Comparison of a composite symptom score assessing both symptom frequency and severity with a score that assesses frequency alone: a preliminary study to develop a practical symptom score to detect gastro-oesophageal reflux disease in a resource-poor setting

Lakmali D. Amarasiri[a], Arunasalam Pathmeswaran[b], Arjuna P. De Silva[c], Anuradha S. Dassanayake[d], Channa D. Ranasinha[d] and Janaka De Silva[c]

Downloaded from http://journals.lww.com/eurojgh by BhDMf5ePHKbH4TTImqenVH0e3gbzxl4SITb2nFaaBPTCCAlhrw2DrKwBWZ1Ia on 01/19/2020

Background There is a need for a practical instrument to screen for gastro-oesophageal reflux disease (GORD) in epidemiological studies.

Objectives To develop a practical score to detect GORD and compare assessment of both symptom frequency and severity with frequency alone.

Methods One hundred patients with upper gastrointestinal symptoms and 150 volunteers with no such past history faced an interviewer-administered questionnaire assessing seven symptoms, graded for frequency and severity. Two scores were generated. Score 1, the sum of frequency of symptoms and score 2, the sum of products of frequency and severity of each. Internal consistency, test–retest reliability and criterion validity against 24-h pH monitoring were assessed. Cut-off scores were generated by receiver operating characteristic curves using scores of half the volunteers and patients selected randomly and validated on the other half.

Results Cut-off scores and area under the curve for score 1 were ≥ 10.5 and 0.93, and score 2 were ≥ 12.5 and 0.93, respectively. The sensitivity and specificity of diagnosing the disease in the remaining participants using score 1 was 89.7 and 92.4% and score 2 was 93.8 and 94.0%, respectively. The instrument had good internal consistency (Cronbach $\alpha = 0.73$) and reliability (intraclass correlation coefficient of scores 1 and 2 were 0.94 and 0.95, respectively). Score 2 showed better correlation with 24-h pH monitoring parameters (Spearman's rank correlation, $P = 0.01$).

Conclusion Our score is valid, reliable and can detect GORD with high sensitivity and specificity. A score assessing both frequency and severity of symptoms correlates better to an objective measure of GORD. *Eur J Gastroenterol Hepatol* 22:662–668 © 2010 Wolters Kluwer Health | Lippincott Williams & Wilkins.

European Journal of Gastroenterology & Hepatology 2010, 22:662–668

Keywords: frequency, gastro-oesophageal reflux disease, reliability, scores, severity, symptom, validity

Departments of [a]Physiology, [b]Public Health, [c]Medicine and [d]Pharmacology, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka

Correspondence to Dr Lakmali D. Amarasiri, MBBS, Lecturer, Department of Physiology, Faculty of Medicine, University of Kelaniya, PO Box 6, Thalagolla Road, Ragama, Sri Lanka
Tel: +94 11 2961128; fax: +94 11 2958337; e-mail: lakmali_g@yahoo.com

Received 21 December 2008 Accepted 12 April 2009

Introduction

Gastro-oesophageal reflux disease (GORD) is a chronic, recurrent disease troublesome both symptom-wise and cost-wise, and greatly affects a patient's quality of life. Overall prevalence of GORD in the Western countries is estimated to be 20–40% and in Asia to be approximately 5–10% [1–3]. Currently, a clinical history of GORD symptoms coupled with upper gastrointestinal endoscopy and/or 24-h pH-metry is used to diagnose the disease [4]. The investigative methods of diagnosing GORD are often invasive, costly and not freely available. The simplest and most cost-effective method to diagnose GORD would be symptom assessment [5]. Furthermore, as GORD is a subjective disease, it can be presented even in the absence of a positive test [6,7]. Indeed GORD is diagnosed even without a positive-objective test, if the symptoms are severe enough to affect the patient's quality of life [8].

There is a need for a good discriminative tool for GORD to be used in an epidemiological setting. A discriminative instrument should be able to differentiate between individuals who have a disease and those who do not, and classify them according to their symptom severity and frequency [9]. The requirements of such a tool are that it must be simple (easy to understand and administer), short, less time-consuming, cost-effective and highly specific [10]. The tool should also have high

sensitivity in GORD patients, cover a range of symptoms, assess symptom frequency and severity, be responsive over short time intervals and be psychometrically validated and translated into many languages with cross-cultural adaptation [11]. A review of all available questionnaires shows that the 'ideal' questionnaire to diagnose GORD is yet to be developed [10,11].

GORD is a concept that is not precisely defined and there is no gold standard to measure it. Symptoms are measured by scales, where answers to a series of questions are combined to give a numerical score. When in measurement of something that is not precisely defined, and when there is no 'gold standard' for comparison, then it would be necessary to compare the scale against a 'near gold standard' or other physiological variables used in the field. This is known as criterion validity [12]. Twenty-four-hour pH monitoring can be considered as the 'near gold standard' in diagnosing GORD, as it has the highest sensitivity [13] and is the only precise method for quantifying the amount of acid reflux over the circadian cycle, it helps identifying patients with atypical presentations of GORD, allows accurate correlation of symptoms with acid reflux [6,14,15], and reproducibility of classification of GORD patients is good using pH monitoring [16].

Clinical impressions are that there is a high frequency of upper gastrointestinal complaints among Sri Lankans. Neither are there any published data on symptom patterns or prevalence of GORD, nor are there any well-defined population-based studies or valid instruments to measure GORD in Sri Lanka. Cultural differences required development of a locally valid and reliable instrument. This study presents a practical scoring system of GORD to be used in the community for future population prevalence studies and assesses whether using symptom frequency and severity rather than using only symptom frequency, correlates better with an objective measure of GORD.

Materials and methods

Study design

This was a cross-sectional validation study carried out at the Professorial Unit Medical Clinics and Endoscopy Unit of the Colombo North Teaching Hospital, Sri Lanka. Patient recruitment was carried out between January 2005 and December 2006. Ethical approval was granted by the Scientific and Ethics Review Committee of the Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka.

Participants, inclusion and exclusion criteria

One hundred consecutive patients who presented with upper gastrointestinal symptoms were recruited as cases. They all underwent upper gastrointestinal endoscopy. Reflux oesophagitis was used as an additional criterion for diagnosis of GORD as endoscopic appearances of the oesophageal mucosa can be used as hard criteria for the definition of the disease [17]. The controls were people who had no past history of upper gastrointestinal symptoms. These included either people accompanying the patients, staff of the hospital or friends and relatives of other patients seeking treatment in the hospital. Additional exclusion criteria were pregnancy, known upper gastrointestinal disease, prior upper digestive tract surgery or those who had taken antireflux medication within the past 7 days [18]. All cases and controls were assessed by a specialist physician to confirm the presence or the absence of upper gastrointestinal symptoms.

Sampling and study power

Sample size was calculated based on earlier studies and according to relevance to our population. Published values for sensitivity and specificity were 90% [18]. We expected our instrument to have similar values. To estimate this within 5% of the true sensitivity or specificity, we calculated that a sample size of 140 was needed in each group. We rounded up this figure to 150 for the controls. A feasibility study showed that owing to available resources, the number of cases maximally possible over the period that the study would be carried out would be 100. Therefore, we decided that a sample size of 100 patients and 150 controls would be required to achieve a sensitivity of 0.90 with a 95% confidence interval ± 0.06.

Development of the questionnaire

The questionnaire developed by us was based on a GORD-specific symptom questionnaire developed and validated by Allen *et al.* [9], and was modified to suit a Sri Lankan population. The questionnaire was developed in English, translated into Sinhalese and then translated back again and reviewed to ensure that the translation process had not altered the meaning.

Item selection

To select the items, we conducted a pilot survey among 50 randomly selected patients presenting to the Medical Clinics and Endoscopy Unit of the Colombo North Teaching Hospital. Questions on the frequency and severity of six symptoms adapted from an earlier questionnaire [9] were discussed with the patients and the volunteers were asked whether they experienced any other symptoms other than the six symptoms adapted from a preexisting questionnaire. The symptom 'belching' was included as a result of this. They were also asked to comment on any questions they had difficulty in understanding. The pilot study showed that most participants found it difficult to express the response to the disease. We used simple descriptive terms to describe symptoms (Table 1).

Table 1 Wording of description of symptoms[a]

Symptom	Description
Heartburn	Burning tightening sensation radiating up from the xiphisternum to the neck
Regurgitation	Episodes of bitter or sour fluid or food coming back to your mouth
Abdominal/chest pain	Upper abdominal pain or chest pain other than the burning pain in the chest or abdomen
Abdominal distension	Feeling of fullness in your upper abdomen
Dysphagia	Difficulty in swallowing
Cough	Episodes of cough during the day following a meal or at night or when you wake up
Belching	Episodes of wind coming up into your mouth

[a]All the symptoms were asked in relation to meals.

The final questionnaire comprised of questions on the frequency and severity of seven symptoms of reflux disease: heartburn, acid or food regurgitation, chest or upper abdominal pain, abdominal distension, difficulty in swallowing, cough and belching. The wording was simple, acceptable to a reading level of a 12-year-old child and unambiguous. We then pretested the questionnaire on 20 patients awaiting upper gastrointestinal endoscopy. For each symptom, the patient was asked two questions: 'how often do you have this problem?' to assess the frequency, and 'how much does this problem bother you?' to assess the severity or 'distress' to the patient [9]. The frequency was scored: 1, if symptoms were absent; 2, if symptoms occurred once a month; 3, if symptoms occurred once a week; 4, if symptoms occurred two to four times a week and 5, if symptoms occurred daily. The severity was scored: 1, if no discomfort at all; 2, if mild discomfort, yet not interfering with day-to-day activities; 3, if moderate discomfort, interfering with day-to-day activities at least once a week and 4, if severe discomfort, interfering with day-to-day activities and sleep more than once a week.

Scoring system

Two scores were generated. Score 1 was based on symptom frequency and was a construct of adding the frequency scores of each symptom with a total maximum score of 35 and a minimum score of 7. Score 2 was calculated as the product of symptom frequency and severity scores. The composite symptom score was the sum of individual symptom scores. This resulted in scores from 1 to 20 for each symptom, with a total maximum score of 140 and a minimum score of 7.

Study procedure

Cases and controls that fulfilled the inclusion and exclusion criteria participated in the study after written voluntary informed consent was obtained. A single investigator administered the questionnaire to all patients and controls. All the symptoms were asked in relation to meals and with a recall period of 4 weeks. A card with the responses was given to the patient (in addition to the investigator reading it out) and he or she was asked to choose the response from it. No patient was taking antireflux medication at the time of administering the questionnaire. The cases were either first-time diagnosed or had stopped medication for 7 days before diagnostic testing.

Measurement of physiological variables

All cases enrolled in the study underwent upper gastrointestinal endoscopy, after an overnight fast, by a trained physician using a fibreoptic endoscope (Olympus Inc., LVU20, Melville, New York, USA). The presence or absence of reflux oesophagitis was noted and graded according to modified Savary–Miller criteria [19].

Twenty-four-hour ambulatory oesophageal pH monitoring was performed on all patients using a monocrystalline antimony electrode (Synetics Medical AB, Stokholm, Sweden). The catheter was positioned 5 cm above the upper margin of the lower oesophageal sphincter, whose position was determined by stationary oesophageal manometry or the pull-through technique. The following parameters were assessed based on the score derived from DeMeester and Johnson: total number of reflux episodes, duration of the longest reflux episode, number of reflux episodes lasting longer than 5 min, total percentage of time $pH < 4$, time $pH < 4$ during upright exposure, time $pH < 4$ during supine exposure and the DeMeester score. A DeMeester score equal to or above 14.72 or a value equal to or above 4.45 for total percentage time $pH < 4$ or both were considered abnormal [15].

Assessment of symptom score

A cut-off score was generated and the symptom score was evaluated by assessing for reliability, consistency and validity.

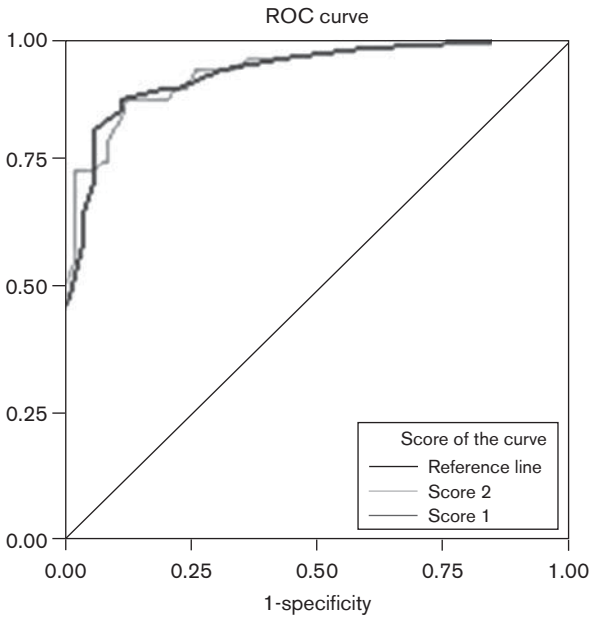
Cut-off score estimation

A cut-off score that best discriminated between these two groups was determined by constructing receiver operating characteristic (ROC) curves [20,21] using data from half the participants. Score values (composite symptom score and frequency-only score) of 51 cases and 74 controls that were selected randomly were utilized for this purpose (Fig. 1). The cut-off point was taken as the score with the highest acceptable sensitivity and specificity. The validity of the cut-off point was determined using data from the other half of the participants.

Assessment of reliability

Reliability indicates the extent to which a scale measures something in a reproducible manner. To assess the reliability, the questionnaire was administered to the cases twice 7 days apart. Intraclass correlation coefficient calculated by analysis of variance was used to compare the responses between days 1 and 7. An intraclass correlation coefficient of 0.7 or above indicates adequate reliability [22].

Fig. 1



Receiver operating characteristic (ROC) curves for scores 1 and 2.

Assessment of consistency

When developing a scale, it is necessary to ascertain that all items in the instrument measure the same thing, and are correlated to each other. Internal consistency was determined by using Cronbach α statistics. A Cronbach α value of 0.7–0.8 is considered to be sufficient to show internal consistency [23].

Assessment of validity

Validity was assessed by several components. Face validity and content validity were assessed by literature survey and expert review. Criterion validity was assessed by using 24-h oesophageal pH-metry as the gold standard, and correlating scale scores against the 24-h oesophageal pH monitoring parameters [12].

Statistical analysis

Participant characteristics and scores were described using descriptive statistics and reported as mean and standard deviation. Scores and physiological variables between test groups were compared by using unpaired *t*-tests. Correlation between symptom scores and physiological variables were compared by means of Spearman's rank correlation. Sample sizes were calculated using EpiInfo [version 6.04 (1996), Centres Disease Control and Prevention, Atlanta, Georgia, USA and World Health Organization, Geneva, Switzerland], and other statistical analyses by SPSS for Windows (version 10, SPSS Inc., Chicago, Illinois, USA). *P* values of less than 0.05 were considered as significant.

Results

Demographics

One hundred patients with upper gastrointestinal symptoms [mean age (range) 35.2 (15–60) years, 46% female] and 150 controls [mean age (range) 33.2 (17–62) years, 53% female] participated in the study. As expected, the mean of both scores 1 and 2 was significantly higher in patients than in controls ($P < 0.001$) (Table 2). The most frequent symptoms reported were heartburn and regurgitation (Table 3). Of the patients, 23 had normal upper gastrointestinal endoscopy and 77 had evidence of reflux oesophagitis.

Assessment of reliability

Both scores showed high reproducibility between scores on days 1 and 7 (Tables 4 and 5).

Assessment of consistency

Internal consistency of the scale items was found to be good with a Cronbach α value of 0.73.

Assessment of validity

Criterion validity was established by correlation of both scores with 24-h pH monitoring parameters of the 100 patients. Correlations above 0.5 were observed for several parameters in relation to score 1 and all parameters in relation to score 2 (Table 6).

Cut-off score

Using score values of 74 controls and 51 cases, who were randomly selected, score 1 (sum of frequency

Table 2 Clinical characteristics of patients and controls

Variable	Patients ($N=100$)	Controls ($N=150$)	*P* value
Mean age (range) (years)	35.2 (15–60)	33.2 (17–62)	0.179
Sex (% female)	46	53	0.256
Symptom score 1, mean (SD)	18.0 (5.3)	9.0 (2.7)	<0.001
Symptom score 2, mean (SD)	38.5 (22.7)	10.9 (5.9)	<0.001

SD, standard deviation.

Table 3 Symptom distribution according to frequency[a]

	Patients ($N=100$)				Controls ($N=150$)			
Symptom/ symptom frequency	Daily	Weekly	Monthly	Never	Daily	Weekly	Monthly	Never
Heartburn	54	22	11	13	3	18	19	110
Regurgitation	48	21	14	17	2	10	20	118
Upper abdominal/ chest pain	24	19	15	42	2	6	11	131
Abdominal distention	31	18	9	42	3	18	17	112
Dysphagia	9	4	6	81	0	0	4	146
Cough	10	5	8	77	0	2	3	145
Belching	18	7	26	49	7	6	22	115

[a]All values given as number of participants.

Table 4 Analysis of variance table for reproducibility of GORD score 1 in the 100 GORD patients

Source of variation	Sum of squares	d.f.	Mean square
Between people	5348.09	99	54.02
Within people	158.50	100	1.58
Between measures	0.60	1	0.60
Residual	157.89	99	1.59
Total	5506.59	199	27.67

Intraclass correlation coefficient for GORD score 1 = (between people − residual/[(between people + residual) + 2(between measures)] = (5348.09 − 157.89)/[(5348.09 + 157.89)+2(0.60)] = 0.94.
GORD, gastro-oesophageal reflux disease.

Table 5 Analysis of variance table for reproducibility of GORD score 2 in the 100 GORD patients

Source of variation	Sum of squares	d.f.	Mean square
Between people	95 124.42	99	960.85
Within people	2371.00	100	23.71
Between measures	128.00	1	128.00
Residual	2243.00	99	22.65
Total	97 495.42	199	489.92

Intraclass correlation coefficient for GORD score 2 = (between people − residual/[(between people + residual) + 2(between measures)] = (95 124.42 − 2243.00)/[(95 124.42 + 2243.00) + 2(128.00)] = 0.95.
GORD, gastro-oesophageal reflux disease.

Table 6 Correlation between scores and 24-h ambulatory oesophageal pH monitoring variables in gastro-oesophageal reflux disease patients (N = 100)

Physiological parameters	Versus score 1	Versus score 2[a]	P value
No. of total reflux episodes	0.572	0.676	0.019
No. of refluxes > 5 min	0.548	0.644	0.037
Longest episode (min)	0.500	0.622	0.010
Total reflux time pH < 4 (min)	0.491	0.651	0.001
Supine reflux time pH < 4 (min)	0.458	0.596	0.097
Upright reflux time pH < 4 (min)	0.442	0.526	< 0.01
DeMeester score	0.590	0.747	< 0.001

[a]Score 2 had a significantly better correlation with the 24-h pH parameters.

of the seven symptoms) had an area under the curve of 0.946, and score 2 (composite score of sum of frequency × severity of symptoms) had an area under the curve of 0.929. When the remaining 49 cases and 76 controls were classified as either abnormal or normal for upper gastrointestinal symptoms using cut-off scores of ≥ 11.5 for score 1 and ≥ 12.5 for score 2, we found that score 1 gave a correct classification of symptomatic GORD in 44 out of 49 cases with a sensitivity and specificity of 89.7 and 92.4%, respectively, and score 2 was able to correctly classify 46 out of 49 cases with a sensitivity and specificity of 93.8 and 94.0%, respectively.

Discussion

This study presents a reliable and valid 7-item questionnaire that is short, easy to administer and able to discriminate between patients and volunteers. It showed good inter-item correlation, and was found to have good test–retest reliability indicating stability of the score.

Several questionnaires have been developed and validated for diagnosis of GORD in epidemiological studies [24–34]. The questionnaire developed by Locke *et al.* [24] had 80 items and was considered too long to be used as a screening instrument. Other questionnaires that were short were either not GORD-specific [25], were GORD-specific but only assessed 'typical symptoms' [26,27], or assessed only frequency and did not include severity of symptoms [18,28,29]. Others that validated the questionnaires against a physician diagnosis [30,31] or the questionnaires on quality of life [32] did not seem objective enough. Out of those developed for diagnostic purposes, the Carllson and Dent questionnaire was shown to be diagnostically poor [33]. The Frequency Scale for Symptoms of GERD (FSSG) and modified FSSG questionnaires were validated against endoscopy as the gold standard [28,29]. However, endoscopy has low sensitivity (30–40%) for GORD [15]. In comparison, 24-h ambulatory pH monitoring has a higher sensitivity (75–96%) and specificity (60–100%) [15] for GORD, and many studies have used it as the gold standard for validation [10,18,34]. We also used ambulatory pH monitoring for validation of our questionnaire, and the symptom scores of the patients showed good correlation with 24-h oesophageal pH-metry parameters (mainly DeMeester score and total percentage time pH < 4). We did not perform either upper gastrointestinal endoscopy or 24-h pH-metry on volunteers, as we deemed it unethical to subject them to invasive testing. We used endoscopy as an additional criterion for diagnosis of GORD [17]. The main problem we encountered during the development of the questionnaire was in the selection of items. There are no specific symptoms of GORD and the existing symptoms may be difficult to express or explain, because of much cultural variability. Although GORD has a whole spectrum of symptomatology, ours included only seven symptoms to keep the screening instrument simple. Heartburn and regurgitation were included as 'typical' symptoms [35]. Chest pain or upper abdominal pain, abdominal distension, dysphagia and cough were included as 'atypical' symptoms. 'Belching' was included as a high proportion of patients attending our gastroenterology units complained of this (as shown by the pilot survey). This instrument was intended to be used as a case-finding tool to be used in other populations of individuals (e.g. asthma patients), in whom atypical symptoms are more prevalent. Therefore, equal weightage was given to all symptoms.

We tested the questionnaire mainly in patients suffering from erosive GOR. Twenty-three patients had had no oesophageal evidence of GORD and 13 had no heartburn.

Those with no oesophagitis had one or both 'typical symptoms'. Eleven of those without heartburn had regurgitation, whereas the other two had evidence of grade I oesophagitis on endoscopy. As all patients had either symptoms or evidence of oesophagitis or both, no participants without oesophagitis or typical symptoms were included as GORD patients in the study.

Although self-administered questionnaires are more convenient, we used a single external investigator to improve the quality and accuracy of data, and to increase the reliability. To minimize errors in responding, we made the questionnaire as simple as possible by providing a short recall period of 4 weeks, using simple wording and providing a cue card. Factor analysis was not performed as the symptoms were obtained directly from a previously validated questionnaire. Predictive values were not quoted, as predictive values depend on a known prevalence and the objective test used in diagnosis. The exact prevalence of GORD in Sri Lanka is unknown. We had used a contrived sample and not a random sample of the population.

Once reliability and validity of the instrument was established, the best cut-off point was determined to assess the ability of the instrument to discriminate between individuals who had upper gastrointestinal symptoms necessitating further action or treatment from those who did not. The performance of the instrument termed as 'diagnostic accuracy' is given by the area under the ROC curve. This area is equal to the probability that a random person with the disease has a higher value of the measurement than a random person without the disease. The score with the higher area under the curve would clearly be the better one. For epidemiological studies, it is important to consider the highest sensitivity when detecting a disease; hence, the cut-off points selected by us for the scores were those that gave the highest sensitivity and specificity. The sensitivity and specificity values obtained by us are similar to that achieved in similar studies [32,34]. Generation of ROC curves was also useful for us to compare the two scores (frequency-only vs. composite score of frequency and severity).

Earlier studies have validated questionnaires using scales with either symptom frequency-alone or with composite scores of symptom frequency and severity [9,18,28,29]. To the best of our knowledge, this is the first study that compares the two types of scores. Although we could not show that one score was superior to the other in detecting GORD cases, we found that consideration of severity of symptoms in addition to frequency correlated better to an objective measure of GORD.

In conclusion, our questionnaire is reliable and valid. A score using both symptom frequency and severity correlates better with an objective measure, such as 24-h ambulatory pH monitoring. We recommend that it can be used as a case-finding tool in epidemiological studies, once its validity in the community setting is established.

Acknowledgements

The authors thank Ms J. Liyanage and Ms R. Subasinghe for assistance with 24-h pH monitoring and oesophageal manometry. This study was supported by research grant RG/2004/M/07 of the National Science Foundation of Sri Lanka.

Conflicts of interests: All the authors declare that there are no financial relationships, academic competition or intellectual commitments, which might bias the work or interfere with objective judgment and results.

References

1 Locke GR, Talley NJ, Fett SL, Zinsmeister AR, Melton LJ III. Prevalence and clinical spectrum of gastro-oesophageal reflux. A population based study in Olmsted County, Minnesota. *Gastroenterology* 1997; **112**:1448–1456.
2 Ho K. Gastro-oesophageal reflux disease is uncommon in Asia: evidence and possible explanations. *World J Gastroenterol* 1999; **5**:4–6.
3 Goh KL. Changing epidemiology of gastroesophageal reflux disease in the Asian-Pacific region: an overview. *J Gastroenterol Hepatol* 2004; **19** (**Suppl 3**):S22–S25.
4 De Vault KR, Castell DO. Updated Guidelines for the diagnosis and treatment of gastro-oesophageal reflux disease. *Am J Gastroenterol* 1999; **94**:1434–1442.
5 McColl E. Best practice in symptom assessment: a review. *Gut* 2004; **53**:49–54.
6 Richter JE, Singh S. Ambulatory esophageal pH monitoring in gastroesophageal reflux disease. *Can J Gastroenterol* 1992; **6**:155–164.
7 Dent J. Definitions of reflux disease and its separation from dyspepsia. *Gut* 2002; **50** (**Suppl IV**):iv17–iv20.
8 Bytzer P. Assessment of reflux symptom severity: methodological options and their attributes. *Gut* 2004; **53**:28–34.
9 Allen CJ, Parameswaran K, Belda J, Anvari M. Reproducibility, validity, and responsiveness of a disease-specific symptom questionnaire for gastrooesophageal reflux disease. *Dis Esophagus* 2000; **13**:265–270.
10 Stanghellini V, Armstrong D, Monnikes H, Bardhan KD. Systematic review: do we need a new gastro-oesophageal reflux disease questionnaire? *Aliment Pharmacol Ther* 2004; **19**:463–479.
11 Fass R. Symptom assessment tools for gastroesophageal reflux disease (GERD) treatment. *J Clin Gastroenterol* 2007; **41**:437–444.
12 Streiner DL, Norman GR. Validity. In: Streiner DL, editor. *Health measurement scales. A practical guide to their development and use*. New York: Oxford University Press; 1995, pp. 104–126.
13 Jiang SP, Huang LW. Role of gastroesophageal reflux disease in asthmatic patients. *Eur Rev Med Pharmacol Sci* 2005; **9**:151–160.
14 De Meester TR, Johnson LF, Joseph GJ, Toscano MS, Hall AW, Skinner DB, *et al*. Symptoms of Gastro-oesophageal reflux in health and disease. *Ann Surg* 1976; **184**:459–470.
15 Jamieson JR, Stein HJ, DeMeester TR, Bonavina L, Schwizer W, Hinder RA, *et al*. Ambulatory 24-h esophageal pH monitoring: normal values, optimal thresholds, specificity, sensitivity, and reproducibility. *Am J Gastroenterol* 1992; **87**:1102–1111.
16 Johnsson F, Joelsson B. Reproducibility of ambulatory oesophageal pH monitoring. *Gut* 1988; **29**:886–889.
17 Fuchs KH, DeMeester TR, Albertucci M. Specificity and sensitivity of objective diagnosis of gastroesophageal reflux disease. *Surgery* 1987; **102**:575–580.
18 Manterola C, Munoz S, Grande L, Bustos L. Initial validation of a questionnaire for detecting gastro-oesophageal reflux disease in epidemiological settings. *J Clin Epidemiol* 2002; **55**:1041–1045.
19 Savary M, Miller G. *The esophagus: handbook and atlas of endoscopy*. Solothurn, Switzerland: Verlag Gassman; 1978. pp. 135–142.

20 Streiner DL, Norman GR. From items to scales. In: Streiner DL, editor. *Health measurement scales. A practical guide to their development and use*. New York: Oxford University Press; 1995. pp. 96–103.
21 Altman DG, Bland J. Diagnostic tests 3: receiver operating characteristic plots. *BMJ* 1994; **309**:188.
22 Streiner DL, Norman GR. Reliability. In: Streiner DL, editor. *Health measurement scales. A practical guide to their development and use*. New York: Oxford University Press; 1995. pp. 144–157.
23 Bland JM, Altman DG. Cronbach's alpha. *BMJ* 1997; **314**:572.
24 Locke GR, Talley NJ, Weaver AL, Zinsmeister AR. A new questionnaire for GORD. *Mayo Clin Proc* 1994; **69**:539–547.
25 Revicki DA, Wood M, Wiklund I, Crawley J. Reliability and validity of the Gastrointestinal Symptom Rating Scale in patients with gastroesophageal reflux disease. *Qual Life Res* 1998; **7**:75–83.
26 Shaw M, Talley NJ, Adlis S, Beebe T, Tomshine P, Healey M. Development of a digestive health status instrument: tests of scaling assumptions, structure and reliability in a primary care population. *Aliment Pharmacol Ther* 1998; **12**:1067–1078.
27 Offman JJ, Shaw M, Sadik K, Grogg A, Emery K, Lee J, *et al*. Identifying patients with gastro-oesophageal reflux disease: validation of a practical screening tool. *Dig Dis Sci* 2002; **47**:1863–1869.
28 Kusano M, Shimoyama Y, Sugimoto S, Kawamura O, Maeda M, Minashi K, *et al*. Development and evaluation of FSSG: frequency scale for the symptoms of GERD. *J Gastroenterol* 2004; **39**:888–891.
29 Shimoyama Y, Kusano M, Sugimoto S, Kawamura O, Maeda M, Minashi K, *et al*. Diagnosis of gastroesophageal reflux disease using a new questionnaire. *J Gastroenterol Hepatol* 2005; **20**:643–647.
30 Shaw MJ, Talley NJ, Beebe TJ, Rockwood T, Carlsson R, Adlis S, *et al*. Initial validation of a diagnostic questionnaire for gastro-oesophageal reflux disease. *Am J Gastroenterol* 2001; **96**:52–57.
31 Kleinman L, Rothman M, Strauss R, Orenstein SR, Nelson S, Vandenplas Y, *et al*. The infant gastroesophageal reflux questionnaire revised: development and validation as an evaluative instrument. *Clin Gastroenterol Hepatol* 2006; **4**:588–596.
32 Wong WM, Lam KF, Lai KC, Hui WM, Hu WHC, Lam CLK, *et al*. A validated symptoms questionnaire (Chinese GERDQ) for the diagnosis of gastro-oesophageal reflux disease in the Chinese population. *Aliment Pharmacol Ther* 2003; **17**:1407–1413.
33 Numans ME, De Wit NJ. Reflux symptoms in general practice: diagnostic evaluation of the Carlsson–Dent gastro-oesophageal reflux disease questionnaire. *Aliment Pharmacol Ther* 2003; **17**:1049–1055.
34 Zimmerman J. Validation of a brief inventory for diagnosis and monitoring of symptomatic gastro-oesophageal reflux. *Scand J Gastroenterol* 2004; **39**:212–216.
35 Klauser AG, Schindlbeck NE, Muller-Lissner SA. Symptoms in gastrooesophageal reflux disease. *Lancet* 1990; **335**:205–208.