

Lack of Correlation Between a Self-Administered Subjective GERD Questionnaire and Pathologic GERD Diagnosed by 24-h Esophageal pH Monitoring

Kevin Chan · Geoffrey Liu · Linda Miller · Clement Ma · Wei Xu ·
Christopher M. Schlachta · Gail Darling

Received: 2 June 2009 / Accepted: 7 December 2009 / Published online: 12 January 2010
© 2010 The Society for Surgery of the Alimentary Tract

Abstract

Introduction Self-reported reflux symptoms do not always correspond to pathologic gastroesophageal reflux disease (GERD). We evaluated whether GERD-related symptoms in the self-reported Mayo-GERD questionnaire (GERDQ) were correlated with current gold standard definitions of pathologic GERD.

Methods Three hundred thirty-six consecutive consenting individuals with GERD symptoms referred for 24-h esophageal pH monitoring completed a baseline GERDQ. Univariate and multivariate analyses identified questions that were most associated with percent total time pH < 4 at distal probe (DT) > 4% or DeMeester score (DS) ≥ 14.7 , two accepted definitions of pathologic GERD. A risk score was created from these analyses, followed by generation of receiver operating characteristic curves and determination of C-statistics, sensitivity, and specificities at various cut points, with prespecified minimal values of each that would be required to meet the definition of “potential clinical utility.”

Results Forty-nine percent of patients were found to have pathologic GERD; half the patients (not necessarily those with pathologic GERD) described suffering from severe or very severe heartburn or acid regurgitation in the past year. Univariate logistic regression analysis identified six of 22 key GERD questions that were significantly related to DT or DS, in addition to age and gender. Three questions (duration of symptoms, nocturnal heartburn, hiatal hernia) along with age and gender remained significant in multivariate analyses. A risk score (RS) was created from these five questions separately for DT and DS. For DT, the C-statistic for RS was 0.75, and at the optimal cut point of ≥ 6 that maximizes sensitivity (SS) and specificity (SP), SS was 68% and SP was 72%. For DS, the C-statistic was 0.73, and at the optimal cut point, SS was 82% and SP 60%. When considering other cut points, the rare extreme case of very low RS (≤ 2) was strongly predictive of lack of pathologic GERD: for DT, SS 100%/SP 18%, negative predictive value (NPV) 100%; and for DS, SS 97%, SP 25%, NPV 88%. However, only 10–15% of patients referred for pH testing had RS scores of ≤ 2 .

Conclusion Self-reported prolonged history of GERD-like symptoms, nocturnal heartburn, history of a hiatus hernia, and male gender were associated with abnormal 24-h esophageal pH monitoring. However, these factors lack clinical utility to predict pathologic GERD in patients referred for pH testing. We found that 51% of patients with severe GERD symptoms do not have true pathological GERD on objective testing. The clinical implications of this study are significant in that treatment with acid-suppressing medication in such patients would be inappropriate.

Financial Support: Kress Family Professorship in Esophageal Cancer, Alan Brown Chair in Molecular Genomics, Postluns Family, Toronto General Hospital, and Princess Margaret Hospital Foundations, CIHR operating grant.

K. Chan · G. Liu
Department of Medical Biophysics and Medicine, Princess
Margaret Hospital/Ontario Cancer Institute, Toronto, ON, Canada

L. Miller · G. Darling (✉)
Department of Surgery (Thoracic Surgery),
Toronto General Hospital, University of Toronto, Suite 9N-955,
200 Elizabeth Street, M5G 2C4 Toronto, ON, Canada
e-mail: Gail.Darling@uhn.on.ca

C. Ma · W. Xu
Biostatistics Department,
Princess Margaret Hospital/Ontario Cancer Institute,
Toronto, ON, Canada

C. M. Schlachta
Department of Surgery, University of Western Ontario,
London, ON, Canada

Keywords Gastroesophageal reflux disease · Esophageal 24-h pH test · Distal time percent · DeMeester score · Risk score

Introduction

Gastroesophageal reflux disease (GERD) is a common medical condition described as a chronic manifestation of acid exposure to the esophagus which causes a myriad of symptoms sufficient to impair quality of life.^{1,2} In a systematic review of 77,671 patients, 25% of adults were reported to have an episode of heartburn at least once a month, and 12% had symptoms weekly, while 5% suffered from daily heartburn.³ Typical symptoms associated with GERD are heartburn and acid regurgitation, which are highly specific but not sensitive for the diagnosis of GERD.⁴ Temporally, acid reflux episodes variably corresponded to GERD symptoms in patients with and without pathologic GERD.^{5,6} Thus, the correlation between symptoms and pathologic GERD has been variable, at best.

Symptom assessment through standardized questionnaires such as the Mayo-GERD questionnaire (GERDQ) allows patients to self-report GERD symptoms and enables clinicians to assess the impact of GERD-related symptoms on patients.⁷ Assessing whether GERDQ can predict pathologic GERD is therefore an appealing extension of the use of this questionnaire. Predictive tools need to be concise, yet the design of the GERDQ is lengthy, consisting of 80 questions. To achieve clinical utility of such a predictive tool would require identifying a smaller subset of questions within the longer questionnaire.

Ambulatory 24-h esophageal pH monitoring has been long recognized as a standard for objectively measuring pathological GERD, with high sensitivity, specificity, and accuracy ranging between 84% and 100%.^{8–10} The uniform pH scoring system identifies six important parameters as predictors of GERD symptoms.¹⁰ In particular, percent total time pH<4 (distal time (DT)) and DeMeester score (DS) are widely accepted as two key quantitative parameters of GERD. DeMeester score is a sum of component scores of the six individual parameters (of which DT is one parameter). If DT>4% or if DS≥14.7, the test is considered diagnostic of pathologic GERD.^{10–12}

We hypothesized that a specific subset of questions which addressed the typical symptoms of GERD within the GERDQ are highly correlated with the presence of pathologic GERD, as defined by DT and/or DS parameters. The primary goal of this study was to determine whether and which components of GERDQ accurately identified patients with pathologic GERD as defined by 24-h esophageal pH testing.

Materials and Methods

Study Design

The study was approved by the University Health Network Research Ethics Board. The study design was a prospective cross-sectional evaluation of consecutively consenting patients (February 2003 to February 2008) with clinical symptoms compatible with gastroesophageal reflux disease (GERD) who were referred to the Esophageal Function Laboratory at Toronto General Hospital for 24-h esophageal pH monitoring. All were naïve to pH testing. Patients eligible for inclusion were those patients with symptoms of GERD who were referred for 24-h pH testing who could read and understand English. Patients were excluded from the study if they had previous antireflux surgery, had a known esophageal motility disorder, or were under evaluation pre- or post-lung transplantation.

After consenting to participate in the study, patients were given the GERDQ questionnaire to complete over the course of the pH testing. Motility and ambulatory 24-h pH testing were performed as outpatients. Patients were required to fast for at least 4 h prior to the testing. All antireflux medications were discontinued 1 week prior to the pH testing. No restriction was placed on patients' daily activities, eating, drinking, or smoking habits. Data from GERDQ questionnaires were kept segregated from pH testing data until the time of analysis.

Mayo-GERD Questionnaire

Mayo-GERDQ is a validated self-administered questionnaire designed to measure symptoms of GERD consisting of 80 questions concerning patients' experience of acid reflux episodes.⁷ The questionnaire addresses four major primary symptoms of GERD including heartburn, acid regurgitation, chest pain, and dysphagia, of which heartburn and acid reflux had the highest specificity for GERD.⁴ In addition, GERDQ also asked questions about atypical symptoms, lifestyle, general quality of life, general medical history, and review of other symptoms. Twenty-two GERDQ questions related to cardinal symptoms of heartburn and acid regurgitation were selected for in-depth analyses. These 22 questions either utilized a Likert scale or were Yes/No dichotomized questions and were classified into eight major categories: duration since the first onset of symptoms, frequency of symptoms, severity of symptoms, nocturnal symptoms, antacid medication, duration of antacid administration, history of gastric or esophageal diseases, smoking and drinking habits.

Manometry and Esophageal pH Monitoring

Esophageal manometry was first performed on all patients to identify the level of the lower esophageal sphincter. An

eight-channel motility catheter was used to assess lower esophageal sphincter (LES) pressure when the patient was supine. Manometric LES values were identified through ten consecutive swallows with 5 cm³ aliquots of water. The amplitude and activity of peristaltic contraction, the upper esophageal sphincter (UES) location, resting and contraction tone, and coordination were also measured.

After removal of the motility assembly, esophageal pH was measured using a COMFORTEC (Sandhill Scientific), two-channel pH probe with 15 cm spacing. The distal sensor was positioned at 5 cm above the manometrically defined LES, while the proximal one was located at 15 cm above the LES. Esophageal pH was monitored and recorded electronically for a 24-h period. Patients were asked to maintain daily normal activities and diet. Using the GERD pH monitoring device (Sandhill Scientific) and Bioview pH software, the number and duration of reflux episodes were measured. Upright and supine acid exposure times (in percentage) were also calculated.

Statistical Analysis

Summary statistics of demographic variables, GERDQ responses, and pH testing results were generated. Because both DT and DS have been used as the reference for defining pathologic GERD, we included both in our primary analyses. Patients were classified as having either pathologic GERD or not on the basis of the standardized cutoffs for DT (no GERD $\leq 4\%$; Pathologic GERD $> 4\%$) or DS (no GERD, < 14.7 ; pathologic GERD, ≥ 14.7). Univariate logistic regression analysis was used to test the association of each of the selected GERDQ questions to pathologic GERD status. Statistically significant predictors of pathologic GERD from univariate analyses were entered into a multivariate logistic regression model using stepwise selection with *p* value cutoffs of 0.20 and 0.15 to enter and remain in the model, respectively.

Using the estimated coefficients (β) to estimate relative weights of each predictor in the multivariate model, a risk score was created of the GERDQ questions in the multivariate model. Receiver operating characteristic (ROC) curves were generated for the risk score (Fig. 1). Sensitivity (SS) as well as specificity (SP) at different risk score cut points were considered, with the final cut points chosen to maximize potential clinical utility.

“Potential clinical utility” for the risk score was considered likely if any of the following conditions were met: (1) *C*-statistic from ROC ≥ 0.85 or (2) having an optimal risk score cut point (which is the cut point which maximizes SS and SP together) such that both SP and SS are over 80%. We planned to pursue validation of the risk scores model using an independent set of samples only if “potential clinical utility” was met.

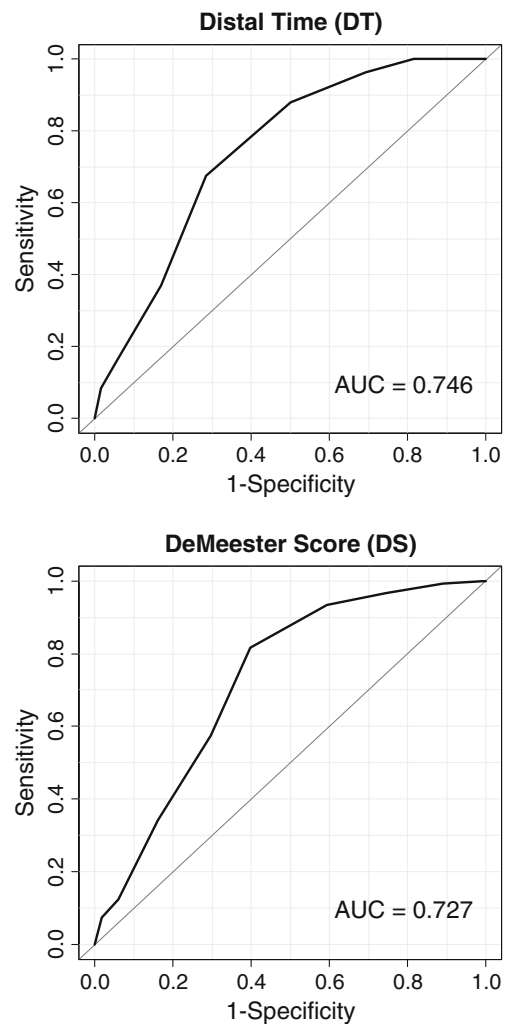


Figure 1 Receiver operator curves (ROC) for risk score, showing results for percent total time pH < 4 at DT and DS separately.

Results

Demographics

Between February 2003 and February 2008, of consecutive patients referred for esophageal motility and pH monitoring, 374 met the inclusion criteria. Of these, 336 agreed to participate (90% annual recruitment rate) and completed both GERDQ and pH testing. Of patients completing GERDQ, 203 (60.4%) were females. The median (range) age was 49.8 (18–85) years (Table 1). Pathologic GERD was diagnosed in 49.4% based on the objective measurement of DT or DS, or both, on esophageal 24-h pH testing.

Symptoms

In this patient population, 48% had episodes of heartburn, and 41% had episodes of acid regurgitation for more than 5 years. Over half of all patients (51%) reported “severe or

Table 1 Demographic and Esophageal pH Study Data of Patients with GERD Referred for 24-h pH Monitoring

Characteristic	Total	Pathologic GERD ^a	No GERD ^a
Total number	336 ^b	166	170
Age in years			
Median (IQR)	49.8 (18.9)	52.5 (18.5)	45.9 (20.1)
Gender			
Female	203 (60%)	89 (44%)	114 (56%)
Male	133 (40%)	77 (58%)	56 (42%)
DT			
Median (IQR)	3.60 (1.92)	7.5 (7.7)	1.0 (1.6)
DeMeester score (DS)			
Median (IQR)	13.1 (1.95)	32.7 (31.0)	4.5 (5.8)
Normal proximal upright time (%)	300/310 (96.8)	147/156 (94.2)	153/154 (99.4)
Normal proximal supine time (%)	261/309 (84.5)	114/156 (73.1)	147/153 (96.1)
Lower esophageal sphincter (LES) resting tone (mmHg)			
Median (interquartile range)	11 (6.0)		
Patients with normal result ≥ 16 to ≤ 30 (%)	51/306 ^c (16.5)	12/152 (7.9)	39/154 (25.3)
Patients with abnormal result < 16 (%)	255/306 ^c (83.5)	140/152 (92.1)	115/154 (74.7)

IQR interquartile range

^a Using combined definition of either DT > 4% or DS ≥ 14.7

^b Denominator = 336 unless otherwise specified

^c Three patients had LES resting tone > 30 mmHg

very severe" heartburn, but only 50% and 54% of these patients had DT > 4% or DS ≥ 14.7 , respectively. Similarly, 48% complained of severe or very severe acid regurgitation, but only 52% and 56% of these patients had DT > 4% or DS ≥ 14.7 , respectively. Over 80% of patients reported being awakened at night from heartburn, and 75% reported nocturnal regurgitation. Daily heartburn occurred in 39% of patients within the previous year, while 47% described heartburn occurring at least once a week in the past year. Symptom improvement with antacids was reported by 58%. Heartburn or acid regurgitation affected daily activities some of the time in 35% of patients and in most or all of the time in 21%. The 227 patients who had at least one episode of heartburn or acid reflux on 24-h pH testing had a median of seven (interquartile range 14) episodes of documented acid reflux based on DT or DS, but only 50% of GERD symptom episodes were correlated with DT > 4% or DS ≥ 14.7 % (interquartile range 84%).

One in four patients had presented to their doctor's offices six times or more in the previous year for GERD symptoms. Ninety percent had already received some sort of diagnostic test by their physician prior to referral for pH testing. Ever-smokers accounted for only 45% of the patients, while fewer than 38% of this patient sample drank alcohol in the last year; however, almost 70% were coffee drinkers.

DT, DS, and Pathologic GERD

The median and interquartile range for DT and DS are presented, for the overall sample and separately by pathologic GERD status, in Table 1. As expected, DT and DS values were strongly correlated (Pearson correlation coefficient = 0.98; $p < 0.0001$). Because of this high correla-

tion, we used either high DT or high DS to define pathologic GERD in our initial demographic comparisons. When comparing pathologic GERD to no-GERD individuals, males had a significantly higher prevalence of pathologic GERD ($p = 0.01$, chi-square test), as did older individuals ($p = 0.008$, t test), when using this combined DT/DS definition of pathologic GERD (Table 1).

Univariate and Multivariate Models

Although we utilized a combined definition of DT/DS for our demographic variables to allow convenient reporting in Table 1, we performed our primary analysis separately for DT and DS, given that each is considered a standard in its own right. Univariate logistic regression analysis identified six out of 22 GERD-related questions with the greatest statistical significance for either DT or DS. These questions, identified as (Q1) through (Q6), were: (Q1) When did heartburn first begin? (Q2) Has heartburn awakened you at night? (Q3) Have you had acid regurgitation in the past year? (Q4) Has acid regurgitation awakened you at night? (Q5) Have you ever had hiatus hernia? and (Q6) Have you ever had disease of esophagus or stomach? In addition to these GERDQ questions, being male (Q7: What gender are you?) and older age (Q8: How old are you at the time of your pH testing?) was also associated with abnormal DT or DS (Table 2).

In multivariate logistic regression analysis, (Q1) through (Q8) were assessed using stepwise selection. Five predictors remained statistically significant or near significant after stepwise selection, (Q1), (Q2), (Q5), (Q7), and (Q8), and these data are presented in Table 3. For (Q1), the original Likert categories were partially collapsed based on

Table 2 Summary of Univariate Analysis—Questions in GERDQ Most Associated with Pathologic GERD (as Defined by D or DS)

Question	Category	% Patients	DT		DS	
			OR (95% CI)	<i>p</i> value	OR (95% CI)	<i>p</i> value
1. When did heartburn first begin?	≤2 year ago	27	Reference	0.002	Reference	0.003
	>2 years and ≤5 years ago	25	1.96 (1.0–4.0)		2.06 (1.0–4.1)	
	More than 5 years ago	48	3.00 (1.6–5.6)		2.91 (1.6–5.4)	
2. Has heartburn awakened you at night?	No	17	Reference	0.004	Reference	0.02
	Yes	83	2.96 (1.4–6.2)		2.21 (1.1–4.4)	
3. Have you had acid regurgitation last year?	No	11	Reference	0.01	Reference	0.003
	Yes	89	2.81 (1.3–6.2)		3.35 (1.5–7.4)	
4. Has acid regurgitation awakened you at night?	No	25	Reference	0.01	Reference	0.02
	Yes	75	2.11 (1.2–3.7)		1.96 (1.1–3.5)	
5. Have you ever had a hiatus hernia?	No	42	Reference	0.0005	Reference	<0.0001
	Yes	58	2.36 (1.5–3.8)		3.15 (1.9–5.1)	
6. Have you ever had disease of esophagus or stomach?	No	78	Reference	0.05	Reference	0.02
	Yes	22	1.75 (1.0–3.1)		1.99 (1.1–3.5)	
7. Gender	Female	60	Reference	0.003	Reference	0.01
	Male	40	1.95 (1.3–3.0)		1.78 (1.1–2.8)	
8. Age	≤42 years old	35	Reference	0.005	Reference	0.002
	>42 and ≤55 years old	31	1.56 (0.9–2.7)		1.56 (0.9–4.6)	
	>55 years old	35	2.49 (1.4–4.3)		2.67 (1.5–4.6)	

the results of the univariate analysis and the frequency of each category (adjacent categories with few individuals were automatically collapsed together). For (Q8), the log-odds of the risk function for age approximated linearity; thus, dividing the sample into tertiles was chosen for convenience in developing the risk score model.

At least one question was not answered by 98 patients (29%), and therefore these patients could not be included in the multivariate analysis. However, using chi-square analysis, there was no difference between the group with missing questions and the group with complete answers in terms of the frequency of pathological GERD either by DT,

Table 3 Multivariate Analysis: Final Models

Questions	Category	Referent	<i>p</i> value	Odds ratio (95% CI)
DT Model				
(Q1) When did heartburn first begin?*	>2 years ago and ≤5 years ago	In the last 2 years	0.16	1.81 (0.8–4.1)
	>5 years ago		0.01	2.55 (1.2–5.3)
(Q2) Has heartburn awakened you at night?	Yes	No	0.01	2.92 (1.3–6.7)
(Q5) Have you had hiatus hernia?	Yes	No	0.02	2.05 (1.1–3.8)
(Q7). What gender are you?	Male	Female	0.0003	3.12 (1.7–5.8)
(Q8) How old are you?	>42 and ≤55 years	≤42 years	0.14	1.74 (0.8–3.6)
	>55 years		0.001	3.41 (1.6–7.2)
DS Model				
(Q1) When did heartburn first begin?*	>2 years ago and ≤5 years ago	In the last 2 years	0.11	1.90 (0.9–4.2)
	>5 years ago		0.02	2.36 (1.2–4.7)
(Q2) Has heartburn awakened you at night?	Yes	No	0.10	1.91 (0.9–4.2)
(Q5) Have you had hiatus hernia?	Yes	No	0.002	2.50 (1.4–4.5)
(Q7) What gender are you?	Male	Female	0.01	2.17 (1.2–4.0)
(Q8) How old are you?	>42 and <55 years	≤42 years	0.18	1.62 (0.8–3.3)
	>55 years		0.003	2.97 (1.4–6.1)

*DT model global *p* value for Q1 *p*=0.04; global *p* value for Q8 *p*=0.005

**DS model global *p* value for Q1 *p*=0.06; global *p* value for Q8 *p*=0.01

DS, or both. Also, answers to the eight questions were similar between the two groups at least for the questions that were answered (data not shown).

Risk Score Creation and Assessment of Risk Score Characteristics

Risk scores were weighted, with the weighting based on estimated β values of each question in the multivariate models (Table 4). The risk score developed for DS and DT using these criteria had a range from 0 through 9, but the weighting was slightly different for DS and DT (Table 4).

ROC curves were generated and yielded *C*-statistics for DT and DS of 0.75 and 0.73, respectively. Thus, this risk score fails criterion (1) for “potential clinical utility.”

The optimal risk score cut point for DT was ≥ 5 , and the optimal cut point for DS was ≥ 6 . At these cut points, SS were 68% and 82% and SP were 72% and 60% for DT and DS, respectively. Thus, this risk score does not meet criterion (2) for “potential clinical utility” either (see Table 5).

Finally, when considering other risk score cut points, values for sensitivity and specificity for all possible cut points were well below 90%, except for the extreme risk score cut point of ≤ 2 , which had SS values of 97–100% (range is reported as DS and DT analyses were performed separately) and negative predictive values of 88–100% in this population. As expected, SP values were very low for this cut point (below 50%). Furthermore, the percentage of individuals (in this population of referred patients for pH testing) that had risk score values ≤ 2 was only 10–15% (see Table 6).

Discussion

Despite a number of GERD questionnaires that were designed to provide reliable assessments of GERD symptoms, comparisons between questionnaires have been difficult, and interpretation of results have varied greatly.^{13–18} Intuitively, clinicians have often assumed that GERD symptoms reported in a questionnaire would accurately represent pathologic GERD. If there is little or no association of GERD symptoms and objective esophageal 24-h pH measurements, misdiagnosis and inappropriate treatment for GERD may result. To our knowledge, our study is the first attempt to develop correlation between subjective self-reported GERD symptoms using GERDQ and objective quantification of key parameters used in 24-h pH testing.

Previous studies have reported that even if reflux symptoms are eliminated completely, it might not ensure normalization of esophageal pH reading.¹⁹ Nor does persistence of GERD symptoms imply pathological GERD. In our study, we identified five key questions that were highly associated with the presence of pathologic GERD in this highly selected patient population who were referred for investigation of reflux symptoms. These questions asked about: (1) the duration since the start of heartburn (at least 2–5 years ago but especially if more than 5 years ago); (2) nocturnal symptoms of heartburn; (3) previous diagnosis of hiatus hernia; (4) being male; and (5) being older, all of which were shown to be statistically significant in the multivariate model. Yet with the rare exception of extremely low risk score values, the resultant risk score derived from these questions was not at all useful in discriminating pathologic GERD from

Table 4 Risk Score Index for the Significant GERDQ Questions from Multivariate Model

Variable	Category	DT		DS	
		Estimated β value from multivariate model	Risk Score Values*	Estimated β value from multivariate model	Risk score values ^a
Q1 When did heartburn first begin?	≤ 2 years ago	0	0	0	0
	>2 years but ≤ 10 years ago	0.59	1	0.64	1
	>10 years ago	0.94	2	0.86	2
Q2 Has heartburn awakened you at night?	No	0	0	0	0
	Yes	1.07	2	0.65	1
Q5 Have you ever had hiatus hernia?	No	0	0	0	0
	Yes	0.71	1	0.91	2
Q7 What gender are you?	Female	0	0	0	0
	Male	1.14	2	0.77	2
Q8 How old are you?	≤ 42 years	0	0	0	0
	>42 and ≤ 55 years	0.55	1	0.48	1
	>55 years	1.23	2	1.09	2

^a β values from 0–0.25 were scored 0; from 0.25–0.75 were scored 1; and from 0.75–1.25 were scored 2. The interquartile cut points (0.25 and 0.75) were chosen for rounding purpose to the nearest 0.50

Table 5 Number of Patients with Risk Score Results by Pathologic GERD, as Defined by DT and DS, Separately Presented

	Risk Score										Total
	0	1	2	3	4	5	6	7	8	9	
DT											
DT≤4%	1	4	19	16	25	28	15	15	5	2	130
DT>4%	0	0	0	4	9	22	33	23	8	9	108
Total	1	4	19	20	34	50	48	38	13	11	238 ^a
DS											
DS<14.7	1	12	17	18	23	12	16	12	5	2	118
DS≥14.7	0	1	3	4	14	29	28	26	6	9	120
Total	1	13	20	22	37	41	44	38	11	11	238 ^a

^a Values do not add up to 336 because of missing values in specific questions. A missing value in any of the five questions renders that sample unanalyzable for risk score.

those with false-positive symptoms. Furthermore, despite self-reported severe symptoms, only approximately half the patients actually had pathologic GERD based on objective testing.

Similar to our study, Klauser et al.⁴ found that GERD-related symptoms, particularly heartburn and acid regurgitation, were highly associated with pathologic GERD but were not particularly discriminatory for predicting pathologic GERD. Similarly, Schlesinger et al.²⁰ reported that 24-h pH monitoring was normal in half of the individuals with reflux symptoms and in 29% with erosive esophagitis. By all of our prespecified criteria (see “Statistical Analysis” section), our risk scores fell short of potential clinical utility for predicting pathologic GERD.

Various subjective diagnostic tools for GERD have been compared to objective 24-h pH monitoring. Klauser et al.⁴ compared personal interview by gastroenterologists to pH testing, where there was some correlation between an experienced gastroenterologist’s subjective assessment and pathologic GERD. Ghoshal et al.²¹ compared another standardized questionnaire (Carlsson–Dent), esophageal biopsy findings, and treatment responses with omeprazole with pH testing and found some correlation between severity of symptoms and severity of pH findings. Our

study focused on comparing a self-reported questionnaire of GERD symptoms, GERDQ, with pH testing results and found a general lack of clinical utility from GERDQ to predict pathologic GERD.

Nocturnal reflux symptoms are often considered one of the key symptoms of GERD, and this association was confirmed in our study. Weigt et al.²² found that individuals with more typical symptoms of heartburn and regurgitation were associated with greater nocturnal esophageal acid breakthrough on pH testing in patients who were already on proton pump inhibitors. Another study reported low specificity (65%) of nocturnal heartburn and greater specificity using nocturnal acid regurgitation (88%) and cough at night (100%), and the study also reported low sensitivity with each of these symptoms.²³ It is likely that both nocturnal acid regurgitation and heartburn are associated with pathologic GERD, and both questions were strongly associated with pathologic GERD in our univariate analysis. However, the tight correlation between these two variables likely led to one of them dropping out of the multivariate model.

Our study found that patients with hiatus hernia were strongly associated with both abnormal DT and DS. This result corresponds to two studies. DeMeester et al.²⁴ reported that acid reflux episodes were found in greater proportion of patients who had a diagnosis of hiatus hernia compared with those without such a diagnosis (83% and 43%, respectively). Jenkinson et al.²³ reported that hiatus hernia alone could detect abnormal nocturnal acid reflux with 79% sensitivity and 76% specificity; furthermore, when hiatus hernia and nocturnal reflux symptoms (heartburn, acid regurgitation) were present together, specificity increased to 100%. Together, these data are all consistent with our present results.

We confirmed that men are significantly more likely to have pathologic GERD than women in findings previously reported.^{25–27} Lin et al.²⁸ presented complementary data whereby, in men and women who had similar pH testing results, women reported greater severity of GERD symptoms (heartburn, acid regurgitation, nocturnal symptoms)

Table 6 Risk Score by Pathologic GERD, Using a Cut Point of ≤2, for DT and DS Separately

	Risk score		Total
	≤2	>2	
DT			
DT≤4%	24	106	130
DT>4%	0	108	108
Total	24	214	238
DS			
DS<14.7%	30	88	118
DS≥14.7%	4	116	120
Total	34	204	238

than men. Richter and DeMeester²⁹ theorized that a greater parietal cell mass in men leads to greater acid secretion in men but does not explain the differences in symptom perception. Lin et al.²⁸ suggested that higher symptom perception and lower pain threshold in women might account for some of these differences. In addition to gender, older men were found to experience longer episodes of reflux than either younger individuals of either gender in one study.²⁶ We confirmed the independent association between increasing age and higher rates of pathologic GERD but did not find an age–gender interaction described in this other study.²⁶

Self-reported GERD questionnaires can be useful. Andersen et al.³⁰ found GERD-related questions to have high sensitivity. Symptom indicators successfully identified almost two thirds of patients with symptoms such as nocturnal heartburn, chest pain, and dysphagia. However, this study compared individuals having benign esophageal disease with individuals having angina pectoris, gastric and duodenal ulcers, or “normal” healthy populations which were vastly different from our underlying patient population. In addition, Shimoyama et al.³¹ also evaluated nine questions from a 50-item questionnaire with a high sensitivity of 80% (compared to the original 50-item questionnaire); this study did not employ pH testing. While endoscopy may be useful to exclude non-GERD cases, there was also a wide variation to accurately diagnose pathologic GERD using the surrogate endoscopic marker of “mucosal breaks,” and this variation depended greatly on endoscopists’ experience.³²

There are several limitations of this study. First, patients were all referred by their physicians for the esophageal motility and pH testing either because of poor response to drug therapy, referring physician’s suspicions that the symptoms were *not* related to GERD, or prior to consideration of antireflux surgery. This would lead to potential selection bias towards both extremes: overrepresentation of severe pathologic GERD cases and overrepresentation of atypical GERD-symptom patients without pathologic GERD. However, under these circumstances, one would have expected a higher chance of identifying a clinical subset of questions that could discriminate pathologic GERD from no GERD, which was not what we found. Our results are further confounded by the fact that physician referral is typically based on the physician’s assessment of a patient’s GERD symptoms, and agreement between physicians’ and patients’ perceptions of GERD symptoms is often poor.³³ Secondly, we assessed only one questionnaire, GERDQ. Although this is a validated questionnaire in other settings, it is possible that other questionnaires could be more discriminatory for pathologic GERD in the setting of referral for pH testing. Despite these concerns, we chose GERDQ because it has been validated

and assesses multiple dimensions of the most specific symptoms of GERD, heartburn and acid regurgitation. Thirdly, we assessed a very specific patient subgroup referred for pH testing as a result of our initial hypothesis. As shown in our results, our patients had a high prevalence of symptoms with specificity for pathologic GERD, including nocturnal symptoms, severe and frequent acid regurgitation, and/or heartburn symptoms often of prolonged duration. Thus, the usefulness of GERDQ in other settings, such as use as a general population screening tool or to correlate with impact on activities of daily living, cannot be generalized from this study. Finally, approximately one third of patients had at least one missing information question, which was probed to determine if results based on missing data were statistically different from the complete data. The potential discrepancy of the data might affect its validity for multivariate analysis. However, as patients with pathologic GERD evident by either $DT > 4\%$ and/or $DS \geq 14.7$ were compared, none of these outcomes was statistically different from the missing and nonmissing groups. When the eight questions were first studied in the univariate analysis, the difference between the missing and nonmissing groups was found to be statistically significant only in the question, “Have you ever had acid regurgitation last year?” ($p = 0.01$). Yet, this question was later discarded, and the overall difference between missing and nonmissing groups remained negligible for the multivariate analysis. It showed that these predictor questions carried equally valid beta coefficients for risk score development regardless of whether complete or missing information were used.

Conclusion

Using a self-administered standardized and validated questionnaire, GERDQ, our study found that abnormal 24-h esophageal pH monitoring was associated with the following characteristics: prolonged history of GERD-like symptoms, nocturnal heartburn, history of a hiatus hernia, and male gender. Despite statistically significant associations, these questions lacked clinical utility to predict pathologic GERD in patients referred for pH testing. Furthermore, pathological GERD as determined by 24-h pH testing was present only in approximately half of the patients despite severe self-reported symptoms.

The clinical implications of this study are significant in as much as patients with GERD symptoms are frequently treated with proton pump inhibitors or other acid-suppressing medications without objective evidence of pathological GERD. Our study demonstrates that 51% of patients with severe GERD symptoms do not have true

pathological GERD on objective testing, and treatment with acid-suppressing medication would be inappropriate. Similarly, patients who have had antireflux surgery who subsequently complain of GERD symptoms should have objective testing before prescribing acid-suppressing medications since symptoms do not correlate with actual acid reflux.

Acknowledgments We acknowledge the help and support of Jennifer Teichman, Jonathan Lui, Connor Small, and Dr Zhuo Chen.

References

- 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(4):481–486.
- Quigley EMM, Hungin APS. Quality-of-life issues in gastro-oesophageal reflux disease. *Aliment Pharmacol Ther* 2005;22 (Suppl 1):41–47. Review
- Moayyedi P, Talley NJ. Gastro-oesophageal reflux disease. *Lancet* 2006;367:2086–2100
- Klauser AG, Schindlbeck NE, Müller-Lissner SA. Symptoms in gastro-oesophageal reflux disease. *Lancet* 1990;335(8683):205–208.
- Colas-Atger E, Bonaz B, Papillon E, Gueddah N, Rolachon A, Bost R, Fournet J. Relationship between acid reflux episodes and gastroesophageal reflux symptoms is very inconsistent. *Dig Dis Sci* 2002;47(3):645–651.
- Shi G, Bruley des Varannes S, Scarpignato C, Le Rhun M, Galmiche JP. Reflux related symptoms in patients with normal oesophageal exposure to acid. *Gut* 1995;37(4):457–464.
- Locke GR, Talley NJ, Weaver AL, Zinsmeister AR. A new questionnaire for gastroesophageal reflux disease. *Mayo Clin Proc* 1994;69(6):539–547.
- Fuchs KH, DeMeester TR, Albertucci M. Specificity and sensitivity of objective diagnosis of gastroesophageal reflux disease. *Surgery* 1987;102(4):575–580.
- Jamieson JR, Stein HJ, DeMeester TR, Bonavina L, Schwizer W, Hinder RA, Albertucci M. Ambulatory 24-h esophageal pH monitoring: normal values, optimal thresholds, specificity, sensitivity, and reproducibility. *Am J Gastroenterol* 1992;87(9):1102–1111.
- Streets CG, DeMeester TR. Ambulatory 24-hour esophageal pH monitoring: why, when, and what to do. *J Clin Gastroenterol* 2003;37(1):3–4.
- Schindlbeck NE, Heinrich C, König A, Dendorfer A, Pace F, Müller-Lissner SA. Optimal thresholds, sensitivity, and specificity of long-term pH-metry for the detection of gastroesophageal reflux disease. *Gastroenterol* 1987;93(1):85–90.
- Johnson LF, DeMeester TR. Development of the 24-hour intra-esophageal pH monitoring composite scoring system. *J Clin Gastroenterol* 1986;8(Suppl 1):52–58.
- Talley NJ, Locke GR III, McNally M, Schleck CD, Zinsmeister AR, Melton LJ III. Impact of gastroesophageal reflux on survival in the community. *Am J Gastroenterol* 2008;103(1):12–19.
- Dimenäs E, Glise H, Hallerbäck B, Hernqvist H, Svedlund J, Wiklund I. Quality of life in patients with upper gastrointestinal symptoms. An improved evaluation of treatment regimens? *Scand J Gastroenterol* 1993;28(8):681–687
- Dimenäs E, Carlsson G, Glise H, Israelsson B, Wiklund I. Relevance of norm values as part of the documentation of quality of life instruments for use in upper gastrointestinal disease. *Scand J Gastroenterol Suppl* 1996;221:8–13.
- Shaw MJ, Beebe TJ, Adlis SA, Talley NJ. Reliability and validity of the digestive health status instrument in samples of community, primary care, and gastroenterology patients. *Aliment Pharmacol Ther* 2001;15(7):981–987.
- Shaw M. Diagnostic utility of reflux disease symptoms. *Gut* 2004;53(Suppl IV):iv25–iv27
- Sandha GS, Hunt RH, Veldhuyzen van Zanten SJ. A systematic overview of the use of diary cards, quality-of-life questionnaires, and psychometric tests in treatment trials of *Helicobacter pylori*-positive and -negative non-ulcer dyspepsia. *Scand J Gastroenterol* 1999;34:244–249.
- Milkes D, Gerson LB, Triadafilopoulos G. Complete elimination of reflux symptoms does not guarantee normalization of intraesophageal and intragastric pH in patients with gastro-oesophageal reflux disease (GERD). *Am J Gastroenterol* 2004;99 (6):997–999.
- Schlesinger PK, Donahue PE, Schmid B, Layden TJ. Limitations of 24-hour intraesophageal pH monitoring in the hospital setting. *Gastroenterol* 1985;89(4):797–804.
- Ghoshal UC, Chourasia D, Tripathi S, Misra A, Singh K. Relationship of severity of gastroesophageal reflux disease with gastric acid secretory profile and esophageal acid exposure during nocturnal acid breakthrough: a study using 24-h dual-channel pH-metry. *Scand J Gastroenterol* 2008;43(6):654–661.
- Weigt J, Kandulski A, Büsch F, Malfertheiner P. Nocturnal gastric acid breakthrough is not associated with night-time gastroesophageal reflux in GERD patients. *Dig Dis* 2009;27 (1):68–73.
- Jenkinson LR, Norris TL, Watson A. Symptoms and endoscopic findings – can they predict abnormal nocturnal acid gastro-oesophageal reflux? *Ann R Coll Surg Engl* 1989;71(2):117–119.
- DeMeester TR, Lafontaine E, Joelsson BE, Skinner DB, Ryan JW, O’Sullivan GC, Brunson BS, Johnson LF. Relationship of a hiatal hernia to the function of the body of the esophagus and the gastroesophageal junction. *J Thorac Cardiovasc Surg* 1981;82 (4):547–558
- Ter RB, Johnston BT, Castell DO. Influence of age and gender on gastroesophageal reflux in symptomatic patients. *Dis Esophagus* 1998;11(2):106–108.
- Richter JE, Bradley LA, DeMeester TR, Wu WC. Normal 24-hr ambulatory esophageal pH values. Influence of study center, pH electrode, age, and gender. *Dig Dis Sci* 1992;37(6):849–856.
- Fass R, Sampliner RE, Mackel C, McGee D, Rappaport W. Age- and gender-related differences in 24-hour esophageal pH monitoring of normal subjects. *Dig Dis Sci* 1993;38(10):1926–1928.
- Lin M, Gerson LB, Lascar R, Davila M, Triadafilopoulos G. Features of gastroesophageal reflux disease in women. *Am J Gastroenterol* 2004;99(8):1442–1447.
- Richter JE, DeMeester TR. Gender but not age affects normal 24-hour esophageal pH values. *Am J Gastroenterol* 1990;85: A1224.
- Andersen LI, Madsen PV, Dalgaard P, Jensen G. Validity of clinical symptoms in benign esophageal disease assessed by questionnaire. *Acta Med Scand* 1987;221(2):171–177.
- Shimoyama Y, Kusano M, Sugimoto S, Kawamura O, Maeda M, Minashi K, Kuribayashi S, Higuchi T, Zai H, Ino K, Horikoshi T, Moki F, Sugiyama T, Toki M, Ohwada T, Mori M. Diagnosis of gastroesophageal reflux disease using a new questionnaire. *J Gastroenterol Hepatol* 2005;20(4):643–647.
- Kusano M, Ino K, Yamada T. Interobserver and intraobserver variation in endoscopic assessment of GERD using the Los Angeles classification. *Gastrointest Endosc* 1999;49(6):700–704
- McColl E, Junghard O, Wiklund I, Revicki DA. Assessing symptoms in gastroesophageal reflux disease: how well do clinicians’ assessments agree with those of their patients? *Am J Gastroenterol* 2005;100:11–18

Discussant

Dr Vic Velanovich (Detroit, MI): I would like to congratulate Dr. Chan and his colleagues for a well-presented study and thank them for getting their manuscript to me in advance of this meeting.

Eleven years ago, we actually did a similar study looking at symptom severity and comparing that to several physiologic measurements, including 24-h PH monitoring, esophagitis by manometry, and esophagitis as returned by endoscopy. And only the level of esophagitis was associated with the symptoms; nothing else was. So the results of your study really are not that surprising.

I do want to talk a little about what the purpose of the study was. It seems to me that you were trying to come up with a handful of questions that could be used to identify patients with pathologic reflux. And clearly, the GERDQ did not fit this bill, which is not surprising.

I do have one big-picture question and a couple of little-picture questions. If this endeavor was successful, that is if you could actually find a handful of questions to identify pathologic reflux, what do you perceive the clinical utility of such a questionnaire would be? Would it somehow change treatment or evaluation strategies?

And would it be more effective than a simple trial of proton-pump inhibitor (PPI) as that seems to be the standard of care now?

As far as the little-picture questions, there are many instruments available to measure GERDQ-related symptoms. So why did you choose this one?

And, more importantly, the instruments are validated usually in total, in complete questionnaires. So to cherry-pick a handful of questions does not necessarily confer that validity onto the questions you selected.

So did you compare the total score with the subset of the 22 questions that you actually ended up selecting and using?

Lastly, I noted that less than one-half of the patients in your study had pathologic reflux? Did you assess for non-acid reflux?

Once again, thank you for the opportunity to review the manuscript and discuss this, and congratulations on a well-done good study.

Closing discussant

Dr Kevin Chan: To answer the first question, I think if the study had been successful, identification of key questions that correlated with pathological GERD would have been useful to clinicians in helping to select patients for

treatment whether medical or surgical or to direct further investigations.

The current practice of a trial of PPIs has resulted in many patients being treated with these medications inappropriately.

To answer your second question, we recognize and appreciate the fact that there are many validated questionnaires available. We chose the Mayo-GERDQ because of the breadth of information it captured regarding the patient's overall health. For the purpose of this study, we chose to analyze only the questions that had relevance to GERD. We did not exclude any questions that may have been related to GERD. The excluded questions related to past history, other nongastrointestinal medical conditions, general health, etc.

The last question is a very good one but the answer is no, we did not assess for non-acid reflux.

Discussant

Dr Tom DeMeester (USC): Dr. Chan's study is of particular interest to me because of the current acceptance that GERD is a chronic progressive disease that passes through distinct stages. The earliest of these stages is the non-erosive stage. These patients can have the symptoms of GERD but have normal esophageal acid exposure. I suspect that some of these patients were part of your study group.

The questions are, do such patients have GERD or just an acid-sensitive esophagus, and is an acid-sensitive esophagus an early stage of GERD before esophageal acid exposure becomes abnormal? Did you look specifically at this group of patients, i.e., those who had symptoms and were pH normal? Did you do any sensitivity testing on such patients? Did you investigate them further to see if they had an acid-sensitive esophagus or something of that nature? Did you assess how their questionnaire scores compare to those who had increased esophageal acid exposure on pH testing? Such patients would clearly affect your results.

Closing discussant

Dr Kevin Chan: Thank you, Dr Demeester. You make a very good point. We have not undertaken further evaluation of the patients with symptoms and negative pH testing. It would be useful to determine whether these patients had nonacid reflux or simply an acid-sensitive esophagus. It is also possible that 48-h pH monitoring may have shown more abnormalities.