ORIGINAL ARTICLE

Taylor & Francis
healthsciences

Validation of a Brief Inventory for Diagnosis and Monitoring of Symptomatic Gastro-oesophageal Reflux

J. Zimmerman
Gastroenterology Unit, Hadassah University Hospital, Kiryat Hadassah, Jerusalem, Israel

Zimmerman J. Validation of a brief inventory for diagnosis and monitoring of symptomatic gastro-oesophageal reflux. Scand J Gastroenterol 2004;39:212–216.

Background: This article presents a brief inventory for the diagnosis and monitoring of GORD symptoms. **Methods:** The inventory consists of five items pertaining to different aspects of GORD to be graded for frequency on a 5-point Likert scale. It was validated on a consecutive group of GORD patients diagnosed either by endoscopy ($n = 25$) or by 24-h ambulatory pH monitoring (acid exposure time $\geq$5%; $n = 233$) and control subjects ($n = 300$). **Results:** Each of the inventory items was significantly associated with GORD ($P < 0.001$). Factor analysis indicated that all items loaded on a single scale with a high reliability (Cronbach alpha 0.88). Each item was weighted by its respective odds ratio in favour of GORD and a total score for the scale was then calculated as the sum of weighted scores on the five individual items. Receiver-operator (ROC) curve analysis implemented on a random sample comprising 67% of the group indicated that a total weighted score of 31.6 (percentage of the maximal possible weighted score) was 91% sensitive and 92% specific in the diagnosis of GORD. This score was then validated on the rest of the sample, where it correctly classified patients and controls with a sensitivity of 89% and specificity of 94%. The score proved stable on repeated administration in controls and in patients with stable symptoms, and decreased by 66% ($P < 0.001$) after 1 month of treatment with omeprazole in patients who reported symptom relief. **Conclusions:** The scale described in this article is a brief, simple and accurate measure, for the diagnosis of GORD as well as for monitoring its symptoms.

Key words: Diagnosis; GORD; heartburn; oesophagus; scales; validation study

Joseph Zimmerman, M.D., Gastroenterology Unit, Hadassah University Hospital, IL-91120 Jerusalem, Israel (fax. +972 2 642 1388, e-mail. zimmerj@vms.huji.ac.il)

Affecting some 19 million people in the United States, gastro-oesophageal reflux disease (GORD) is probably the most prevalent functional gastrointestinal disorder in the industrialized world (1, 2). GORD is associated with significant complications involving the oesophagus (3) as well as the upper and lower respiratory tract (4). Furthermore, it is an established risk factor for adenocarcinoma of the oesophagus (5). With an annual direct cost of $9.3 billion in the United States, it is also a significant economic burden, exceeding that incurred from gallbladder disease, colorectal cancer and peptic ulcer disease (1). Heartburn, the hallmark symptom of GORD, is common, with an incidence of 10%–48% (6, 7). However, only a fraction of the people complaining of heartburn actually suffer from GORD (8). Moreover, different persons may use the term 'heartburn' to describe disparate symptoms (9). Therefore the use of questionnaires describing the various symptoms associated with GORD in simple words rather in terms such as 'heartburn' has been proposed as a potential diagnostic aid in clinical practice (9). In this article I describe a brief inventory for the diagnosis and monitoring of GORD symptoms. It is shorter and simpler than previously published instruments. Its responsiveness to change in GORD symptoms renders it appropriate for clinical studies. Validation of this tool is described.

Methods

The protocol of this study was approved by the ethics committee of the Hadassah University Hospital.

Assessment of symptoms

The inventory used in this study comprises five items chosen from previously published literature (9) or obtained by discussion with GORD patients and with clinical experts. These items, pertaining to various aspects of GORD, are to be graded for frequency on a 5-point Likert scale, where '0' indicates an asymptomatic state and '4' indicates daily symptoms. The questionnaire is given in the appendix to this article.

Patients and controls

I studied prospectively a consecutive group of ambulatory patients and a group of control subjects. Exclusion criteria included use of H_2 blockers or proton-pump inhibitors,

DOI 10.1080/00365520310005333

pregnancy, known malignancy, previous upper digestive tract surgery and psychiatric disease.

The patient group consisted of symptomatic, untreated patients in whom GORD was established either by upper gastrointestinal endoscopy or by ambulatory 24-h pH monitoring. Patients were requested to fill out the questionnaires on their visit to the clinic, as part of the clinical assessment. Their demographic data and medical history were obtained by a direct interview, as well as from the medical records. The endoscopic criterion for diagnosis was erosive oesophagitis (at least grade A by the LA classification (10)). Patients undergoing pH monitoring had had normal findings on upper gastrointestinal endoscopy. For the purpose of this study, these patients were defined as suffering from GORD if the total acid exposure of the distal oesophagus was at least 5% of the monitoring period. Ambulatory pH monitoring was performed with the Synectics Digitrapper Mk III system (Synectics Medical AB, Stockholm, Sweden). The procedure was initiated in the fasting state, between 0800 and 1000 h. An antimony pH electrode was passed transnasally and positioned 5 cm above the upper border of the lower oesophageal sphincter. Acid reflux was defined as a drop in pH below 4.0. (11). Data were analyzed as previously described (11).

The control subjects were a random sample of university students in the Depts. of Humanities and Social Sciences, hospital administrative staff and laboratory personnel who agreed to fill out a health questionnaire anonymously for research purposes. Control subjects who had been investigated for GORD or dyspeptic symptoms, or who were currently taking H_2 blockers or proton-pump inhibitors or who reported having received these medications in the preceding year, were excluded. Both the patients as well as the control subjects were unaware of the purpose of this study. For the assessment of test–retest reliability, questionnaires were re-administered to a group of control subjects and patients with stable symptoms 14 days after they had filled out the first questionnaire. In patients, no change in medications has been made in the interim. For assessment of sensitivity to change in GORD symptoms, questionnaires were re-administered to patients after 1 month of therapy with the proton-pump inhibitor omeprazole (20 mg) daily. At that time, all reported amelioration of symptoms.

Statistical analysis

The association between individual items of the inventory and group (patients or controls) was investigated using logistic regression. To test for scale homogeneity, principal component analysis of the questionnaire items was performed. Scale reliability was checked using the Cronbach alpha coefficient. Since individual items differed in their association with GORD, each item was weighted by its odds ratio in favour of GORD (12). The weighted score on an item was its actual score multiplied by the weight of that item. A weighted total score for the scale, the sum of the weighted scores on the five individual items of the inventory, was computed. For further analysis, that score was expressed as a percentage of the maximal possible score on the weighted scale and then tested for its value in discriminating between patients and controls, for its stability and for its sensitivity to changes in symptoms following therapy. For the assessment of diagnostic value, the total sample of patients and controls was randomly divided into two subgroups, one comprising two-thirds of the sample (development group), and the other, the remaining one-third (validation group). Receiver-operator (ROC) curve analysis was carried out on the development group to determine the cut point score that discriminated best between GORD patients and controls. Validation of this classification was carried out on the validation group. Sensitivity and specificity measures were calculated for both groups. All statistical tests were two-tailed. A P value of 0.05 or less was considered to indicate statistical significance. The statistical analysis was performed using SPSS software (13).

Table I. Demographic data of patients and controls

Group	n	M/F	Smokers (%)	Age, mean $\pm s_{\bar{x}}$
GORD	258	171/87	18.3	42.5 ± 0.9
Control	300	145/155	18.6	37.5 ± 0.6

GORD = Gastro-oesophageal reflux disease.
$s_{\bar{x}}$ = standard error of the mean.

Results

Study population

The demographic data of the patients and controls are given in Table I. Of the GORD patients, 25 had an endoscopic diagnosis of at least grade A reflux oesophagitis according to the LA classification (10); the rest had endoscopically normal mucosa but on ambulatory pH monitoring were found to have a pH lower than 4 for at least 5% of the monitoring time (22.9 ± 0.8 h; mean $\pm$ standard deviation (s), range 20.5–24.0). The mean ($\pm$ standard error of the mean ($s_{\bar{x}}$)) acid exposure of the distal oesophagus was 12.6 ± 0.5% of the total monitoring period (range 5.0%–37.7%). None of the patients had been diagnosed with scleroderma or any other condition known to be associated with severe gastrointestinal dysmotility.

Table II gives the association of individual items with

Table II. Individual item association with GORD[1]

Item	Odds ratio	95% CI[2]	P
1	5.61	4.39–7.18	<0.001
2	4.15	3.33–5.18	<0.001
3	1.94	1.68–2.24	<0.001
4	4.05	3.25–5.06	<0.001
5	5.23	4.02–6.81	<0.001

[1]Odds ratios were calculated by logistic regression.
[2]Confidence intervals.

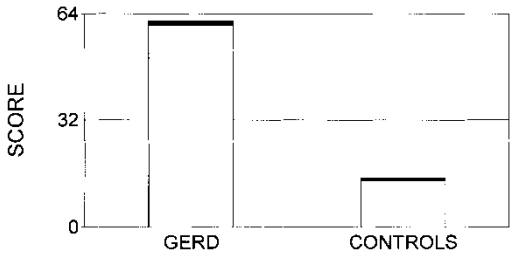

Fig. 1. Total scores on the weighted reflux scale in patients and controls. Scores are expressed as percentages of the maximal possible score (mean, standard error of the mean). The difference between patients and controls was significant ($P < 0.001$).

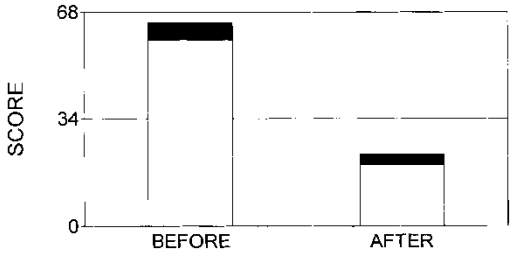

Fig. 3. Response to treatment. Questionnaires were re-administered to 14 patients after 1 month of therapy with omeprazole (20 mg) daily. All reported symptom relief. Scores are expressed as percentages of the maximal possible score on the weighted scale (mean, standard error of the mean). The decrease in mean scores was significant ($P < 0.001$, paired t test).

GORD. Each of the items in the inventory was strongly predictive of GORD. Principal components analysis of the inventory items showed that all loaded on a single factor, which accounted for 67.1 of the variance. Reliability of the scale formed by the five items was high, as indicated by Cronbach alpha coefficient of 0.88. Fig. 1 shows the weighted scale scores, expressed as a percentage of the maximal possible weighted score (a score of 84), in GORD patients and controls. The mean total score of GORD patients was 4-fold higher than that of controls ($P < 0.001$). In patients who underwent pH monitoring, the weighted scores correlated significantly with the total acid exposure time as well as with acid exposure in the upright position and in the 2-h postprandial state ($r = 0.20$, 0.19 and 0.13, adjusted for age, sex and smoking, $P = 0.002$, 0.005 and 0.047, respectively). The stability of the weighted scores is demonstrated in Fig. 2. A group of 28 control subjects and patients with stable symptoms filled out the questionnaire on 2 different occasions, 2 weeks apart. The two scores were similar and were highly correlated ($r = 0.94$; $P < 0.001$). For responsiveness to changes in symptoms, patients filled out the questionnaire before and at the end of a 1-month course of treatment with omeprazole (20 mg) daily. These patients reported amelioration of GORD symptoms. Following treatment and symptom relief, the mean scores decrease by $65.9\% \pm 5.7\%$ (mean individual percentage change value and $s_{\bar{x}}$). This decrease was highly significant (Fig. 3).

Fig. 4 shows the ROC curve of the weighted score, implemented for the classification of people either as GORD patients or controls on a random sample consisting of 371 cases (67% of the group). The area under the ROC curve was significantly higher than 0.5, corroborating the diagnostic value of the score. From the coordinates of this curve, it was determined that a score of 31.6 (percentage of the maximal possible score on the weighted scale) resulted in optimal classification of cases and controls, with a sensitivity of 90.5% and a specificity of 91.7% (Table III). That score was then validated on the remaining 187 cases, where it

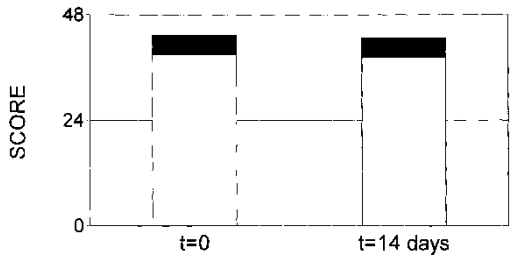

Fig. 2. Test–retest reliability of the total score on the weighted reflux scale. Questionnaires were re-administered to a random sample of 28 control subjects and patients with stable symptoms 14 days after they had filled out the questionnaire. In patients, no change in medications was made in the interim. Scores are expressed as percentages of the maximal possible score on the scale (mean, standard error of the mean). The correlation between the scores on the two occasions was high ($r = 0.94$; $P < 0.001$).

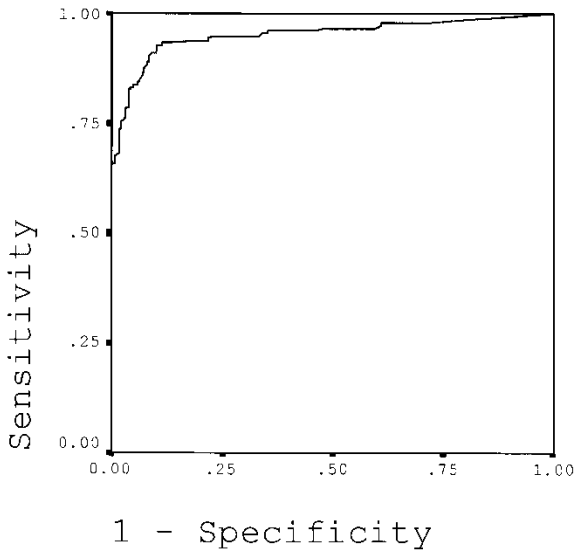

Fig. 4. ROC curve of the weighted total score (expressed as percentage of the maximal possible score on the weighted scale) in discriminating between GORD patients and controls. The area under the curve was 0.952 (standard error 0.012; 95% confidence interval 0.929–0.976; $P < 0.001$). A score of 31.6 resulted in optimal classification of cases and controls.

Table III. Diagnostic parameters of the reflux score[1]

	%	95% CI[2]
Sensitivity	90.5	87.5–93.5
Specificity	91.7	88.9–94.5
Positive predictive value	91.0	88.1–93.9
Negative predictive value	91.2	88.3–94.1

ROC curve analysis was implemented on a sample of 371 patients and controls. The weighted score that optimally discriminated between patients and controls was 31.6 (percentage of the maximal possible score on the weighted scale). This table compiles the diagnostic parameters using this score as a cut-off point.

95% confidence intervals.

correctly classified patients and controls with a sensitivity of 88.6% and a specificity of 94.4%.

Discussion

The increasing prevalence of GORD and its complications heightened interest in this condition as a public health problem. Diagnosis of GORD may be established by endoscopy, when typical erosive changes in the distal oesophagus are demonstrated (10). Endoscopy has a low sensitivity, but is highly specific for GORD when such changes are observed. Only 30%–40% of patients with GORD have oesophagitis (9, 14). In symptomatic patients in whom endoscopy yielded normal findings, ambulatory oesophageal pH monitoring may be considered the 'gold standard' (15). However, this procedure may, occasionally, yield inconsistent results (11, 15). Amelioration of symptoms during a therapeutic trial with a proton-pump inhibitor has been recently proposed as an additional diagnostic criterion (16, 17). Despite all these methods, a comprehensive review has underscored the importance of symptomatology in establishing the diagnosis of GORD (11). The routine use of short, self-administered questionnaires in clinical care has been recently advocated as an aid in the differentiation between GORD and dyspepsia (9).

Several papers describing the development and validation of various symptom inventories for diagnosis of GORD have been published recently (18–21). The highest sensitivity and specificity, 92% and 95%, respectively, were reported by Manetrola et al., who used ambulatory pH monitoring as the diagnostic criterion (21). However, external validation of the instrument was not reported. Moreover, the instrument used in that study was not self-administered and responsiveness to changes in symptoms was not evaluated. The questionnaire validated by Carlsson et al. (18) had a high sensitivity at 92%, but low specificity (19%) when the criteria for diagnosis were either erosive oesophagitis by endoscopy, or distal oesophageal pH below 4 for more than 4% of the 24-h monitoring period. Two other studies relied on specialty physician diagnosis as the 'gold standard' for diagnosis of GORD (19, 20). Differentiation between GORD and reflux-like dyspepsia on the basis of clinical history alone is problematic, even when elicited by a 'specialty physician'. The sensitivity and specificity of this diagnostic criterion in one study were 78% and 60%, respectively (22). Moreover, the scales of three of the above-mentioned diagnostic instruments (18–20) include items pertaining to the use of over-the-counter medications for digestive symptoms. While this approach may be warranted in a population where self-medication is highly prevalent, it may be inappropriate in a community where it is not.

In the present study, patients were defined as suffering from GORD only if they met stringent criteria. In addition to symptoms compatible with GORD, either unequivocal erosive changes in the distal oesophagus on endoscopy, or, in the absence of such changes, clear-cut evidence of acid reflux was requested for diagnosis. Gastro-oesophageal reflux is a physiological phenomenon of a continuous nature. The upper limit of normal for acid exposure of the distal oesophagus on ambulatory pH monitoring is 3.5% of the time (23). In order to exclude borderline cases, only patients in whom acid exposure was $\geq$5% were included in this study.

The instrument presented and validated in the present study is based on five items of high face validity. Its criterion validity was demonstrated by the tight association of each of the items with a diagnosis of GORD, as established either by endoscopy or by pH monitoring. Within patients undergoing pH monitoring, construct validity was further corroborated by the significant correlation between the extent of acid exposure of the distal oesophagus and the weighted total scores. The predictive performance of the cut point score in classifying the members of the validation group was excellent. Test–retest reliability was high, indicating stability of the weighted score, and sensitivity to change was demonstrated. This inventory compares with, or exceeds, other reported instruments in its diagnostic accuracy. It has the advantages of briefness, simplicity and self-administration. Its documented responsiveness to changes in symptoms renders it useful as a monitoring device in clinical practice and research.

Acknowledgement

The dedicated assistance of Yardena Yatzkan Israelit and Myrna Ittah is gratefully acknowledged.

References

1. Sandler RS, Everhart JE, Donowitz M, Adams E, Cronin K, Goodman C, et al. The burden of selected digestive disease in the United States. Gastroenterology 2002;122:1500–11.
2. Talley NJ, Boyce P, Jones M. Identification of distinct upper and lower gastrointestinal symptom groupings in an urban population. Gut 1998;42:690–5.
3. Mantynen T, Farkkila M, Kunnamo I, Mecklin JP, Juhola M, Voutilainen M. The impact of upper GI endoscopy referral volume on the diagnosis of gastroesophageal reflux disease and its complications: A 1-year cross-sectional study in a referral area with 260,000 inhabitants. Am J Gastroenterol 2002;97:2524–9.
4. el-Serag HB, Sonnenberg A. Comorbid occurrence of laryngeal

or pulmonary disease with esophagitis in United States military veterans. Gastroenterology 1997;113:755–60.
5. Lagergren J, Bergstrom R, Lindgren A, Nyren O. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. N Engl J Med 1999;340:825–31.
6. Nebel OT, Fornes MF, Castell DO. Symptomatic gastroesophageal reflux: incidence and precipitating factors. Am J Dig Dis 1976;21:953–6.
7. Heading RC. Prevalence of upper gastrointestinal symptoms in the general population: a systematic review. Scand J Gastroenterol 1999;34 Suppl 231:3–8.
8. Spechler SJ. Epidemiology and natural history of gastroesophageal reflux disease. Digestion 1992;51 Suppl 1:24–9.
9. Dent J. Definitions of reflux disease and its separation from dyspepsia. Gut 2002;50:iv17–iv20.
10. Lundell LR, Dent J, Bennett JR, Blum AL, Armstrong D, Galmiche JP, et al. Endoscopic assessment of oesophagitis: clinical and functional correlates and further validation of the Los Angeles classification. Gut 1999;45:172–80.
11. Kahrilas PJ, Quigley EM. Clinical esophageal pH recording: a technical review for practice guideline development. Gastroenterology 1996;110:1982–96.
12. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. Oxford, New York: Oxford University Press; 1995. p. 85–8.
13. Norusis MJ. SPSS 10.0. Chicago IL: SPSS Inc.; 1999.
14. Sonnenberg A, El-Serag HB. Clinical epidemiology and natural history of gastroesophageal reflux disease. Yale J Biol Med 1999; 72:81–92.
15. Dhiman RK, Saraswat VA, Naik SR. Ambulatory esophageal pH monitoring. Technique, interpretations and clinical indications. Dig Dis Sci 2002;47:241–50.
16. Johnsson F, Solhaug WJ-H, Hernqvist H, Bengtsson L. One-week omeprazole treatment in the diagnosis of gastroesophageal reflux disease. Scand J Gastroenterol 1998;33:15–30.
17. Fass R, Ofman JJ, Gralnek IM, Johnson C, Camargo E, Sampliner RE, et al. Clinical and economic assessment of the omeprazole test in patients with symptoms suggestive of gastroesophageal reflux disease. Arch Intern Med 1999; 159: 2161–8.
18. Carlsson R, Dent J, Bolling-Sternevald E, Johnsson F, Junghard O, Lauritsen K, et al. The usefulness of a structured questionnaire in the assessment of symptomatic gastroesphageal reflux disease. Scand J Gastroenterol 1998;33:1023–9.
19. Ofman JJ, Shaw M, Sadik K, Grogg A, Emery K, Lee J, et al. Identifying patients with gastroesophageal reflux disease: validation of a practical screening tool. Dig Dis Sci 2002; 47: 1863–9.
20. Shaw MJ, Talley NJ, Beebe TJ, Rockwood T, Carlsson R, Adlis S, et al. Initial validation of a diagnostic questionnaire for gastroesophageal reflux disease. Am J Gastroenterol 2001; 96: 52–7.
21. Manterola C, Muňoz S, Grande L, Bustos L. Initial validation of a questionnaire for detecting gastroesophageal reflux disease in epidemiological settings. J Clin Epidemiol 2002;55:1041–5.
22. Klauser AG, Schindlbeck NE, Muller-Lissner SA. Symptoms in gastro-oesophageal reflux disease. Lancet 1990;8683:205–8.
23. Kahrilas PJ, Pandolfino JE. Gastroesophageal reflux disease and its complications, including Barrett's metaplasia. In: Feldman M, Friedman LS, Sleisenger MH, editors. Sleisenger & Fordtran's gastrointestinal and liver diseases. 7th ed. Philadelphia: Saunders; 2002. p. 609–10.

Appendix 1

The Brief Reflux Inventory

How often do you experience the following?

(1) A feeling of pain, pressure or burning that starts in the stomach and spreads up the front of your chest.

(2) A burning sensation deep in the throat.

(3) A bitter, salty or sour taste in the mouth.

(4) A feeling that something you ate a while ago is coming back up.

(5) Are you ever woken up at night by a feeling of heartburn, coughing or choking?

Response scales:

0 = never; 1 = rarely; 2 = once a month to once a week; 3 = at least twice a week; 4 = daily.

Received 29 July 2003
Accepted 10 November 2003