DISEASES OF THE ESOPHAGUS

Original article

The development of the GERD-HRQL symptom severity instrument

V. Velanovich

Division of General Surgery, Henry Ford Hospital, Detroit, Michigan, USA

SUMMARY. **The Gastroesophageal Reflux Disease-Health Related Quality of Life (GERD-HRQL) instrument was introduced approximately 10 years ago to provide a quantitative method of measuring symptom severity in gastroesophageal reflux disease (GERD). Since that time the instrument has been used to assess treatment response to medication, endoscopic procedures, and surgery for GERD. However, the development of the instrument has progressed over the course of several years, and there is no one source which reviews this progress. The purpose of this article is to summarize the development and testing of the GERD-HRQL. The GERD-HRQL was initially developed to measure the typical symptoms of GERD. It was initially determined to have face validity and subsequent studies assessed its content validity, criterion validity, concurrent validity, predictive validity and construct validity. Reliability was determined by the test-retest method. Responsiveness was determined by the effects of treatment. This instrument is practical, with little administrative burden. There are few missing responses. Because there are 51 possible scores, the instrument has a high level of precision; and because of the response anchors, cannot have a floor effect, and only 4/372 patients reached the highest score of 50, implying little ceiling effect. The instrument has been translated into several languages, and appears valid, reliable and practical in each.**

KEY WORDS: **gastroesophageal reflux disease, quality of life instruments, symptom assessment, the GERD-HRQL.**

INTRODUCTION

The early 1990s saw the development and dissemination of laparoscopic antireflux surgery for the treatment of gastroesophageal reflux disease.[1,2] The outcomes were generally measured with qualitative scales of 'poor, fair, good, or excellent', or some derivation thereof. At that time, there were very few instruments specifically designed to measure symptom severity in GERD. This lack of a good instrument inhibited progress of GERD research due to the inability to quantitatively compare the magnitude of symptomatic improvement. Specifically, good symptom severity instruments allow the clinician or researcher to characterize the impact of GERD or its treatment in terms that are of value to the patient, may be used as independent predictors of surgical outcomes, may be indicators of the severity of disease, and can provide information on the quality of care.[3] In order to meet this need, the Gastroesophageal Reflux Disease-Health Related Quality of Life (GERD-HRQL) instrument was development to assess symptomatic outcomes for the typical symptoms of GERD. This instrument is one of the most frequently used of the symptom severity instruments, and has been recommended for use by the European Association for Endoscopic Surgery.[4] Nevertheless, the entire development of the GERD-HRQL has not been documented in one source, hence the purpose of this article is to trace its development with special attention to the important attributes of a quality of life instrument.

OBJECTIVE OF THE INSTRUMENT

The Scientific Advisory Committee of the Medical Outcomes Trust has put forward recommendations to assess health status and quality of life instruments.[5] One of the key attributes is the 'Conceptual and measurement model.' This is the rationale for and the description of the concept and the populations that a measure is intended to assess. The review

Address correspondence to: Vic Velanovich, MD, Division of General Surgery, K-8; Henry Ford Hospital; 2799 West Grand Blvd., Detroit, MI 48202-2689 USA. Email: vvelano1@hfhs.org

criteria include the concept to be measured, conceptual and empiric basis for item content, target population, information on dimensionality, evidence of scale variability, intended level of measurement, and rationale for deriving a scale score. Not all of these are needed for each instrument.

For the GERD-HRQL, the primary purpose was to measure symptomatic change as a result of medical or surgical treatment of GERD. The theoretical basis of the instrument was the quantification of the 'typical' symptoms of GERD with the target population being patients with GERD-like symptoms seeking medical attention. In the mid-twentieth century, objectively determining the presence of pathologic reflux was problematic. Not all patients with GERD-like symptoms in fact have GERD.[6] Therefore, failure to achieve good symptomatic results after antireflux surgery may have been due to inappropriate patient selection. This led to the development of physiologic testing for GERD with endoscopy, esophageal manometry, and 24 h esophageal pH testing.[7] And, in fact, the combination of the 'typical' symptoms of GERD with abnormal 24 h esophageal pH monitoring has been the best predictor of symptomatic improvement after antireflux surgery.[8] Given this physiologic 'gold standard' for the diagnosis and outcome prediction of GERD, it was felt that the symptom severity questionnaire should incorporate the 24-h esophageal pH monitoring criteria.

As stated, the primary purpose in the development of the GERD-HRQL was to measure symptomatic improvement of both medical and surgical treatment for GERD. However, there were secondary considerations used in instrument development. Specifically, practicality as reflected by low administrative burden and simplicity in scoring, and appropriateness as reflected by responsiveness and interpretability were considered highly desirable. In addition, another goal was to keep the instrument short, essentially to one page, to allow for ease of use by patients, and self-explanatory to the patient so that the use of research assistance was not necessary.

METHODS OF TESTING THE INSTRUMENT

The typical symptoms of GERD include heartburn and regurgitation, occurring both during the night, frequently waking the patient up from sleep, and also occurring during the day, frequently associated with meals. As the disease progresses, strictures can form leading to dysphagia. These symptoms have a great impact on a patient's quality of life.[9,10] The 24-h pH probe measures the total number of reflux episodes, the longest reflux episode, supine reflux, upright reflux, the total time the intraesophageal pH is < 4, and the longest reflux episode. Initially, nine items were chosen for the GERD-HRQL. These items were chosen by the author based on clinical interviews of dozens of patients with GERD to reflect the progressive severity of typical GERD. In this sense, the instrument as initially designed had *face validity*.[11] That is, the instrument appears to cover the issues of the disease as determined by those familiar with the disease. To this, an additional item related to bloating was added, as this is one of the side-effects of antireflux surgery and a measure of pre-existing gastroparesis, which commonly occurs in GERD patients. All subsequent studies have revalidated the GERD-HRQL with this additional item. In addition to the items reflecting the symptoms, an additional item was added with regard to use of medication as a measure of the effect of medication usage on quality of life, which has often been overlooked by other investigators. Table 1 presents the GERD-HRQL questionnaire. The final instrument contains a total of 10 scaled items which are scored, and a patient-reported global satisfaction assessment which is not added to the total GERD-HRQL score.[11]

Once the items of the questionnaire were chosen, a scaling system was needed to be devised to allow for increase in severity of symptoms to be appropriately quantified across patients. A main concern was that of floor and ceiling effects. The 'floor' effect is when a patient reports that he/she is at the lowest score possible as measured by the instrument,

Table 1   The Gastroesophageal Reflux Disease-Health Related Quality of Life instrument

---

• **Scale**: No symptoms = **0**; Symptoms noticeable, but not bothersome = **1**; Symptoms noticeable and bothersome, but not every day = **2**; Symptoms bothersome every day = **3**; Symptoms affect daily activities = **4**; Symptoms are incapacitating, unable to do daily activities = **5**

• **Questions**

__ 1. How bad is your heartburn?	0 1 2 3 4 5
__ 2. Heartburn when lying down?	0 1 2 3 4 5
__ 3. Heartburn when standing up?	0 1 2 3 4 5
__ 4. Heartburn after meals?	0 1 2 3 4 5
__ 5. Does heartburn change your diet?	0 1 2 3 4 5
__ 6. Does heartburn wake you from sleep?	0 1 2 3 4 5
__ 7. Do you have difficulty swallowing?	0 1 2 3 4 5
__ 8. Do you have pain with swallowing?	0 1 2 3 4 5
__ 9. Do you have bloating or gassy feelings?	0 1 2 3 4 5
__ 10. If you take medication, does this affect your daily life?	0 1 2 3 4 5
__ How satisfied are you with your present condition? Satisfied __ Neutral __ Dissatisfied __	

---

but later he/she reports symptoms beyond this lowest point. The 'ceiling' effect is at the opposite end of the scale. As an example of a ceiling effect is in the visual analog scale (VAS) for pain. In this instrument the worst possible score is 10 (from a scale of 0 to 10). Yet, the patient reports that his pain is 'an 11 out of 10.' This response cannot be measured by the VAS. Only four of 372 patients who completed the instrument scored 50, implying very little ceiling effect. In addition, it was important to insure that patients could understand the response scale. This was approached by having the numerical Likert-type responses attached to an anchor, whereby each patient can assess the severity of his or her own symptoms on an ordinal scale. Because the severity would be defined in laymen's terms, the responses would be standardized from patient to patient. The scale and anchors were defined in such a manner that zero was defined as no symptoms; (and therefore since patients could not be better than asymptomatic, the problem with the floor effect is avoided) and 5 was defined as 'incapacitating, unable to do daily activities' (and this was felt to be an adequate ceiling, since it is unlikely that patients could be any worse than completely incapacitated from the reflux). The figure also shows the scale with the anchors. The total GERD-HRQL score is derived by simply adding the individual item scores. No transformation of the raw scores to a scaled score is required, thereby insuring practicality. Therefore, the best possible total GERD-HRQL score is 0 (asymptomatic in all items) and the worst possible score is 50 (incapacitated in all items). Because the total GERD-HRQL score has 51 possible scores, it has a high level of precision,[12] especially compared to other GERD instruments. Item 11 pertaining to satisfaction has no numerical score and it is not reflected in the total GERD-HRQL score. This item should be interpreted as a patient-reported 'global' assessment of his or her present condition with respect to GERD.

VALIDITY OF THE GERD-HRQL

Whether the GERD-HRQL has validity or not has been questioned.[13] Let us look carefully at the types of validity used in quality of life research. Fayers and Machin[12] have defined three broad types of validity, and within these, subtypes. 'Content validity' relates to the adequacy of the content of the instrument to the quality of life characteristics it intends to measure. An aspect of content validity is 'face validity'; that is, whether the instrument appears to cover the issues of the disease as determined by those familiar with the disease (as mentioned above). The GERD-HRQL was intended to measure the typical symptoms of reflux; therefore,

patients had to feel that it measured the symptoms they were experiencing. Patients were asked to assess the GERD-HRQL and the SF-36 with the following questions:

1 Which questionnaire do you like best?
2 Which questionnaire was easier to understand?
3 Which questionnaire was more reflective with your problems of reflux?
4 Given the choice, which questionnaire would you rather fill out?

The GERD-HRQL was chosen more often by patients for all of these questions and particularly for question #3, 85% of patients felt that the GERD-HRQL better reflected their problems with reflux than the SF-36 and 68% of patients would rather fill out the GERD-HRQL rather than the SF-36.[14] From the standpoint of patient preference, the GERD-HRQL was a better questionnaire. More importantly, patients felt that the GERD-HRQL better reflected their problems with GERD and this supports the instrument's content and face validity.

'Criterion validity' involves measuring the instrument against a 'gold standard.' At the time the instrument was developed, there was no gold standard questionnaire for GERD. It was felt that the gold standard was physiologic assessment. Therefore, the instrument incorporates aspects of the physiologic goal standard of the 24-h pH probe; hence, it does have criterion validity in this sense. In addition, the instrument was assessed by comparing it to other physiologic standards such as endoscopically demonstrated esophagitis, results of the 24-h pH probe and esophageal manometry. Triadalfilopoulos[15] has shown that there is a correlation between question #1 of the GERD-HRQL (How bad is your heartburn) and the percentage of time that the pH < 4 by 24 h esophageal pH monitoring. In another study, it has been shown that as esophagitis grade increases, so does the total GERD-HRQL score.[16] This correlation with esophagitis grade and 24 h esophageal pH monitoring meet the criteria of criterion validity subtype of 'concurrent validity.' The total GERD-HRQL score did not correlate with the DeMeester score.[16] It is believed that this reflects the fact that only three of the items (#1–3) directly correlate to the aspects recorded by 24-h pH monitoring. Another subtype of criterion validity is 'predictive validity.' The GERD-HRQL has been shown to predict which patients would chose antireflux surgery and which patients would continue with medical management.[11]

Lastly, 'construct validity' is an assessment of the degree to which an instrument measures the theoretical construct that it was designed to measure. A subtype of construct validity is 'known-groups' validity; that is, it would be expected that similar groups would have similar scores and differing groups would have different scores. In the case of

GERD, symptomatic improvement is a primary outcome endpoint. This is why patients seek medical attention. Therefore, an instrument measuring GERD symptoms must be able to differentiate patients who are satisfied with their present level of symptoms and those who are not. The GERD-HRQL has been shown to do this.[11,14] In addition, we would expect patients who have had treatment for GERD to have better scores, which it true for the GERD-HRQL for both medical and surgical therapy,[11] and patients who have concomitant esophageal disorders to have worse scores, as was demonstrated with patients with non-specific esophageal motility disorders.[17]

'Concurrent validity' is agreement with the 'true' value. As this is not possible with most quality of life instruments because the 'true' value is not discernable, another way to address this is to compare them with other instruments. Instruments that measure the same phenomenon should have similar results. The GERD-HRQL was compared with another instrument which measures symptom severity, the quality of life questionnaire for patients undergoing antireflux surgery (QOLARS), and was found to correlate.[18]

Another subtype of construct validity is 'discriminant validity', in which instruments which do not measure the same aspects of quality of life would have scores which poorly correlate. When comparing the responses to the GERD-HRQL to the SF-36, it was shown using both univariate and multivariate analysis that the total GERD-HRQL score was a better predictor of patient satisfaction with level of reflux symptoms than the SF-36 and there was little correlation between the scores of the SF-36, and the GERD-HRQL.[14] Moreover, the range of scores had little overlap between the satisfied and the dissatisfied groups. Therefore, given these findings, validity of the GERD-HRQL has been assessed.

RELIABILITY OF THE GERD-HRQL

Reliability is the degree to which an instrument is free from random error.[5,12] Another way of stating this is that the instrument should give the same score at the same level of symptoms. Reliability of the instrument was assessed with the test-retest standard.[11] When patients at the same level of their reported symptom severity had retaken the test in two consecutive visits, the average difference of the scores was less than seven points. This difference was less than the difference between the total scores between the satisfied and dissatisfied patients. Therefore, there is stability in patient scores from test-to-test at the same level of patient-perceived symptoms.

RESPONSIVENESS OF THE GERD-HRQL

The strength of the GERD-HRQL is its sensitivity to change (responsiveness) to the effect of treatment. This attribute is an instrument's ability to detect change over time.[5] This characteristic is important in assessing the efficacy of treatments,[19] particularly surgical treatments.[20] The GERD-HRQL total score does improve (that is, reduces in score) with both medical and surgical treatment.[11] In addition, the magnitude of improvement reflects the initial severity of the score. The score shows that there is similar improvement in patients who have undergone laparoscopic versus open antireflux surgery both for the total score and for the first six items of the instrument individually. With respect to laparoscopy, there was also improvements in items number 8, 9, and 10. So for both medical, laparoscopic surgical treatment, and open surgical treatment, the total GERD-HRQL score and most of the individual item scores were responsive to improvements in patients' symptoms.[21] In addition, we see that when patients are less satisfied with antireflux surgery, such as those with chronic pain syndromes or psychoemotional problems,[22,23] the magnitude of the change is less. Also, the GERD-HRQL has been used in a number of studies evaluating new endoscopic treatments with similar responsiveness as with surgery.[24]

PRACTICALITY OF THE GERD-HRQL

Another strength in the GERD-HRQL is its practicality. The instrument has a total of 11 items, 10 of which are related to the scale and are included in assessing the total GERD-HRQL. Item number 11 is a global item related to patient satisfaction. The instrument is generally administered by simply handing it to the patient during an office visit or can be easily given over the phone in less than 2 minutes. Patients find it easy to understand and there are relatively few unanswered points when assessing the questionnaire. It has been my experience that the number of unanswered items is in the 1–2% range (unpubl. data). Few self-administered questionnaires have such a low unanswered question rate.

LIMITATIONS OF THE GERD-HRQL

Although the GERD-HRQL is an appropriate instrument to measure the severity of the typical symptoms of GERD, it does have important limitations. The GERD-HRQL is not appropriate for measuring the atypical symptoms of GERD. Specifically, there are no items for respiratory or laryngeal symptoms and none for chest pain as an

independent symptom from heartburn. Also, the GERD-HRQL does not measure the symptoms or effects of laryngopharyngeal reflux as a separate clinical entity. Other instruments have been developed for this purpose.[25] In addition, the GERD-HRQL is not an appropriate instrument for the measurement of the effects of GERD on lifestyle or other activities of daily living. There does exit another instrument which measures such problems.[26]

As the GERD-HRQL focuses on the typical symptoms of GERD, investigators may need to supplement its use with other quality of life (QoL) instruments. For example, to assess the effects of GERD on other aspects of QoL or to be able to assess the QoL effects of GERD as compared to other diseases, a generic instrument would be most appropriate. Such instruments as the SF-36, the Psychological General Well-Being, or the Sickness Impact Profile have been used in GERD and many other disease processes.[3] Therefore, whether to use the GERD-HRQL alone or in combination with other instruments will depend entirely on the purpose of the investigator or clinician.

CONCLUSION

In conclusion, the GERD-HRQL has found a place in the assessment of symptom severity in gastro-esophageal reflux disease. It is reliable, valid, and practical for this purpose. Further areas of research include additional comparisons with other instruments as well as further studies in the areas of physiologic testing.

References

1 Dallemagne B, Weerts J M, Jehaes C, Markiewicz S, Lombard R. Laparoscopic nissen fundoplication: preliminary report. Surg Laparosc Endosc 1991; 1: 138–43.

2 Hinder R A, Filipi C J, Wetscher G, Neary P, DeMeester T R, Perdikis G. Laparoscopic Nissen fundoplication is an effective treatment for gastroesophageal reflux disease. Ann Surg 1994; 220: 472–81.

3 Wood-Dauphinee S, Korolija D. Symptoms, health-related quality of life and patient satisfaction: using these patient-reported outcomes in people with gastroesophageal reflux disease. In: Granderath F A, Kamolz T, Pointner R, (eds). Gastroesophageal Reflux Disease: Principles of Disease, Diagnosis, and Treatment. Vienna: Springer, 2006; 269–86.

4 Korolija D, Sauerland S, Wood-Dauphinee S et al. Evaluation of quality of life after laparoscopic surgery: evidence-based guidelines of the European Association of Endoscopic Surgery. Surg Endosc 2004; 18: 879–97.

5 Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality of life instruments: attributes and review criteria. Qual Life Res 2002; 11: 193–205.

6 Costantini M, Crookes P F, Bremner R M et al. Value of physiologic assessment of foregut symptoms in a surgical practice. Surgery 1993; 114: 780–6.

7 Stein H J, DeMeester T R, Hinder R A. Outpatient physiologic testing and surgical management of foregut motility disorders. Curr Prob Surg 1992; 24: 415–555.

8 Campos G M, Peters J H, DeMeester T R et al. Multivariate analysis of factors predicting outcome after laparoscopic Nissen fundoplication. J Gastrointest Surg 1999; 3: 292–300.

9 Kamolz T, Pointner R, Velanovich V. The impact of gastroesophageal reflux disease on quality of life. Surg Endosc 2003; 17: 1193–9.

10 Kamolz T, Velanovich V. Quality of life in gastro-oesophageal reflux disease. Digest Liv Dis 2001; 33: 737–9.

11 Velanovich V, Vallance S R, Gusz J R, Tapice F V, Hurkabus M A. Quality of life scale for gastroesophageal reflux disease. J AM Coll Surg 1996; 183: 217–24.

12 Fayers P M, Machin D. Quality of life: assessment, analysis, and interpretation. Chichester: John Wiley & Sons, Ltd, 2000.

13 Borganokor M R, Irvine E J. Quality of life measurement in gastrointestinal and live disorders. Gus 2000; 47: 444–54.

14 Velanovich V. Comparison of generic (SF-36) vs. disease-specific (GERD-HRQL) quality-of-life scales for gastroesophageal reflux disease. J Gastrointest Surg 1998; 2: 141–5.

15 Triadalfilopoulos G. Changes in GERD symptom scores correlate with improvement in esophageal acid exposure after the Stretta procedure. Surg Endosc 2004; 18: 1038–44.

16 Velanovich V, Karmy-Jones R. Measuring gastroesophageal reflux disease: relationship between healthy-related quality of life scores and physiologic parameters. Am Surg 1998; 64: 649–53.

17 Velanovich V, Mahatme A. Effects of manometrically discovered nonspecific motility disorders of the esophagus on the outcomes of antireflux surgery. J Gastrointest Surg 2004; 8: 335–41.

18 Zeman Z, Rozsa S, Tarko E, Tihanyi T. Psychometric documentation of a quality of life questionnaire for patients undergoing antireflux surgery (QOLARS). Surg Endosc 2005; 19: 257–61.

19 Fitzpatrick R, Ziebland S, Jenkinson C et al. Importance of sensitivity to change as a criterion for selecting health status measures. Qual Health Care 1992; 1: 88–93.

20 Velanovich V. Using quality of life instruments to assess surgical outcomes. Surgery 1999; 126: 1–4.

21 Velanovich V. Comparison of symptomatic and quality of life outcomes of laparoscopic versus open antireflux surgery. Surgery 1999; 126: 782–9.

22 Velanovich V. The effect of chronic pain syndromes and psychoemotional disorders on symptomatic and quality of life outcomes of antireflux surgery. J Gastrointest Surg 2003; 7: 53–8.

23 Velanovich V. Using quality of life measurements to predict patient satisfaction outcomes for antireflux surgery. Arch Surg 2004; 139: 621–6.

24 Kamolz T, Velanovich V. The impact of disease and treatment on health-related quality of life in patients suffering from GERD. In: Granderath F A, Kamolz T, Pointner R, (eds). Gastroesophageal Reflux Disease: Principles of Disease, Diagnosis, and Treatment. Vienna: Springer, 2006; 287–98.

25 Westcott C J, Hopkins M B, Bach K et al. Fundoplication for laryngopharyngeal reflux. J Am Coll Surg 2004; 199: 23–30.

26 Liu J Y, Woloshin S, Laycock W S et al. Symptoms and treatment burden of gastroesophageal reflux disease: validating the GERD Assessment Scales. Arch Intern Med 2004; 164: 2058–64.