

Physicians and Patients Measure Different Dimension on Assessment for Gastroesophageal Reflux Disease-Related Symptoms

Juan Carlos Lopez-Alvarenga,^{1,3,4*} Sergio Sobrino-Cossio,² Ronnie Fass⁵ and Jose A Vargas-Romero³

¹Core of Biostatistics, Clinical Research Department, Hospital General de Mexico, Mexico City, Mexico; ²Department of Endoscopy, Instituto Nacional de Cancerologia, Mexico City, Mexico; ³Nycomed Mexico SA de CV, Mexico City, Mexico; ⁴Adjunt Scientist, Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, Texas, USA; ⁵Section of Gastroenterology, Department of Medicine, Southern Arizona VA Health Care System, Tucson, Arizona, USA

Background/Aims

Gastroesophageal reflux disease is a highly prevalent disease. Assessing treatment efficacy is critical in that clinical endpoints are properly evaluated. Clinical tools for symptoms severity assessment should be discriminative, predictive and evaluative.

Methods

In this study we compared a patient-oriented symptoms evaluation (ReQuest™) vs a structured interview assessment initiated by a physician (sickness impact profile [SIP]). Both questionnaires were analyzed in a multidimensional space using latent factors. Five dimensions were found: 1 for the short ReQuest™ questionnaire and 4 for SIP.

Results

We included 1,522 women and 1,296 men; mean age was 36 ± 7 years, and mean body mass index was 26 ± 4 . The score questionnaire assessment evaluation by physicians and patients did not correlate between them (between $r = 0.03$ and 0.26) except nausea and sleep disorder ($r = 0.45$ and 0.51) but both were sensitive enough to detect changes after treatment ($P < 0.05$). Medical specialty of the physician showed effect on the score of both, ReQuest™ and SIP evaluation. Questionnaire variance decomposition due to specialist was only 2% ($P < 0.05$).

Conclusions

While both evaluations are orthogonal (non-correlated), meaning patients and physicians measured diverse aspects of the same disease, they both were able to measure patient's improvement with treatment.

(J Neurogastroenterol Motil 2011;17:381-386)

Key Words

Monitoring, Physiologic; Pantoprazole; Questionnaires

Received: July 20, 2011 Revised: September 2, 2011 Accepted: September 7, 2011

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Correspondence: Juan Carlos Lopez-Alvarenga, MD, PhD

Core of biostatistics, Dirección de Investigación, Unidad 301-2do piso, Hospital General de Mexico, Dr. Balmis 148 Col. Doctores cp 06726, Mexico City, Mexico
Tel: +52-5004-3842, Fax: +52-5004-3843, E-mail: jclalvar@yahoo.com

Financial support: None.

Conflicts of interest: Nycomed Mexico SA de CV supported this study. Competing Interests: Dr. Vargas-Romero is Medical Director of Nycomed SA de CV. Dr. Lopez-Alvarenga is Biometrical advisor of Nycomed SA de CV.

Introduction

Gastroesophageal reflux disease (GERD) is a highly prevalent gastrointestinal disorder. It is critical that clinical endpoints are properly evaluated in order to assess treatment efficacy. Clinical tools for assessment of symptoms severity should be discriminative, predictive and evaluative.¹

Any evaluative scale should measure clinical changes (ie, symptoms severity) over the time and under treatment (sensitivity to detect changes); the scales should be objective and interpretable. Very few clinical tools have been developed and validated to evaluate the severity of GERD, and so far, there is no gold standard to determine the responsiveness of a measurement scale in GERD. Despite the subjectivity, GERD questionnaires such as Carlsson-Dent are sensitive enough to detect people with GERD than endoscopy in a region with low prevalence.² The subjective assessment of symptoms intensity as related to functional disease indicates that self-assessment is more accurate than physician evaluation. This concept is founded on the idea that only the person who suffers the pain knows its intensity, compared to a physician who can make an approximation of the patient's pain. Consequently, some researchers only pay attention to the patient's score of symptoms severity.³

Clinimetrics is a word derived from biometrics, and it refers to measurement of clinical symptoms and procedures which are helpful for diagnosis, characterization or evaluation of clinical entities. Any clinimetric measurement has its own variance, and the source of variation regarding symptoms intensity can be associated with factors like race, sex or age. Additionally, other factors external to patients can influence the results, such as medical specialty of the physician administering the interview.

In this study, we tested 2 hypotheses regarding the clinimetric approach to GERD assessment: (1) symptoms intensity score would correlate between physician and patient measurements. To test this, we compared the physician (from several specialties) measurement using a Likert scale of symptoms intensity with sickness impact profile (SIP) vs patients' measurement with ReQuestTM. We assumed that both scales would have high internal reliability and a positive coefficient of correlation. We measured their sensitivity to detect changes in symptoms severity after 4 weeks of pantoprazole magnesium (40 mg orally per day). (2) Physician specialty, patients' disease and characteristics are important source of variation in symptoms intensity scores. We used a variance decomposition of scores to determine whether the

amount of information (attributable percentage) of patients and physician specialty contributed to the scores.

Materials and Methods

Study Sample

This was a nation widely conducted study in private, state and hospital-based practices in Mexico. We included subjects between 18 to 45 years old (we excluded older people to decrease the probability of malignancy), with clinical history of heartburn, acid regurgitation, or both during the previous 3 months. These symptoms were reliable indicators of the presence of GERD, as has been found in other questionnaires.^{4,5} The present study included a wide sample of GERD patients with few clinical restrictions, and this design supported external validity of our results for daily clinical practice.

Questionnaires

We asked physician to conduct a structured interview (SIP) driven by questions that explore symptoms severity associated with GERD. This SIP has 18 questions divided into classical symptoms of GERD (heartburn and regurgitation), dyspepsia and extraesophageal manifestations of GERD. The interview asked patients to rate severity of heartburn, regurgitation, retching, halitosis, flatulence, nausea, sialorrhea, globus, discomfort, chest pain, dyspnea, chronic cough, early satiety and sleeping disturbances. The scale included categories such as "never," "rare," "sometimes," "most of the time" and "always." All of them were scored on a Likert scale of 4 points. We had demonstrated that this structured interview has enough sensitivity to detect group differences for symptoms severity or its changes by proton pump inhibitor treatment.^{6,7}

The patient directed-instrument, ReQuestTM, is a validated questionnaire with high internal consistency (Cronbach α = 0.90) and test-retest reliability (intraclass correlation coefficient between 0.86 for short version and 0.94 for long version).⁸ The short version has 6 questions with a visual analog scale of 10 cm long. The 6 questions vary in scope from general well being (quality of life) to acid-related complaints, upper abdominal-related complaints, lower abdominal-related complaints, nausea and sleep disorders.

Statistical Methods

Data were analyzed with SPSS version 15 and Statistica ver-

sion 8.0. The internal reliability was measured with Cronbach alpha coefficient,⁹ calculated for ReQuestTM and SIP separately.

Correlation between Likert scores and the ReQuestTM was calculated by Spearman's test. Dimensionality was assessed by latent factor analysis with a varimax rotation and Kaiser normalization of variables. Those factors with eigenvalue greater than 1 were considered as non compressed and to be able to represent one or more original variables, so they were considered significant because they explained an important amount of the variability in the data. The extracted scores were used as dependent variables.

Differences between specialties were analyzed using analysis of variance (ANOVA) and only those with $P < 0.05$ were considered for post hoc test with Fisher method (data expressed as mean \pm SEM). To understand the source of variance, we performed variance components analysis using type III ANOVA to calculate mean squares and the percentage of explained variance. The specialty of the physician interviewer was considered as a random effect; sex was fixed factor; age and body mass index were considered as covariates.

Ethics

All applicable international regulations concerning the ethical participation of our volunteers were followed during this research. The study and the informed consent document were ap-

proved by the Ethical Committee for Research of the Dr. Maximiliano Ruiz Castañeda General Hospital of Naucalpan; and the Center of Bioethics of the Medicine Faculty of the University of Guanajuato, Mexico. All subjects provided informed consent.

Results

Studied Sample

We invited 1,306 physicians to participate from four branches of medicine (gastroenterology, $n = 157$; internal medicine, $n = 218$; general surgery, $n = 65$; and general practitioner, $n = 866$). They evaluated 3,665 patients and 2,818 (77%) were included in this sample who completed the ReQuestTM questionnaire and were suitable for the analysis. We included 1,522 (54%) women and 1,296 men; their mean age was 36 ± 7 (SD) years, and the mean body mass index was 26 ± 4 (SD).

Internal Consistency and Correlation Between Sickness Impact Profile and ReQuestTM

The Cronbach alpha coefficient showed high internal consistency for ReQuestTM (0.87, $P < 0.05$) and SIP (0.84, $P < 0.05$). Most of the correlation coefficients between ReQuestTM

Table 1. Matrix of Correlations between Likert Scores and ReQuest[®] Score Calculated by Spearman's Test (N = 3,200)

Symptom (SIP)	Complaint (ReQuest TM)					
	Quality of life	Acid	Upper abdominal	Lower abdominal	Nausea	Sleep disorders
Heartburn	0.21	0.24	0.18	0.15	0.10	0.15
Regurgitation	0.21	0.23	0.19	0.17	0.13	0.20
Belching	0.14	0.15	0.15	0.16	0.13	0.14
Sialorrhea	0.08	0.06	0.08	0.19	0.12	0.16
Globus	0.08	0.06	0.07	0.14	0.07	0.13
Dysphagia	0.10	0.08	0.10	0.17	0.11	0.13
Odynophagia	0.11	0.08	0.08	0.15	0.10	0.14
Retching	0.07	0.04	0.09	0.17	0.18	0.14
Halitosis	0.10	0.10	0.13	0.19	0.14	0.17
Gastric pain	0.16	0.13	0.23	0.22	0.15	0.16
Early satiety	0.11	0.09	0.15	0.23	0.17	0.16
Nausea	0.15	0.14	0.18	0.23	0.45	0.20
Flatus	0.08	0.04	0.09	0.26	0.12	0.14
Chest oppression	0.09	0.07	0.14	0.16	0.09	0.13
Dyspnea	0.08	0.04	0.08	0.13	0.07	0.16
Chronic cough	0.05	0.03	0.04	0.12	0.09	0.17
Dysphonia	0.03	0.01	0.04	0.11	0.04	0.13
Sleep disorders	0.19	0.14	0.17	0.22	0.17	0.51

All correlations showed P -value < 0.05 .

and SIP were small (between 0.03 and 0.26); only nausea ($r = 0.45, P < 0.001$) and sleep disorders ($r = 0.51, P < 0.001$) were highly correlated. Table 1 shows this matrix correlation.

Dimensionality of Scales

The factor analysis including both measurements showed 5 main factors, and one of these factors (F1) was the ReQuest™ itself. The other four factors can be described as F2: dyspepsia; F3: sleep disorders and cough; F4: symptoms associated with larynx-upper esophagus region such as hoarseness, odynophagia, dysphagia and globus; and F5: classical GERD symptom (heartburn and regurgitation). The obtained factors, calculated by principal component analysis, were orthogonal to each other;

which meant by definition that they did not have correlation between calculated scores.

Variance Decomposition of the Scores

The ANOVA found differences in symptoms intensity and ReQuest™ scores by physician specialty (Table 2). Gastroenterologists gave the highest scores for heartburn and regurgitation; surgeons gave the highest scores for symptoms associated with upper respiratory or esophageal symptoms like odynophagia, dysphagia and globus. General practitioners gave the highest scores for other symptoms like flatus and dyspepsia (Table 2). The sex adjustment showed that women had higher scores for symptoms severity than men (data not shown) as has been found in other

Table 2. Comparison of Sickness Impact Profile and ReQuest™ Scores by Physician Specialty

Variable	Gastroenterology (n = 395)	Internal medicine (n = 624)	Surgeon (n = 133)	General practitioner (n = 2,048)	P-value
Demographics					
Age (yr)	35.5 ± 0.3 ^a	37.2 ± 0.2 ^b	36.6 ± 0.5 ^c	35.6 ± 0.1 ^{a,c}	< 0.001
Sex (M/F)	172/223	285/339	61/72	961/1,087	0.653
BMI	25.5 ± 0.2 ^c	26.2 ± 0.1 ^b	25.6 ± 0.4 ^{b,c}	26.4 ± 0.1 ^b	0.002
Likert score (SIP)					
Heartburn	2.38 ± 0.03 ^b	2.21 ± 0.02 ^c	2.00 ± 0.06 ^a	2.24 ± 0.01 ^c	< 0.001
Regurgitation	2.14 ± 0.04 ^b	2.13 ± 0.03 ^b	1.94 ± 0.06 ^c	2.16 ± 0.01 ^b	0.024
Belching	1.49 ± 0.05	1.53 ± 0.03	1.35 ± 0.08	1.56 ± 0.02	0.062
Sialorrhea	0.66 ± 0.04	0.67 ± 0.03	0.60 ± 0.08	0.71 ± 0.02	0.358
Globus	0.74 ± 0.04 ^a	0.76 ± 0.03 ^a	1.10 ± 0.09 ^b	0.87 ± 0.02 ^c	< 0.001
Dysphagia	0.50 ± 0.04 ^a	0.56 ± 0.03 ^a	0.90 ± 0.08 ^b	0.74 ± 0.02 ^c	< 0.001
Odynophagia	0.36 ± 0.03 ^a	0.47 ± 0.03 ^c	0.89 ± 0.08 ^b	0.53 ± 0.01 ^c	< 0.001
Retching	0.43 ± 0.03	0.48 ± 0.03	0.52 ± 0.07	0.52 ± 0.01	0.235
Halitosis	0.84 ± 0.05 ^c	0.89 ± 0.04 ^c	0.90 ± 0.08 ^{b,c}	1.00 ± 0.02 ^b	0.005
Gastric pain	1.78 ± 0.05 ^c	1.81 ± 0.03 ^c	1.45 ± 0.09 ^a	1.89 ± 0.02 ^b	< 0.001
Early satiety	1.08 ± 0.05 ^{b,c}	1.03 ± 0.04 ^{b,c}	0.93 ± 0.08 ^c	1.12 ± 0.02 ^b	0.062
Nausea	1.06 ± 0.05 ^b	1.04 ± 0.03 ^b	0.82 ± 0.08 ^c	1.04 ± 0.02 ^b	0.090
Flatus	0.95 ± 0.05 ^c	0.95 ± 0.04 ^c	0.93 ± 0.08 ^c	1.12 ± 0.02 ^b	< 0.001
Chest oppression	1.05 ± 0.05 ^{b,c}	1.13 ± 0.04 ^b	0.88 ± 0.08 ^c	1.16 ± 0.02 ^b	0.016
Dyspnea	0.30 ± 0.03	0.37 ± 0.03	0.29 ± 0.05	0.34 ± 0.01	0.426
Chronic cough	0.41 ± 0.04 ^b	0.49 ± 0.03 ^b	0.77 ± 0.08 ^c	0.45 ± 0.01 ^b	< 0.001
Dysphonia	0.45 ± 0.04 ^c	0.54 ± 0.03 ^c	0.77 ± 0.08 ^b	0.48 ± 0.01 ^c	< 0.001
Sleep disorders	1.00 ± 0.05 ^c	1.06 ± 0.04 ^c	1.03 ± 0.09 ^{b,c}	1.22 ± 0.02 ^b	< 0.001
ReQuest™					
Quality of life	6.12 ± 0.14	6.17 ± 0.11	5.85 ± 0.23	6.33 ± 0.06	0.127
Acid complaints	6.24 ± 0.15	6.25 ± 0.12	5.71 ± 0.25	6.39 ± 0.06	0.068
Upper abdominal or stomach somplaints	5.82 ± 0.15 ^b	5.97 ± 0.12 ^b	5.18 ± 0.26 ^c	6.10 ± 0.06 ^b	0.003
Lower abdominal or digestive complaints	4.42 ± 0.16 ^c	4.57 ± 0.13 ^{b,c}	4.40 ± 0.26 ^{b,c}	4.82 ± 0.06 ^b	0.036
Nausea	3.80 ± 0.16 ^{b,c}	4.01 ± 0.13 ^{b,c}	3.22 ± 0.25 ^c	4.04 ± 0.07 ^b	0.025
Sleep disorders	4.09 ± 0.17 ^c	4.23 ± 0.13 ^{b,c}	3.84 ± 0.29 ^c	4.48 ± 0.07 ^b	0.033

^{a,b,c}Indicate homogeneous groups using Scheffe contrast, where $b > c > a$.

BMI, body mass index; SIP, sickness impact profile.

Data were expressed as mean ± SEM.

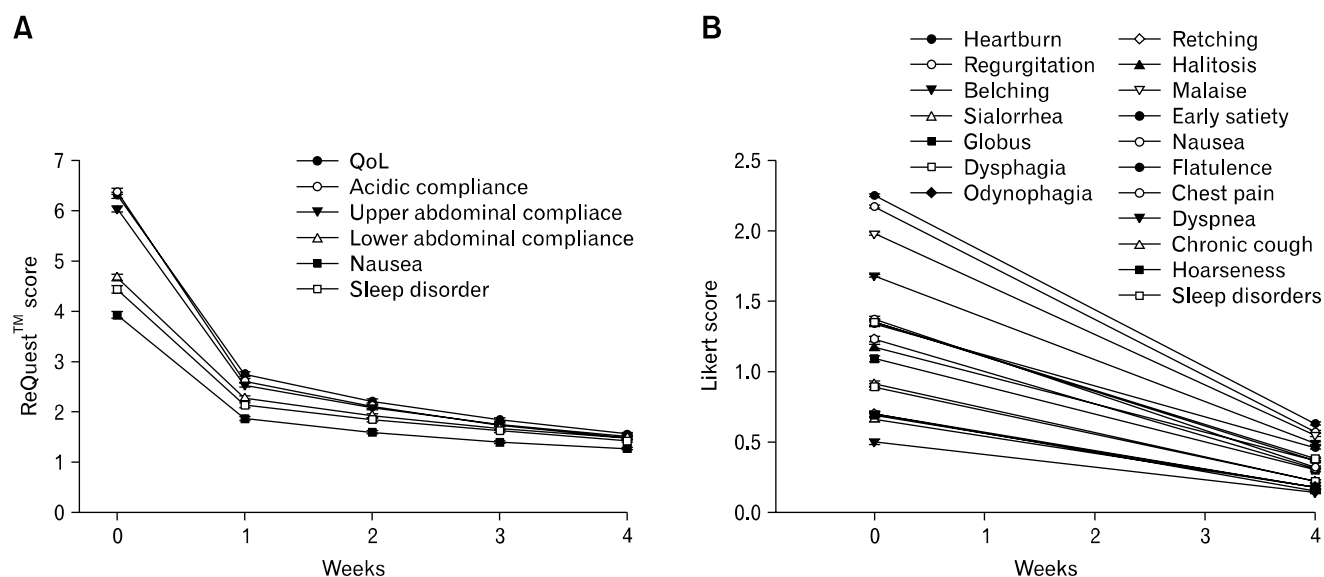


Figure. Symptoms improvement after 4 weeks of treatment with pinverium bromide plus symeticon. (A) It Shows patient evaluation with ReQuest™. (B) It shows physician evaluation with a structured interview, despite the lack of correlation both measurements improved significantly with treatment. Both assessment showed significant amelioration of symptoms ($P < 0.05$ for all variables).

studies.

The variance decomposition showed the effect of specialty in less than 2% compared to the total variance. Interestingly, despite the lack of correlation between both evaluations, SIP and ReQuest™, showed a clinically and statistically significant response after the 4 weeks pantoprazole treatment (Figure).

Discussion

This study demonstrated that symptoms severity of GERD assessed by physicians did not correlate with symptoms severity assessed by patients. This could be interpreted as assessment of different aspects (2 different points of view of the same patients' symptoms). Interestingly, the physician-administered and patient-administered tools have high internal reliability for measurement of the same disease, but they were non-correlated as we demonstrated orthogonality in a multiple dimensional-mathematical space (latent factor analysis). In other words, they measure diverse dimensions, and one of them is occupied by all ReQuest™ questions, while the other four factors were measured by the physician evaluation with SIP. However, both tools recorded improvement of patients' symptoms after 4 weeks of pantoprazole magnesium treatment, which meant that GERD measured in multidimensional traits improved with treatment.

Clinimetry Should Be Evaluated by Physicians or Patients?

We expected a positive correlation between the questions of the structured interview and ReQuest™, but instead we found that evaluation of SIP and ReQuest™ was orthogonal. Thus, even though both physicians and patients were evaluating the same disease, the dimensions measured were not the same. After treatment, measurements by both tools showed improvement.

Neither of the measurements showed any great advantage over the other in evaluating symptoms severity during treatment. Moreover, the multidimensionality of the evaluation helped to check improvement not only for classical GERD symptoms (heartburn and regurgitation) but also for extraesophageal and overlapping symptoms. Some of the score variations in both tools were due to: perception associated with the medical specialty of the physician and women's perception of symptoms severity. We think that in a larger sample size, the effect of the medical specialty of the participating physicians would diminish the total error variation. For a smaller sample size, the stratification of specialties becomes the rule. Perhaps then the next questions should be, when to evaluate the symptoms, and who should initiate the evaluation: physician or patients?

Advantages of the Method

The methods we used had some interesting components. We

invited a large number of centers to participate in the study in order to increase external validity of our results for daily clinical practice.

It has been recommended that patient-driven evaluation of symptoms is the most effective. The logic of this suggestion is based on the subjectivity of evaluation about severity on any functional disease. Patients know better than physicians how they feel.¹⁰ We compared 2 tools that should be antagonistic from this point of view, a tool filled by patients and another tool filled by physician. The high alpha coefficient support that the structured questions and the ReQuestTM have a coherent structure, the lack of correlation with the orthogonal dimension corroborated by factor analysis, which is consistent with ReQuestTM being an unidimensional measurement, while the structured interview covers 4 dimensions. We analyzed the variance composition of the latent factor scores and found the medical specialty of the physician influenced the perception of symptoms severity of GERD.

Limitations of the Study

The relationship between ReQuestTM and the structured interview remains unclear; despite the fact that they were uncorrelated, both tools measured improvement with treatment, meaning that they had a communality affected by treatment. We will study further the co-decomposition of variance between these tools.

The factor analysis should be used to rebuild the structured interview. The factor scores obtained are non-dimensional, however, they can be helpful to select a new clusters of questions to evaluate symptoms intensity for GERD.

Conclusions

We conclude that the patient- and physician-driven tools do not correlate because they measure diverse orthogonal dimensions. However, both tools are sensitive enough to detect favorable changes in symptoms severity associated with treatment with pantoprazol magnesium after 4 weeks.

Acknowledgements

Dr. Vargas-Romero conceptualized the study with critical

analysis and wrote the paper, Dr. Lopez-Alvarenga and Dr. Sobrino-Cossio analyzed and wrote the paper. Dr. Ronnie Fass made critical review of the manuscript and included structural modifications to the paper.

References

1. Stanghellini V, Armstrong D, Mönnikes H, Bardhan KD. Systematic review: Do we need a new gastro-oesophageal reflux disease questionnaire? *Aliment Pharmacol Ther* 2004;19:463-479.
2. Netinatsunton N, Attasaranya S, Ovarlarnporn B, Sangnil S, Boonviriyi S, Piratvisuth T. The Value of Carlsson-dent questionnaire in diagnosis of gastroesophageal reflux disease in area with low prevalence of gastroesophageal reflux disease. *J Neurogastroenterol Motil* 2011;17:164-168.
3. Irvine EJ, Feagan BG, Wong CJ. Does self-administration of a quality of life index for inflammatory bowel disease change the results? *J Clin Epidemiol* 1996;49:1177-1185.
4. Locke GR, Talley NJ, Weaver AL, Zinsmeister AR. A new questionnaire for gastroesophageal reflux disease. *Mayo Clin Proc* 1994;69:539-547.
5. Moreno Elola-Olaso C, Rey E, Rodríguez-Artalejo F, Locke GR 3rd, Díaz-Rubio M. Adaptation and validation of a gastroesophageal reflux questionnaire for use on a Spanish population. *Rev Esp Enferm Dig* 2002;94:745-758.
6. Lopez-Colombo A, Lopez-Alvarenga JC, Vargas J, et al. Do prior pregnancies modify the intensity of symptoms related to GERD? A report of the Mexican GERD working group. *Gut* 2008;57(suppl 2):A312.
7. Lopez L, Lopez-Alvarenga JC, Comuzzie AG, Gonzalez J, Crespo Y, Vargas J. Nighttime GERD symptoms associated with dyspepsia, esophageal discomfort, and extraesophageal complaints improvement after treatment with Pantoprazole magnesium; 40 mg daily for 28 days. *Gastroenterology* 2008;134(suppl 1):A176-A177.
8. Mönnikes H, Bardhan KD, Stanghellini V, Berghöfer P, Bethke TD, Armstrong D. Evaluation of GERD symptoms during therapy. Part II. Psychometric evaluation and validation of the new questionnaire ReQuest in erosive GERD. *Digestion* 2004;69:238-244.
9. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297-334.
10. Brandt LJ, Bjorkman D, Fennerty MB, et al. Systematic review on the management of irritable bowel syndrome in North America. *Am J Gastroenterol* 2002;97(11 suppl):S7-S26.