, and nausea/vomiting).

The frequency of symptoms was noted using a 5-point Likert scale ranging from "1" (no symptoms) to "5" (daily symptoms), and severity was ranked using a 5-point Likert scale from "not at all or very weak" to "extremely severe." Scores were calculated into a sum score for each symptom frequency and severity, with higher scores indicating greater severity of symptoms. All items were given equal weight.

Reliability and Validity of Self-evaluation Questionnaire for Gastroesophageal Reflux Disease

Test–retest reliability was performed to evaluate the stability of the SEQ-GERD. Subjects were instructed to complete the questionnaire twice in a 2-week interval before starting treatment with medication. Construct validity was evaluated using external correlation of SEQ-GERD with the Patient Assessment of GI Symptom Severity Index (PAGI-SYM), the Reflux Disease Questionnaire (RDQ), and a disease specific health-related quality of life (QOL) instrument (eg, GERD-QOL).[11,12] An English version of GERD-QOL is valid and reliable disease-specific QOL instrument for GERD patients in Asia. We also underwent the validation work for Korean version of GERD-QOL. Negative correlation with GERD-QOL was expected as the GERD-QOL measures QOL in the reverse direction.

Discriminant validity was evaluated by examining the relationship between the SEQ-GERD subscale scores and dyspeptic symptoms, including early satiety, post-prandial fullness, nausea, vomiting, abdominal pain, and bloating.

The OTE scale was a 5-point Likert scale, defined as follows: 1 point, disappearance of symptoms; 2 points, greatly improved symptoms; 3 points, disappearance of symptoms $\geq 50\%$; 4 points, no change in symptoms; and 5 points, worsening of symptoms. Responsiveness was defined as a patient with symptom responsiveness of $\geq 50\%$ after drug treatment.

Statistical Methods

A sample size of 22 patients was required to achieve acceptable reliability.[13] Internal consistency was assessed using Cronbach's alpha; a value of 0.7 or above is considered to indicate a high degree of internal consistency.[14] Test–retest reliability was evaluated using intra-class correlation coefficients (ICC) and Pearson correlation coefficients. An ICC of 0.7 or above was considered sufficient to demonstrate reliability. Constructive validity was assessed by means of Pearson and Spearman correlation coefficients between the SEQ-GERD and the GERD-subscale scores of the PAGI-SYM, RDQ, and GERD-QOL. Discriminant validity was also assessed using ANOVA to compare groups of patients with different GERD symptom severities.

Responders were defined as patients having improved symptoms of 50% or more after the termination of treatment compared to before treatment. The independent t test or ANOVA test was used to compare the difference of changes in the mean score SEQ-GERD scores from baseline to 4 weeks according to scale of OTE.

Responsiveness was calculated as between-subject change and was evaluated on the basis of changes in GERD symptoms. Independent t tests were used to compare the changes in SEQ-GERD scores from baseline to 4 weeks between responders versus non-responders. The responsiveness index was obtained by dividing the standard deviation of the mean of the pre-post difference in the responsive group by that of the non-responsive group.[15,16] P-values < 0.05 are statistically significant.

Results

Study Participants and Clinical Characteristics

A total of 83 patients with GERD were included in this study. The mean age of the subjects was 46.1 years (SD 14.4 years, range 19-75 years) and 61.4% of the participants were women. The mean GERD symptom duration was 2.8 years (SD 0.9 years) (Table 1). All patients underwent endoscopy within 6 months of starting the study and 56 patients had reflux esophagitis of Los Angeles (LA) classification grade A, while 27 subjects had normal or minimal changes in reflux esophagitis.

Table 1. Study Participants and Clinical Characteristics

Variables	Mean ± SD (range)
Gender (female) (n)	83 (51 [61.4%])
Age (yr)	46.1 ± 14.4 (19-76)
Symptom duration (yr)	2.8 ± 0.9 (1-5)
Endoscopic findings (n)	
LA-A	56 (66.7%)
Normal or minimal change	27 (33.3%)
BMI (kg/m^2)	22.3 ± 3.1 (16.4-32.5)
Current smoker (n)	12 (14.5%)
Alcohol user[a] (n)	17 (20.5%)
Coffee user[b] (n)	14 (16.9%)

[a]Intake of ≥ 24 g alcohol at least 3 times per week or more.
[b]Intake of > 3 cups of coffee per day.
SD, standard deviation; LA-A, Los Angeles classification grade A; BMI, body mass index.

For assessment of test–retest stability, 41 patients repeated the SEQ-GERD 2 weeks after their initial evaluation. Fifty-one patients underwent drug treatment with lansoprazole (30 mg) for 4 weeks, and finally, 44 patients completed the third SEQ-GERD questionnaire. Nine patients completed both tests. The severity of GERD assessed by a 5-point Likert scale was evaluated at the first visit. Thirty-one patients (37.3%) complained of no or mild symptoms, 37 (44.6%) complained of moderate symptoms, and 15 (18.1%) complained of severe or very severe symptoms.

Reliability

Internal consistency was calculated using Cronbach's alpha. The Cronbach alpha coefficient of the typical GERD symptoms (heartburn or acid regurgitation) was 0.81 and that of the 4 major GERD symptoms (heartburn, retrosternal pain, acid regurgitation, and non-acid regurgitation) was 0.87. The Cronbach alpha ranged from 0.61 (for nighttime GERD symptoms) to 0.94 (for bloating/distension) (Table 2). The SEQ-GERD showed high test–retest reliability, demonstrated by ICCs ranging for the SEQ-GERD different subscales. The ICC of postprandial fullness or early satiation was 0.85 and that of epigastric pain was 0.74 (Table 3).

Validity

Using the SEQ-GERD, heartburn highly correlated with retrosternal pain ($r = 0.674$, $P < 0.001$, but it showed modest correlation with acid regurgitation ($r = 0.309$, $P = 0.004$) and poor correlation with early satiety and postprandial fullness (Fig. 1). Awakening at night due to heartburn or acid regurgitation highly correlated with typical GERD symptoms using the SEQ-GERD (r

Table 2. Internal Consistency and Reliability of the Self-evaluation Questionnaire for Gastroesophageal Reflux Disease

Subscale	Cronbach's alpha	P-value
Two typical symptoms (heartburn and acid regurgitation)	0.813	< 0.001
Four major GERD symptoms (heartburn, retrosternal chest pain, acid regurgitation, and non-acid regurgitation)	0.872	< 0.001
Six GERD, symptoms including two nighttime symptoms[a]	0.914	< 0.001
Nighttime GERD symptoms only	0.613	< 0.001
Extra-esophageal symptoms (dysphagia, chronic cough, hoarseness, and sore throat)	0.850	< 0.001
Epigastric pain or soreness	0.902	< 0.001
Postprandial distress (early satiation and postprandial fullness)	0.861	< 0.001
Dyspepsia (epigastric pain or soreness, early satiation, and postprandial fullness)	0.863	< 0.001
Nausea/vomiting	0.752	< 0.001
Bloating or distension	0.941	< 0.001

[a]Sleep disturbance due to GERD at night or heartburn upon awakening in the morning.
GERD, gastroesophageal reflux disease.
Sum of the frequency and severity of each subscale.

**Table 3.** Test–Retest Reliability of the Self-evaluation Questionnaire for Gastroesophageal Reflux Disease

Subscale	ICC	P-value	Pearson's r	P-value
Two typical GERD symptoms (heartburn and acid regurgitation)	0.824	< 0.001	0.639	< 0.001
Four major GERD symptoms[a]	0.798	< 0.001	0.665	< 0.001
Six GERD symptoms including the 2 nighttime symptoms[b]	0.828	< 0.001	0.712	< 0.001
Nighttime GERD symptoms	0.825	< 0.001	0.712	< 0.001
Extra-esophageal symptoms (dysphagia, chronic cough, hoarseness, and sore throat)	0.952	< 0.001	0.912	< 0.001
Epigastric pain or soreness	0.738	< 0.001	0.592	< 0.001
Postprandial distress (early satiation and postprandial fullness)	0.848	< 0.001	0.736	< 0.001
Dyspepsia (epigastric pain, soreness, early satiation, or postprandial fullness)	0.788	< 0.001	0.654	< 0.001
Nausea/vomiting	0.673	< 0.001	0.539	0.001
Bloating or distension	0.807	< 0.001	0.680	< 0.001

[a]Heartburn, retrosternal chest pain, acid regurgitation, and non-acid regurgitation.
[b]Sleep disturbance due to GERD at night or heartburn upon awakening in the morning.
ICC, intra-class correlation coefficient; GERD, gastroesophageal reflux disease.
Sum of frequency and severity of each subscale.
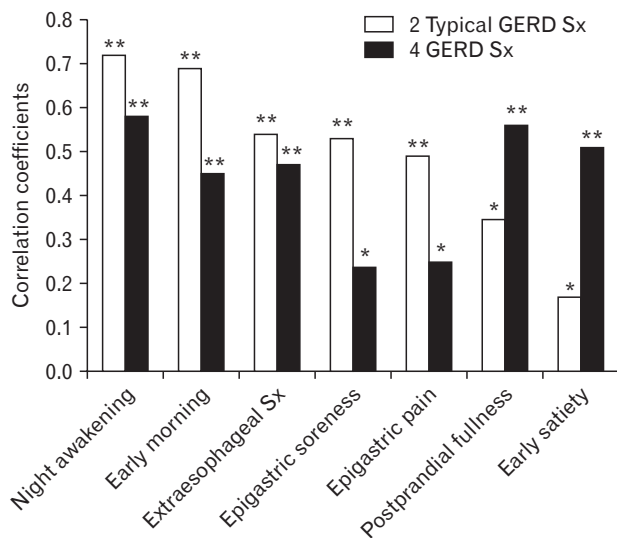
Figure 1. Correlation between the 2 typical symptoms and 4 main symptoms of self-evaluation questionnaire for gastroesophageal reflux disease (GERD) and other upper gastrointestinal symptoms. $*P < 0.05$, $**P < 0.001$. Sx, symptomes.

$= 0.719, P < 0.001$).

The constructive validity of the SEQ-GERD was evaluated through its correlation with the PAGI-SYM GERD subscale scores, RDQ scores, and GERD-QOL scores. The typical GERD symptoms of the SEQ-GERD highly correlated with the PAGI-SYM GERD subscale ($r = 0.758, P < 0.001$), and the 4 major GERD symptoms showed higher correlation with the PAGI-SYM subscale ($r = 0.894, P < 0.001$). The typical GERD

symptoms included in the SEQ-GERD also significantly correlated with the RDQ (typical 2 symptoms, $r = 0.793, P < 0.001$; 4 major symptoms $r = 0.877, P < 0.001$) (Table 4). Moreover, the SEQ-GERD symptoms were also correlated with GERD-QOL; as symptom severity increased, patient reports of QOL decreased significantly. The strongest correlations were observed for 4 major GERD symptoms of the SEQ-GERD and the psychologic well-being domain of GERD-QOL ($r = -0.604, P < 0.001$).

Discriminant validity was determined by examining whether differences in SEQ-GERD scores were significantly different according to self-reported GERD severity and through correlation analysis with the SEQ-GERD and subscales other than those of the PAGI-SYM. The mean scores of the 4 major SEQ-GERD symptoms increased significantly according to the severity of GERD (11.1 ± 2.72, no or mild group [n = 31]; 15.8 ± 4.0, moderate group [n = 37]; and 20.1 ± 8.3, severe/very severe group [n = 15]) ($P < 0.001$). However, the 4 major SEQ-GERD symptoms showed poor correlation with lower abdominal pain scales ($P = 0.351$) and low to moderate correlation with the PAGI-SYM dyspepsia and nausea scales ($-0.241 \leq r \leq 0.560$).

Between-subject changes were evaluated on the basis of changes in clinical status and observed changes in GI-related symptoms (Table 5). A total of 44 patients with GERD were treated with 30 mg lansoprazole for 4 weeks; the responsiveness was subsequently measured using the SEQ-GERD. The mean score of the SEQ-GERD at week 4 was significantly decreased in the responder group compared to the non-responder group (-2.67 ± 4.68 vs 1.80 ± 4.71, Mann-Whitney U test, $P = 0.049$).

Table 4. Construct Validity of the Self-evaluation Questionnaire for Gastroesophageal Reflux Disease Evaluated Through Comparison to the Patient Assessment of Gastrointestinal Symptom Severity Index Gastroesophageal Reflux Disease-subscale, Reflux Disease Questionnaire, and Health-related Quality of Life Questionnaire for Gastroesophageal Reflux Disease

Variables	Typical GERD symptoms	Four major GERD symptoms[a]	Six GERD symptoms[b]
PAGI-SYM_GERD subscale	0.758 (< 0.001)	0.894 (< 0.001)	0.822 (< 0.001)
Postprandial fullness	0.311 (0.048)	0.297 (0.006)	0.550 (< 0.001)
Early satiety	0.284 (0.009)	0.241 (0.028)	0.173 (0.119)
Bloating	0.310 (0.004)	0.351(0.001)	0.347 (0.001)
Epigastric pain	0.476 (< 0.001)	0.665 (< 0.001)	0.631 (< 0.001)
Lower abdominal pain	0.019 (0.867)	0.033 (0.766)	0.550 (< 0.001)
Nausea	0.362 (0.001)	0.560 (< 0.001)	0.428 (< 0.001)
Vomiting	0.389 (< 0.001)	0.404 (< 0.001)	0.550 (< 0.001)
RDQ-GERD	0.793 (< 0.001)	0.877 (< 0.001)	0.842 (< 0.001)
GERD-QOL_DA	−0.507 (< 0.001)	−0.556 (0.025)	−0.305 (0.030)
QOL_TE	−0.379 (< 0.001)	−0.580 (0.018)	−0.188 (0.187)
QOL_Diet	−0.544 (< 0.001)	−0.556 (0.025)	−0.249 (0.078)
QOL_PW	−0.516 (< 0.001)	−0.604 (0.013)	−0.303 (0.031)
QOL_Total	−0.552 (< 0.001)	−0.536 (0.032)	−0.306 (0.029)

[a]Heartburn, retrosternal chest pain, acid regurgitation, and non-acid regurgitation.
[b]Symptoms of 4 gastroesophageal reflux disease (GERD) combined with nighttime symptoms.
PAGI-SYM, patient assessment of gastrointestinal symptom severity index; RDQ, reflux disease questionnaire; GERD-QOL, health-related quality of life questionnaire for GERD; DA, daily activity; TE, treatment effect; PW, psychological well-being; SEQ-GERD, self-reported questionnaire for GERD.
Sum of frequency and severity of each subscale. Data were presented as correlation coefficient (*P*-value).

Table 5. Changes in the Mean Score of the Self-evaluation Questionnaire for Gastroesophageal Reflux Disease by Overall Treatment Effect

SEQ-GERD	No symptoms (n = 4)	Markedly improved (n = 18)	Improved (n = 17)	No change or worse (n = 5)	*P*-value
Typical GERD symptoms					
Baseline	8.00 ± 3.27	8.89 ± 3.27	8.59 ± 2.50	7.40 ± 2.40	0.760
At 4 weeks	4.00 ± 0.00	7.11 ± 2.72	6.94 ± 2.01	8.40 ± 2.88	0.050
Difference	−4.00 ± 3.27	−1.77 ± 2.49	−1.65 ± 2.05	1.00 ± 2.24	0.020
Four major GERD symptoms[a]					
Baseline	14.75 ± 6.70	16.00 ± 5.36	15.00 ± 4.74	14.00 ± 3.32	0.850
At 4 weeks	8.00 ± 0.00	13.39 ± 4.39	13.23 ± 5.43	15.80 ± 5.81	0.130
Difference	−6.75 ± 6.70	−2.61 ± 5.01	−1.76 ± 3.42	1.80 ± 4.71	0.048
Six GERD symptoms[b]					
Baseline	22.25 ± 10.21	23.28 ± 7.79	22.59 ± 5.96	20.20 ± 4.92	0.860
At 4 weeks	12.00 ± 0.00	19.00 ± 5.70	18.65 ± 6.32	21.20 ± 8.29	0.140
Difference	−10.25 ± 10.21	−4.28 ± 6.51	−3.94 ± 5.73	1.00 ± 5.20	0.090

[a]Heartburn, retrosternal chest pain, acid regurgitation, and non-acid regurgitation.
[b]Four gastroesophageal reflux disease (GERD) symptoms combined with nighttime symptoms.
SEQ-GERD, self-reported questionnaire for GERD.
Changes were mean differences according to the overall treatment effectiveness.

SEQ-GERD demonstrated a high responsiveness after 4weeks of treatment, with the responsiveness index exceeding the predefined level of 0.80 (Table 6).

Validation of the Korean Version of the Gastroesophageal Reflux Disease Quality of Life

The reliability coefficient for the GERD-QOL was 0.948 (*P* = 0.001) for the total score; the subscale scores were also highly

Table 6. Responsiveness of the Self-reported Questionnaire for Gastroesophageal Reflux Disease, Patient Assessment of Gastrointestinal Symptom Severity Index, Reflux Disease Questionnaire, and Health-related Quality of Life Questionnaire for Gastroesophageal Reflux Disease

	Mean change	Responsiveness index
SEQ-GERD	−1.63	1.09
PAGI-SYM GERD subscale	−0.49	1.36
RDQ	−1.51	1.87
GERD-QOL	11.40	1.96

reliable (daily activity, 0.915 ($P < 0.001$); treatment effect, 0.882 ($P < 0.001$); and diet, 0.841 ($P < 0.001$). The Pearson's correlation between the total GERD-QOL score and the global GERD-related QOL rated using a 5-point Likert scale was considered satisfactory ($r = 0.448$, $P < 0.001$). There was significant correlation between the total GERD-QOL score and global GERD symptom severity ($r = -0.531$, $P < 0.001$). In the severe GERD group, the diet-related QOL subscale was significantly lower than that of other groups (severe, $52.8 \pm 30.3\%$; moderate, $77.8 \pm 23.8\%$; and mild, $86.6 \pm 16.5\%$; $P = 0.003$) and the psychological well-being subscale score of the moderate to severe group was significantly lower than that of the mild group (mild, $83.3 \pm 22.7\%$; moderate, $56.3 \pm 28.6\%$; and severe, $36.1 \pm 29.6\%$; $P < 0.001$) (Fig. 2).

Discussion

Symptom patterns determine the diagnostic category and the scope of potential therapeutic measures for GERD. The present study documented the psychometric validity of the SEQ-GERD, which was developed in line with the FDA PRO guidelines. The SEQ-GERD showed high reliability, validity, and responsiveness to symptom changes in patients with GERD.

A high degree of accuracy and complete reporting of psychometric properties are essential for clinical trials, especially for medical product development. Although there are many PROs for GERD, there are few validated PRO instruments in Asia.[17] Therefore, we developed a symptom questionnaire for GERD. Patients with GERD complain of various symptoms, and there are differences in symptom nuances between languages and regions. Pictorial representations have been employed in patient health communication, and are demonstrated to improve comprehension and recall of the provided information.[18] Symptoms of heartburn are expressed differently in different languages; thus, we added cartoon style images familiar to Koreans to include in the SEQ-GERD question-
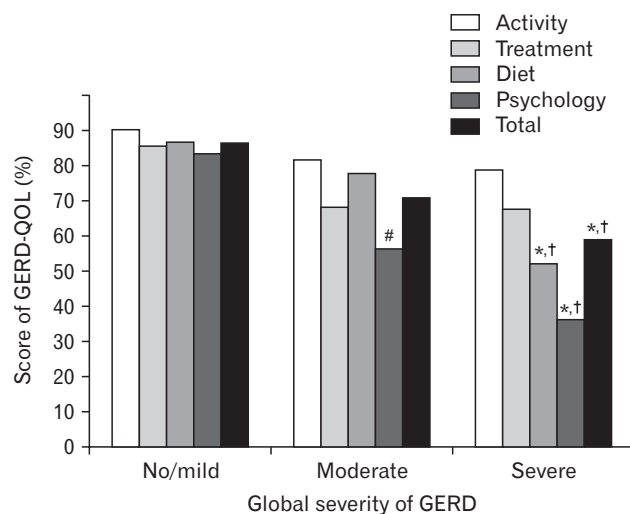
Figure 2. Total and subscale scores of the Korean version of the health-related quality of life questionnaire for gastroesophageal reflux disease (GERD-QOL) instrument according to the severity of GERD. The GERD-QOL includes subscales measuring daily activity, treatment effect, diet, and psychological well-being. *$P < 0.05$ mild vs severe, #$P < 0.05$ mild vs moderate, †$P < 0.05$ moderate vs severe.

naire.

To develop a conceptual framework for the development of the SEQ-GERD, a literature review was performed. The final questionnaire included 2 chest pain symptoms, 2 non-acid regurgitation symptoms, and 2 nighttime symptoms within a 1-week recall period. It showed high reliability and validity for individual scores and the GERD domain composite scores. Among the 6 symptoms, acid regurgitation showed the highest internal consistency and test–retest reliability; however, it was modestly correlated with heartburn (Pearson correlation coefficient, $r = 0.314$, $P = 0.004$, data not-presented). We compared the 2 typical GERD symptoms (ie, heartburn and acid regurgitation) with the 4 main GERD symptoms, as well as the 6 total symptoms; as the number of GERD-related symptoms increased in our analyses, the degree of internal consistency and repeatability increased. GERD symptoms are difficult to distinguish from upper abdominal symptoms and commonly overlap with upper GI symptoms in actual practice. The composite score of the 4 main GERD symptoms showed higher correlation with nighttime GERD symptoms and the extra-esophageal symptoms, but showed relatively lower correlation with epigastric pain and soreness. These results demonstrate that the SEQ-GERD is particularly effective in distinguishing GERD from upper abdominal symptoms, such as dyspepsia. In terms of responsiveness, the

changes in symptom scores according to the OTE after the drug trial showed the clearest difference using the typical 2 symptoms and the 4 main GERD symptoms; the statistical significance became smaller as the number of symptoms increased. We determined that the 4 major GERD symptoms were more appropriate for inclusion in the PRO instrument than the 2 typical symptoms, as these were more sensitive and effective in judging therapeutic effect.

The GERD subscale of PAGI-SYM is made up of 7 items, combining bitter or sour taste in the mouth with 6 items including heartburn, chest discomfort and regurgitation during the day and when lying down respectively.[13] Compared to the 4 symptoms of our tool, the GERD domain of PAGI-SYM contains 3 questionnaires of GERD symptoms during lying down, and the response to the drug treatment may be different because of heterogeneous characteristics of nighttime symptoms. There is some need to compare these 2 questionnaire tools in future research involving more patients.

We also performed a linguistic validation of the Korean version of the GERD-QOL, which was a valid and highly reliable tool for the evaluation of GERD-specific QOL. The Korean version was easy to understand and showed high reliability and validity, which were not affected by cultural or ethnic variations.

There remain some limitations to the validation of the SEQ-GERD. The study population was heterogeneous with regard to endoscopic GERD grades. In Asia, endoscopic severe reflux esophagitis is only detected in small number of patients.[2,19,20] While symptomatic endoscopic reflux esophagitis patients are a subset of the overall GERD patient population, 50-80% of patients have NERD. In the present study, only 67% of patients had mild endoscopic reflux esophagitis and no patients had GERD of grade C and D according to the LA classification.

In conclusion, the present study demonstrated that the SEQ-GERD clearly exceeds the requirement for psychometric validation, as it has a high degree of reliability with high external validity compared with the PAGI-SYM, RDQ, and GERD-QOL. Moreover, this instrument shows strong evidence for accurately determining the therapeutic responsiveness to drug treatment.

Supplementary Material

Note: To access the supplementary table mentioned in this article, visit the online version of *Journal of Neurogastroenterology and Motility* at http://www.jnmjournal.org/, and at https://doi.org/10.5056/jnm18085.

References

1. Dent J, El-Serag HB, Wallander MA, Johansson S. Epidemiology of gastro-oesophageal reflux disease: a systematic review. Gut 2005;54:710-717.
2. Jung HK. Epidemiology of gastroesophageal reflux disease in Asia: a systematic review. J Neurogastroenterol Motil 2011;17:14-27.
3. Hershcovici T, Fass R. Nonerosive reflux disease (NERD) - an update. J Neurogastroenterol Motil 2010;16:8-21.
4. Zagari RM, Fuccio L, Wallander MA, et al. Gastro-oesophageal reflux symptoms, oesophagitis and Barrett's oesophagus in the general population: the loiano-monghidoro study. Gut 2008;57:1354-1359.
5. von Gruenigen VE, Huang HQ, Gil KM, et al. Assessment of factors that contribute to decreased quality of life in gynecologic oncology group ovarian cancer trials. Cancer 2009;115:4857-4864.
6. US department of health and human services food and drug administration. Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance 2006.
7. Services USDoHaH, Administration FaD. Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims. Available from URL: http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf. (assessed 1 september, 2018).
8. Speight J, Barendse SM. FDA guidance on patient reported outcomes. BMJ 2010;340:c2921.
9. Moss SF, Armstrong D, Arnold R, et al. GERD 2003 – a consensus on the way ahead. Digestion 2003;67:111-117.
10. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research; U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research; U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006;4:79.
11. Shaw M, Dent J, Beebe T, et al. The reflux disease questionnaire: a mea-

sure for assessment of treatment response in clinical trials. Health Qual Life Outcomes 2008;6:31.

12. Chan Y, Ching JY, Cheung CM, et al. Development and validation of a disease-specific quality of life questionnaire for gastro-oesophageal reflux disease: the GERD-QOL questionnaire. Aliment Pharmacol Ther 2010;31:452-460.

13. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med 1998;17:101-110.

14. Mesbah M, Cole BF, Lee MLT. Statistical methods for quality of life studies. Boston: Kluwer academic 2002:63-70.

15. U.S. Department of health and human services FDA center for drug evaluation and research. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006;4:79.

16. Mönnikes H, Pfaffenberger B, Gatz G, Hein J, Bardhan KD. Novel measurement of rapid treatment success with ReQuest: first and sustained symptom relief as outcome parameters in patients with endoscopy-negative GERD receiving 20 mg pantoprazole or 20 mg esomeprazole. Digestion 2007;75(suppl 1):62-68.

17. Nakada K, Matsuhashi N, Iwakiri K, et al. Development and validation of a simple and multifaceted instrument, GERD-TEST, for the clinical evaluation of gastroesophageal reflux and dyspeptic symptoms. World J Gastroenterol 2017;23:5216-5228.

18. Michielutte R, Bahnson J, Dignan MB, Schroeder EM. The use of illustrations and narrative text style to improve readability of a health education brochure. J Cancer Educ 1992;7:251-260.

19. Choi JY, Jung HK, Song EM, Shim KN, Sung SA. Determinants of symptoms in gastroesophageal reflux disease: nonerosive reflux disease, symptomatic, and silent erosive reflux disease. Eur J Gastroenterol Hepatol 2013;25:764-771.

20. Ko SH, Baeg MK, Jung HS, Kim P, Choi MG. Russian Caucasians have a higher risk of erosive reflux disease compared with East Asians: a direct endoscopic comparison. Neurogastroenterol Motil 2017;29:e13002.