ELSEVIER

Journal of Clinical Epidemiology 55 (2002) 1041–1045

Journal of Clinical Epidemiology

Initial validation of a questionnaire for detecting gastroesophageal reflux disease in epidemiological settings

Carlos Manterola[a,b,*], Sergio Muñoz[b], Luis Grande[c], Luis Bustos[b]

[a]*Department of Surgery, Universidad de La Frontera, Casilla 54-D, Temuco, Chile*
[b]*Clinical Epidemiology Unit, Faculty of Medicine, Universidad de La Frontera, Temuco, Chile*
[c]*Department of Surgery, Universidad de Barcelona, Barcelona, Spain*

Received 25 September 2001; received in revised form 6 June 2002; accepted 6 June 2002

Abstract

There is a high prevalence of gastroesophageal reflux disease (GERD) in the general population. Prevalence studies are scarce, and there is a lack of valid instruments for measuring them. The aim of this paper is to validate a questionnaire for detecting GERD. A validity study design with pathologic GERD patients and controls was used. A sample of 240 subjects age and sex paired was selected in the ratio of 3:1 (patients to controls). The initial structured questionnaire contained a variety of GERD symptoms. Internal consistency, interobserver reliability, criteria validity using 24-h esophageal pH monitoring, construct validity, and extreme group validation were assessed. Sensitivity, specificity, and predictive values were also obtained in different cutoff points of the definitive scale. A total of 180 confirmed GERD patients and 60 controls were included in the study. Mean age in years was 45 ± 13, with no statistical difference by gender (67% were female). Internal consistency of 0.75 and interobserver reliability of 0.87 was achieved in building the scale. Extreme group validation was highly significant by assessing the scale score with 24-h esophageal pH monitoring ($P < .0001$). At cutoff point 3 of the scale and with a correct classification of subjects of 92.4%, sensitivity, specificity, positive, and negative predictive values were 92, 95, 98, and 79%, respectively. The conclusion of this article is that a reliable and valid instrument was built to detect GERD.

Keywords: Esophagus; Gastroesophageal reflux disease; Scales; Validity; Reliability

1. Introduction

Gastro-esophageal reflux disease (GERD) is an increasing public health problem due to length and quality of life expectations of the population. Even though GERD symptoms may not be recognized by a proportion of the population, considered as normal, 44% of adults report monthly and 7% daily symptoms [1,2]. Studies in symptomatic consulting subjects show a prevalence of esophagitis due to reflux ranging from 5 to 10% [3], and of Barrett esophagus from 1 to 5% [4,5] .

Different clinical manifestations may be presented in GERD. Typical symptoms are heartburn, regurgitation, and dysphagia, while nontypical manifestations are nocturnal cough, dysphonia, asthma, chest pain, and other less frequent and nonspecific symptoms. Patients may consult either in the early stages of the disease or never, assuming that GERD symptoms are normal [6].

The most important and least expensive mechanism for diagnosing GERD is the clinical history, and to complete the evaluation of these patients several diagnostic tests are available. However, the test with the greatest sensitivity and specificity, and therefore considered as the gold standard, is the 24-h esophageal pH monitoring. This method physiologically quantifies the presence of GERD, allowing correlation between the symptoms and the changes in esophageal pH, and reflects esophageal clearance [7].

Reports on the prevalence of GERD are scarce, and studies using tools to measure and detect GERD are even less typical [1,8–17], although studies evaluating health-related quality of life for heartburn and GERD have recently been published and facts evaluated [18,19]. The objective of this article is to present a valid, reliable questionnaire to determine GERD for later application in population prevalence studies.

2. Material and methods

2.1. Design

The design was a cross-sectional validity study.

* Corresponding author. Tel.: 56-45-325760; fax: +56-45-325761.
E-mail address: cmantero@ufro.cl (C. Manterola).

0895-4356/02/$ – see front matter
PII: S0895-4356(02)00454-7

2.2. Subjects

Two groups of subjects were included in the study. Cases were patients having requested medical consultation, diagnosed with GERD when they fulfilled at least two of the following criteria: clinical, radiologic, endoscopic, and histologic. Controls were relatives of GERD patients with an absence of upper digestive symptoms. Inclusion criteria were to be resident of the city of Temuco, Chile, aged 15 or more, and who gave informed consent to participate in the study. Exclusion criteria were to be pregnant, to have either malignant or benign esophageal disease, GERD-related disease, gastric and duodenal disease, previous upper digestive tract surgery, and mental disease.

2.3. Sample

The sample size was determined according to feasibility criteria [20]. A total of 10 subjects per item in the questionnaire were included. The questionnaire initially consisted of 11 items, requiring a minimum sample size of 110 subjects. The final sample size was 240 subjects with a patient/control ratio of 3:1, matched by sex and age (matching criteria was used because the disease is sex and age related). Patient recruitment was done from January 1998 to December 2000 in a reference health center ("Hospital Clínico de La Frontera," Temuco, Chile). A prior pilot study was conducted in similar subjects to detect errors in item construction and/or item comprehension [20]. A preliminary report of validity and reliability studies of the scale was published [21].

2.4. Measurements

A structured questionnaire was designed as the basis for scale development. Items were selected from review of the literature and consultation to thematic experts from Chile, Spain, Uruguay, the United Kingdom, and Italy.

The instrument was divided into the following three sections: bioclinical characteristics, common GERD symptom frequency (heartburn, regurgitation, and dysphagia), and uncommon GERD symptom frequency (chest pain, nocturnal cough, dysphonia, and asthma). Two versions of the instrument with permuted questions were used at the pilot phase of the questionnaire to minimize the likelihood of bias in answer tendency [22].

2.5. Study protocol

Subjects fulfilling the inclusion/exclusion criteria were included in the study after voluntary informed consent was signed. The Hospital Ethics Committee approved the study. Two blinded clinical staff members, specially trained for this study, having at least 2 years of experience in the field, applied the instrument twice within 24 h: first for use in the study, and second to confirm reliability (both applications were made before GERD diagnosis was confirmed). Twenty-four-hour esophageal pH monitoring was done by an external specialist with clinical experience using a monocrystalline antimony pH catheter (Synectics Medical AB, Stockholm, Sweden), which was positioned 5 cm above the upper margin of the lower esophageal sphincter previously located by manometry. The reference electrode was attached to the upper chest surface after light skin abrasion and application of conductive gel (Elektroden gel; PPG, Hellige, Germany). Data acquisition was performed using a portable data logger (Diggitrapper MD; Synectics Medical AB, Stockholm, Sweden). Patients were instructed to note meals, snacks, supine periods, and occurrence of symptoms, using a written diary. The following variables were included: total number of reflux episodes, total number of episodes >5 min., the length of the longest episode, total time of pH < 4.0, time of pH < 4.0 in supine position, percent reflux time, percent reflux time at night, reflux index, and the DeMeester score, as variables with high discriminatory capacity to diagnose GERD [7].

2.6. Reliability assessment

Exploratory factor analysis was done to determine the number of underlying factors associated with GERD symptoms. Internal consistency was determined by using Cronbach alpha statistics [23,24]. Kappa statistics were used to assess interobserver reliability (minimum value = 0.70) [25,26].

2.7. Validity assessment

Validity was evaluated in different ways as suggested by several authors [20,27,28]. Content and face validity were assessed by literature review of clinical manifestations of GERD, analyzing other available GERD scale tools, and ex-

Table 1
Bioclinical variables in patients and controls ($N = 240$)

	Controls ($N = 60$)		Cases ($N = 180$)		
Variable	Average	95% CI	Average	95% CI	p
Age (years)	47.0	(43.8, 50.1)	44.7	(42.6, 46.7)	.2401
Weight (kg)	68.1	(65.6, 70.6)	67.5	(65.6, 69.5)	.7665
Height (cm)	164.6	(162.4, 166.8)	162.8	(161.5, 164.1)	.1694
Length of symptoms (months)	7.9	(3.5, 12.2)	38.3	(34.1, 42.5)	.0002
Symptoms score (points)[a]	0.5	(0.2, 0.8)	6.6	(6.1, 7.0)	.0001
Sex, case number (%)					.8950
Female	40 (67)		120 (67)		
Male	20 (33)		60 (33)		

[a]Explanation for this variable may be observed in Table 4.

Table 2
Items distribution according to appearance frequency

Items	Frequency (%)			
	Daily	Weekly	Monthly	Never
Heartburn	34.2	19.6	12.5	33.7
Regurgitation	35.8	19.6	14.2	30.4
	Permanent	Occasional	Never	
Chest pain	5.4	31.3	63.3	
Nocturnal cough	17.9	25.8	56.3	
		Yes	No	
Dysphagia		36.7	63.3	
Dysphonia		35.4	64.6	
Asthma		14.2	85.8	

pert consultation. Criteria validity was studied vs. 24-h esophageal pH monitoring. Regression analysis was performed to correlate scale values with the DeMeester score. Construct validity was determined by extreme group comparison. Concurrent validity was assessed vs. 24-h esophageal pH monitoring variables. Logistic regression was used to obtain sensitivity, specificity, and predictive values to establish the final scale cutoff point.

3. Results

Two hundred and forty subjects were studied—180 patients with GERD and 60 controls; 160 females and 80 males—same case/control proportion. The mean age was 45 years, with an average weight of 67.7 kg and a mean height of 163.2 cm. Obviously, the length and the severity of symptoms significantly differed in cases and controls (Table 1).

The most frequent symptoms reported were regurgitation (69.6%) and heartburn (66.3%), while asthma was reported by a minority of patients (Table 2). When analyzing the number of symptoms reported by both groups, 72.9% of the controls did not manifest symptoms, in comparison with 80.5% of the cases who reported at least three symptoms, regardless of their frequency ($P < .001$).

3.1. Reliability

Cronbach alpha value for internal consistency of the scale items was 0.75, considered very good because the rule of thumb of 0.70 was overpassed. Interobserver reliability study was conducted by two independent observers with two applications of the scale to the same study subject. Table 3 shows an agreement degree >90% and kappa values >0.87. On establishing internal consistency and item reliability, a definitive scale was built for all the items with a score ranging from 0 to 13 (according to the existence of symptoms and their frequency). The mean value in the whole series was 5.1. The application of factor analysis with Varimax rotation allowed to verified that a single scale is generated, to which the following items conformed: heartburn, regurgitation, dysphagia, chest pain, nocturnal cough, dysphonia, and asthma (Table 4).

Table 3
Interobserver reliability study (kappa) ($N = 240$)

Items	Agreement (%)	Kappa
Heartburn	90.8	0.87
Regurgitation	91.2	0.88
Dysphagia	98.3	0.96
Chest pain	96.6	0.93
Nocturnal cough	97.1	0.95
Dysphonia	96.6	0.93
Asthma	98.3	0.93

Table 4
Scale and score

Items	Frequency	Score
Heartburn	Daily	3
	At least once per week	2
	At least once per month	1
	Never	0
Regurgitation	Daily	3
	At least once per week	2
	At least once per month	1
	Never	0
Dysphagia	Yes	1
	No	0
Chest pain	Permanent	2
	Occasional	1
	Never	0
Nocturnal cough	Permanent	2
	Occasional	1
	Never	0
Dysphonia	Yes	1
	No	0
Asthma	Yes	1
	No	0

Maximum score = 13 points.
Minimum score = 0 point.
Case average score = 6.6 points (6.1, 6.9).
Control average score = 0.5 points (0.2, 0.8).

3.2. Validity

Criteria validity was determined applying a linear regression model between the scale and the final DeMeester score, achieving high significance ($P < .001$). Concurrent validity of the scale was then determined against the other variables of 24-h esophageal monitoring, by means of the Spearman correlation. Correlation's >0.50 was observed (Table 5).

Table 5
Concurrent validity: correlation between the scale and 24-h esophageal pH monitoring variables ($n=240$)

Variables	Scale
No. total reflux episodes	0.5037
No. of greater than 5 min episodes	0.5473
Longest episode (min)	0.5107
Total reflux time (min)	0.5470
Supine reflux time (min)	0.5052
Total reflux time percentage (%)	0.5386
Supine reflux time percentage (%)	0.3393
Reflux index (reflux/hour)	0.6137
Final score (points)	0.5355

Table 6
Scale discrimination cutoff points (logistic regression)

Parameters	Cutoff points 1	2	3	4	5
Sensitivity (%)	100.0	95.5	91.6	84.4	73.7
Specificity (%)	72.9	86.4	94.9	96.6	98.3
Positive predictive value (%)	91.8	95.5	98.2	98.7	99.3
Negative predictive value (%)	[a]	86.4	78.9	67.1	55.2
Correct classification (%)	93.3	93.3	92.4	87.4	79.8
Under curve area (%)	50.0	90.9	93.3	90.5	86.0
Study subjects/Scale association (OR)	[b]	136.3	204.1	153.7	162.9

[a] = Not calculable.
OR = Odds ratio.
[b] = Predicts failure perfectly.

Later, construct validity in extreme groups was analyzed. First, the relationship between the scale score and the study groups was studied with a difference of 0.5 points for controls vs. 6.6 points for patients ($P < .001$). A logistic regression model was applied to study the association between the scale and the previously generated extreme groups (0 and 10.1 points, respectively), obtaining a statistically significant association ($P < .001$).

The best cutoff point of the scale for GERD was determined. It was observed that at cutoff points 2 and 3 of the scale adequate values of sensitivity, specificity, and predictive values were obtained (Table 6). ROC curves were built to calculate the area under the curve at different cutoff points (Fig. 1). Finally, a logistic regression model was applied with the scale and the cutoff points. The best option to differentiate healthy subject from patients was to use the cutoff point 3, with an odds ratio of 204 (Table 6).

4. Discussion

GERD is a complex clinical entity, because its manifestations may be multiple and on occasions difficult to interpret. To outline the diagnosis, the presence of symptoms is not enough, requiring diagnostic tests, such as 24-h esophageal pH monitoring, considered as the gold standard. However, the prevalence of GERD is important because this information may be useful to acquire knowledge of epidemiologic and economic aspects of this disease.

As mentioned previously, scanty knowledge is available in relation to the different epidemiologic aspects of GERD. The literature is scarce, and most reports present evident methodologic defects. One of the main criticisms of the reports available is the lack of valid and reliable measurement tools [1,8–13,16]. Thus, not only is the reproducibility capacity of these measurements ignored, but rather, it cannot be determined whether these measurements correspond to the true state of the phenomenon being measured. It is therefore important to design a valid, reliable scale that allows further studies of GERD, either observational or experimental. An exception is the questionnaire designed by Shaw et al. [17], which had been validated on a clinical basis and was designed as a clinical tool in a secondary health care. This questionnaire, however, may have a serious selection bias because despite the limitations of 24-h esophageal pH metry, it continues to be the gold standard for GERD.

In the design of the scale we implemented many items such as symptoms and well known GERD manifestations. However, there are several GERD symptoms that are difficult to interpret such as chronic sinusitis, pharyngolaryngitis, bronchial spasms, recurrent pneumonitis, sleep apnea, recurrent otitis, hoarseness, and dental erosions that were not considered in the scale design.

In the first statistical step, factor analysis was carried out, and it was observed that the items were strongly correlated with the clinical manifestations, because of the single, closely related domain between them.

The DeMeester score variable was used as a criteria validity study because this parameter constitutes a mathemati-

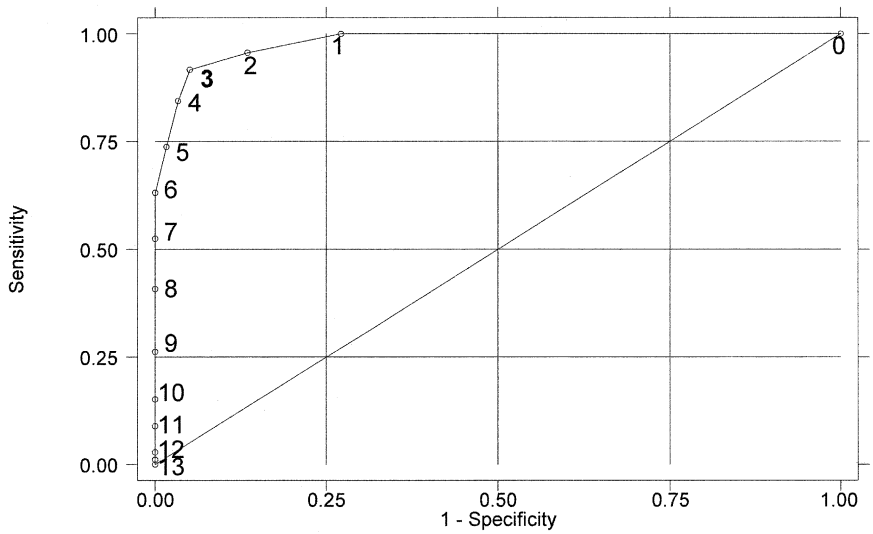

Fig. 1. Definitive scale ROC curve. Area under ROC curve = 0.9806.

cal index that includes other 24-h esophageal pH monitoring variables, making it the best variable to represent the other variables, as later demonstrated in the current validity study.

On completion of reliability and validity studies, the best cutoff point was determined to define the scale discrimination of illness. Both cutoff points 2 and 3 of the scale showed good sensitivity, specificity, and predictive values. However, Table 6 suggests that cutoff point 3 gives the best discrimination parameters with sensitivity, specificity, and positive predictive values greater than 91%. Nonetheless, for prevalence studies it is important for the measurement tool used to have greatest sensitivity possible, while also taking specificity into account, because this may lead to an increase in the number of false negative subjects. Nevertheless, the selection of a cutoff point depends on the desired use of the measurement tool and the decision to be made thereafter as a result of its use.

The next step of our investigation line is to prospectively test the questionnaire. In fact, we have recently begun an epidemiologic survey of GERD in a general population in Temuco (Chile) and Barcelona (Spain).

In conclusion, this article provided a valid, reliable scale to detect GERD symptoms, with a cutoff point providing correct classification of 92% of the subjects, sensitivity greater than 91%, and a specificity of almost 95%. The generated instrument is very simple to administer, is useful as a diagnostic test but not for health-related quality of life, and its utility for assessing long-term history of GERD or even treatment could be proven in future studies.

Acknowledgments

This work was supported by grants DID-UFRO 99-20 and EP/2102, Universidad de La Frontera, Temuco, Chile.

References

[1] The Gallup Organization. Gallup Survey on heartburn across America. Princeton, NJ: The Gallup Organization; 1988; p. 24.
[2] Nebel OT, Fornes MF, Castell DO. Symptomatic gastroesophageal reflux: incidence and precipitating factors. Am J Dig Dis 1976;21: 953–6.
[3] Trimble KC, Douglas S, Pryde A. Clinical characteristics and natural history of symptomatic but not excess gastroesophageal reflux. Dig Dis Sci 1995;40:1098–104.
[4] Cameron AJ. Epidemiology of columnar-lined esophagus and adenocarcinoma. Gastroenterol Clin North Am 1997;26:487–94.
[5] Freys SM, Fuchs KH, Heimbucher J, Fein M, Thiede A. Epidemiology and pathophysiology of Barrett esophagus. Zentralbl Chir 2000; 125:406–13.
[6] McDougall NI, Johnston BT, Kee F, Collins JS, McFarland RJ, Love AH. Natural history of reflux oesophagitis: a 10 year follow up of its effect on patient symptomatology and quality of life. Gut 1996;38: 481–6.
[7] DeMeester TR, Wang CI, Wernly JA, Pellegrini CA, Little AG, Klementschitsch P, Bermudez G, Johnson LF, Skinner DB. Technique, indications, and clinical use of 24 hour esophageal pH monitoring. J Thorac Cardiovasc Surg 1980;79:656–70.
[8] Csendes A, Valenzuela J, Becker P, Arraztoa JA, Medina E. Prevalence of esophageal and gastrointestinal symptoms in Chilean adults. Rev Med Chil 1989;117:146–9.
[9] Mold JW, Reed LE, Davis AB, Allen ML, Decktor DL, Robinson M. Prevalence of gastroesophageal reflux in elderly patients in a primary care setting. Am J Gastroenterol 1991;86:965–70.
[10] Ruth M, Mansson Y, Sandberg N. The prevalence of symptoms suggestive of esophageal disorders. Scand J Gastroenterol 1991;26:73–81.
[11] Raiha IJ, Impivaara O, Seppala M, Sourander LB. Prevalence and characteristics of symptomatic gastroesophageal reflux disease in the elderly. J Am Geriatr Soc 1992;40:1209–11.
[12] Kay L, Jorgensen T, Jensen KH. Epidemiology of abdominal symptoms in a random population: prevalence, incidence and natural history. Eur J Epidemiol 1994;10:559–66.
[13] Bohmer CJ, Klinkenberg-Knol EC, Niezen-de Boer RC, Meuwissen SG. The prevalence of gastro-esophageal reflux disease based on non-specific symptoms in institutionalised, intellectually disabled individuals. Eur J Gastroenterol Hepatol 1997;9:187–90.
[14] 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.
[15] Frank L, Kleinman L, Ganoczy D, McQuaid K, Sloan S, Eggleston A, Tougas G, Farup C. Upper gastrointestinal symptoms in North America: prevalence and relationship to healthcare utilization and quality of life. Dig Dis Sci 2000;45:809–18.
[16] Bohmer CJ, Klinkenberg-Knol EC, Niezen-de Boer MC, Meuwissen SG. Gastroesophageal reflux disease in mentally retarded persons: prevalence, diagnosis and treatment. Ned Tijdschr Geneeskd 2000; 144:1156–61.
[17] Shaw MJ, Talley NJ, Beebe TJ, Rockwood T, Carlsson R, Adlis S, Fendrick AM, Jones R, Dent J, Bytzer P. Initial validation of a diagnostic questionnaire for gastroesophageal reflux disease. Am J Gastroenterol 2001;96:52–7.
[18] Talley NJ, Fullerton S, Junghard O, Wiklund I. Quality of life in patients with endoscopy-negative heartburn reliability and sensitivity of disease-specific instruments. Am J Gastroenterol 2001;96:1998–2004.
[19] Colwell HH, Mathias SD, Pasta DJ, Henning JM, Hunt RH. Development of a health-related quality-of-life questionnaire for individuals with gastroesophageal reflux disease: a validation study. Dig Dis Sci 1999;44:1376–83.
[20] 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. p. 104–62.
[21] Manterola C, Pérez O. Prevalence of gastroesophageal reflux disease symptoms in a student population. Rev Chil Cir 1998;50:513–7.
[22] Manterola C, Muñoz S, Grande L, Riedemann P. Construction and validation of a gastroesophageal reflux symptom scale. Preliminary report. Rev Med Chil 1999;127:1213–22.
[23] Streiner DL, Norman GR. Selecting the items. In: Streiner DL, editor. Health measurement scales. A practical guide to their development and use. New York: Oxford University Press; 1995. p. 54–68.
[24] De Vellis RF. Reliability. In: De Vellis RF, editor. Scale development. Theory and applications. Newbury Park, CA: Sage Publications; 1991. p. 24–42.
[25] Streiner DL, Norman GR. Devising the items. In: Streiner DL, editor. Health measurement scales. A practical guide to their development and use. New York: Oxford University Press; 1995. p. 15–27.
[26] Muñoz SR, Bangdiwala SI. Interpretation of kappa and B statistics measures of agreement. J Appl Stat 1997;24:105–11.
[27] Fletcher RH, Fletcher SW, Wagner EH. Clinical epidemiology. The essentials. 3rd ed. Baltimore: Williams & Wilkins; 1996. p. 22–3.
[28] Hennekens CH, Buring JE. Epidemiology in medicine. 2nd ed. Boston, MA: Little Brown and Company; 1987.