Journal of Pediatric Gastroenterology and Nutrition
41:178–185 © August 2005 Lippincott Williams & Wilkins, Philadelphia

Age-Specific Questionnaires Distinguish GERD Symptom Frequency and Severity in Infants and Young Children: Development and Initial Validation

*Linda Deal, †Benjamin D. Gold, ‡David A. Gremse, §Harland S. Winter, *Suzanne B. Peters, *Polly D. Fraga, *Michael E. Mack, *Stefanie M. Gaylord, ‖Vasundhara Tolia, and ¶Joseph F. Fitzgerald

**Wyeth Research, Collegeville, Pennsylvania; †Emory University School of Medicine, Atlanta, Georgia; ‡University of South Alabama, Mobile, Alabama; §MassGeneral Hospital for Children, Boston, Massachusetts; ‖Children's Hospital of Michigan, Detroit, Michigan; ¶Indiana School of Medicine and James Whitcomb Riley Hospital for Children, Indianapolis, Indiana*

ABSTRACT

Two gastroesophageal reflux disease (GERD) symptom questionnaires were developed and tested prospectively in a pilot study conducted in infants (1 through 11 months) and young children (1 through 4 years) with and without a clinical diagnosis of GERD. A pediatric gastroenterologist made the clinical diagnosis of GERD. Parents or guardians at 4 study sites completed the questionnaires, providing information on the frequency and severity of symptoms appropriate to the 2 age cohorts. In infants, symptoms assessed were back arching, choking or gagging, hiccups, irritability, refusal to feed and vomiting or regurgitation. In young children, symptoms assessed were abdominal pain, burping or belching, choking when eating, difficulty swallowing, refusal to eat and vomiting or regurgitation. Respondents were asked to describe additional symptoms. Symptom frequency was the number of occurrences of each symptom in the 7 days before completion of the questionnaire. Symptom severity was rated from 1 (not at all severe) to 7 (most severe). An individual symptom score was calculated as the product of symptom frequency and severity scores. The composite symptom score was the sum of the individual symptom scores. The mean composite symptom and individual symptom scores were higher in infants ($P < 0.001$ and $P < 0.05$, respectively) and young children ($P < 0.001$ and $P < 0.05$, respectively) with GERD than controls. Vomiting/regurgitation was particularly prevalent in infants with GERD (90%). Both groups with GERD were more likely to experience greater severity of symptoms. We found the GERD Symptom Questionnaire useful in distinguishing infants and young children with symptomatic GERD from healthy children. ***JPGN 41:178–185, 2005.*** © 2005 Lippincott Williams & Wilkins

INTRODUCTION

The North American Society for Pediatric Gastroenterology, Hepatology and Nutrition clinical practice guidelines for pediatric GERD define gastroesophageal reflux (GER) as the passage of gastric contents into the esophagus and gastroesophageal reflux disease (GERD) as symptoms or complications resulting from the retrograde passage of gastric refluxate (1). There are limited pediatric studies by which to assess the true prevalence of GERD in children. Available data suggest that the prevalence of GER is age-dependent with the highest prevalence in infants younger than 6 months (2). Although there are few longitudinal studies, the prevalence of GERD, which may be a lifelong condition after infancy, appears to increase with age (3–6). The poor understanding of the epidemiology of childhood GERD may be a result of the lack of population-based studies, variable clinical presentations, lack of a consistent diagnostic approach and poor or inconsistent ''case'' definitions (3).

GERD may present with a variety of esophageal and extra-esophageal symptoms with an impact on a child's health and quality of life (1,3). Symptoms of GERD in children differ with age (7) and depend in part on the specific complications of GERD (8). The presenting signs and symptoms include spitting up/regurgitation, vomiting, poor weight gain, dysphagia, abdominal or substernal pain, esophagitis and respiratory disorders (9).

Existing questionnaires designed to aid pediatricians in decisions on GERD diagnosis and treatment in infants (10,11) and children (4–6) were not used in this study for several reasons. Although the Infant Gastroesophageal Reflux Questionnaire (I-GERQ) developed by Orenstein et al. (10) is validated, until recently (12) it was of limited

Received May 17, 2004; accepted May 16, 2005.

Address correspondence and reprint requests to Linda Deal, Associate Director, Wyeth Global Health Outcomes Assessment 500 Arcola Road Collegeville, PA 19426-3930. (e-mail: deall@wyeth.com).

Wyeth Research sponsored this study.

use for evaluating the response of GERD symptoms to therapeutic treatment because of the absence of a defined recall period for framing the assessment. Furthermore, although both the original I-GERQ and its short form, the Infant Gastroesophageal Reflux Questionnaire-Shortened Form (IGER-SF) adapted by Nelson et al. (4), assess symptom frequency and quality, severity assessments are somewhat limited, making response to treatment difficult to follow. Finally, the median time (20 minutes) for completion of the I-GERQ (11) limits its practical use and the shorter version (IGER-SF) lacks documented validation (4). Because the frequency and severity of GERD symptoms varies with age we studied two age cohorts: infants 1 through 11 months and children 1 through 4 years. Neither of the existing questionnaires contains questions specific to the 1 through 4 year-old cohort.

Our primary objective was to develop questionnaires that were valid to assess the response of GERD symptoms to therapeutic treatment over a defined time in two age cohorts; infants 1 through 11 months (GSQ-I) and children 1 through 4 years (GSQ-YC). As part of the development and validation process, a pilot study was conducted to: (1) confirm the appropriateness and relevance of the selected GERD symptoms (ie, to test the face and content validity), (2) test the adequacy of frequency and severity scales (ie, to test the range of measurement that the scales permitted), (3) test the questionnaire ease of use, and (4) confirm that symptom scores were higher in children with a clinical diagnosis of GERD than in healthy children without any gastrointestinal disorder (ie, to evaluate the sensitivity and specificity).

METHODS

Questionnaire Development

Questionnaires were designed to be completed by a parent/guardian on behalf of the study participant.

Selection of Symptoms

Preliminary symptoms were derived from the literature and refined by the study investigators and expert consultants to form the pilot set of symptoms. For infants, the symptoms assessed were back arching, choking/gagging, episodes of hiccups, irritability/fussiness, refusal to feed and vomiting/regurgitation. For young children, the symptoms assessed were abdominal/belly pain, burping/belching, choking when eating, difficulty swallowing, refusal to eat and vomiting/regurgitation. Symptom descriptors in the questionnaire were reviewed and edited for clarity and appropriateness. To ensure that important symptoms were not missed, the questionnaires included an ''other'' symptom category that allowed open-ended responses by parents/guardians.

Scale Development

The GSQ-I and GSQ-YC were designed to assess frequency and usual severity of symptoms in the preceding 7 days. This recall period was selected to balance the burden of the time spent completing the questionnaire against recall bias associated with longer periods. Although symptom distress (ie, the degree of ''bother'' associated with a symptom) and symptom duration were also considered for inclusion in the questionnaire, they were not included because of their high correlation with symptom severity (13).

Parents/guardians were asked to indicate the frequency of each symptom (or episode of symptoms) in the previous 7 days by providing a whole number of zero or greater. There was no restriction on the maximum number of occurrences. Respondents also rated the usual severity of each occurrence on a scale from 1 (not all severe) to 7 (most severe). Label descriptors (anchors) were provided only for the extreme values of the severity scale. If a symptom frequency was zero, respondents were instructed to skip the severity assessment for that symptom.

To accommodate the frequency and severity scales, certain symptoms such as hiccups were defined based on episode. For example, rather than counting each hiccup, parent/guardians were instructed to count each episode of hiccups (defined as a series of hiccups with a clear beginning and ending). Respondents were instructed that the average length of the episode and the average number of hiccups within the episode should be considered in the severity rating.

Scoring

An individual score for each symptom (ISS) and an overall or composite symptom score (CSS) were calculated. The ISS was defined as the product of the number of times the symptom occurred (frequency) and the severity of that symptom (ranging from 1 ''not at all severe'' to 7 ''most severe'') in the previous 7 days. The CSS was calculated as the sum of the ISSs. Calculation of ISSs and CSS allowed identification of the extent to which an individual symptom contributed to the overall symptom complex.

Validity Assessments

Face and content validity: These terms are used to determine whether a scale appears reasonable. Face validity ascertains whether a questionnaire appears to assess the desired qualities; content validity ascertains whether a questionnaire appears to characterize all the relevant or important content and domains (the different, measurable characteristics of the subject being studied) (14). This assessment was accomplished during development of the instruments by expert review of GERD symptoms for infants and young children, the symptom definitions, and the severity scales, and by including an ''other'' symptom category on the questionnaire. Symptoms entered in the ''other'' section with great frequency for the clinically diagnosed GERD patients were reviewed to determine the need to include them in the final questionnaires. Similarly, if queried symptoms appeared with very low frequency or had statistically similar prevalence and frequency in both the clinically diagnosed group and controls, these symptoms might be dropped from the questionnaire.

Measurement range: Scales assessing a full range of health status allow for more precision in measuring differences in health or changes in health over time and have more power for hypothesis testing (15). We assessed the range of measurement of this questionnaire by investigating the distribution of scores in a sample from the healthy population and in samples from the two GERD cohorts.

A ceiling (floor) effect results when a maximum (minimum) scale value is selected with great frequency. To avoid this effect,

respondents were not restricted to a maximum frequency on the questionnaires. Although we did set a minimum value of zero for the frequency scale, we were not concerned about a floor effect because frequent values of zero would indicate that the symptom should be removed from the questionnaire. We set both a minimum and a maximum value for severity scales. Any clustering toward the extreme values (skewness) would indicate a ceiling or floor effect and would suggest that either the range should be expanded or that more or alternate anchor labels be provided.

Ease of Use

To evaluate the ease of use of the questionnaires, confusion reported by respondents regarding symptoms, their descriptors or the scales was noted, as were missing responses or text in numerical fields. The study coordinators at each site observed the approximate time required for completion the questionnaires.

Sensitivity and Specificity

The test of sensitivity for the questionnaire was its ability to detect the presence of GERD symptoms and disease (ie, true cases) when symptom and disease were truly present as assessed by an experienced pediatric gastroenterologist. The test for specificity was its ability to detect the absence of GERD symptoms and disease when truly absent. This study included patients with symptomatic GERD and healthy controls. By comparing the proportion of reported symptoms between each age cohort and its control group we could determine if the symptoms were indicative of GERD and not common or present among the subjects in the control group. Because we knew which patients had a prior diagnosis of GERD, our intent was to determine a CSS criterion that would achieve reasonable sensitivity and specificity for further studies employing these questionnaires.

Pilot Study

Design

This was a nontreatment, multiple-center, parallel-design, pilot study conducted at four centers in the United States to assess two questionnaires. Enrollment of approximately 60 infants (1 through 11 months) and approximately 60 young children (1 through 4 years) with and without symptomatic GERD (40 GERD patients, 20 healthy controls per age cohort) was planned. Questionnaires were completed at a single study visit by the parents/guardians of infants (GSQ-I) and young children (GSQ-YC) clinically diagnosed with GERD and a control group of ''well visit'' patients with no gastrointestinal complaints, in each of the two age cohorts. Respondents reported medications taken in the 7 days before completion of the questionnaire. There were no restrictions on prior or concomitant medications. Data were collected on the method of diagnosis for patients with GERD. A study coordinator or the investigator, or both, was available for questions by respondents to facilitate completion of the questionnaires. All sites had institutional review board (IRB) approval. An IRB-approved informed consent form was completed by the parent/guardian before questionnaire completion.

Population

This study included 128 male and female subjects aged 1 month through 4 years who had had not yet celebrated their fifth birthday) with and without GERD. The GERD population was defined as patients clinically diagnosed by an experienced pediatric gastroenterologist with symptomatic GERD as demonstrated by the presence of typical GERD symptoms with or without confirmatory tests. The control populations were healthy infants and young children with no gastrointestinal disorder.

TABLE 1. *Demographic and baseline characteristics: infants 1 through 11 months of age*

	Healthy	GERD	Total
	(n = 23)	(n = 41)	(n = 64)
Age (months)			
Mean + SD	5.1 ± 3.1	5.6 ± 2.5	5.4 ± 2.7
Median (range)	4.2 (0.9–11.4)	5.1 (1.4–11.4)	4.8 (0.9–11.4)
Sex, n (%)			
Boys	9 (39)	24 (59)	33 (52)
Girls	14 (61)	17 (41)	31 (48)
Ethnic origin, n (%)			
Asian/Asian American	2 (9)	0	2 (3)
Black	11 (48)	6 (15)	17 (27)
Hispanic	0	2 (5)	2 (3)
Biracial	0	1 (2)	1 (2)
White	10 (43)	32 (78)	42 (66)
Method of diagnosis, n (%)			
Symptoms	N/A	40 (98)	
Upper GI	N/A	9 (22)	
pH	N/A	6 (15)	
Endoscopy +/− histology	N/A	4 (9.8)	
Oxypneumogram	N/A	1 (2.4)	
Radionuclide study	N/A	1 (2.4)	

GI, gastrointestinal study.
Not all percentages add up to 100% because some patients were diagnosed by more than one criterion.

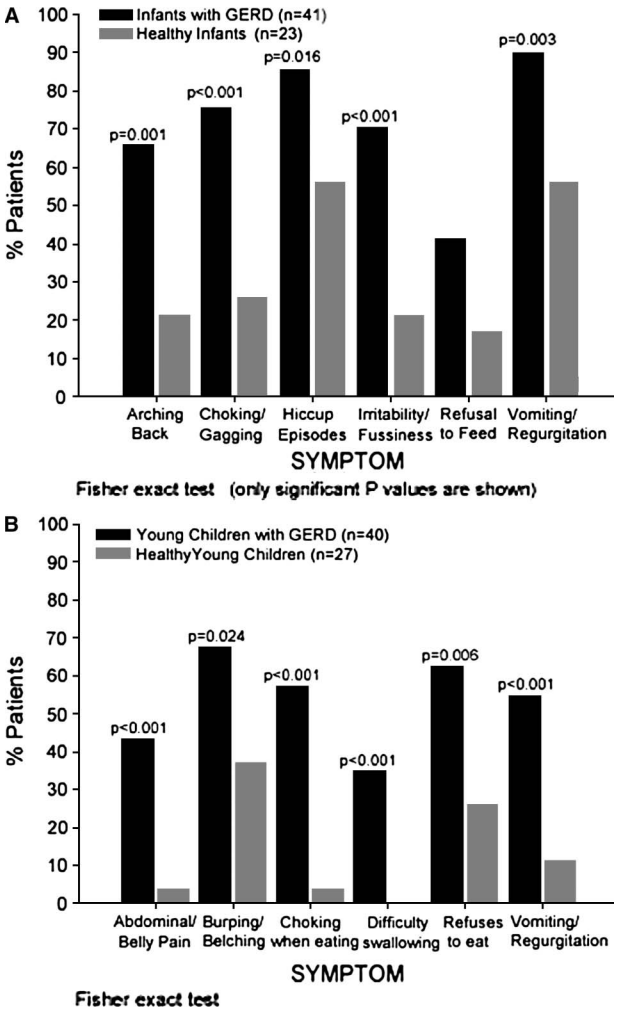

FIG. 1a. Prevalence of symptoms among GERD infants and age-matched controls 1 through 11 months of age. FIG. 1b. Prevalence of symptoms among GERD children and age-matched controls 1 through 4 years of age.

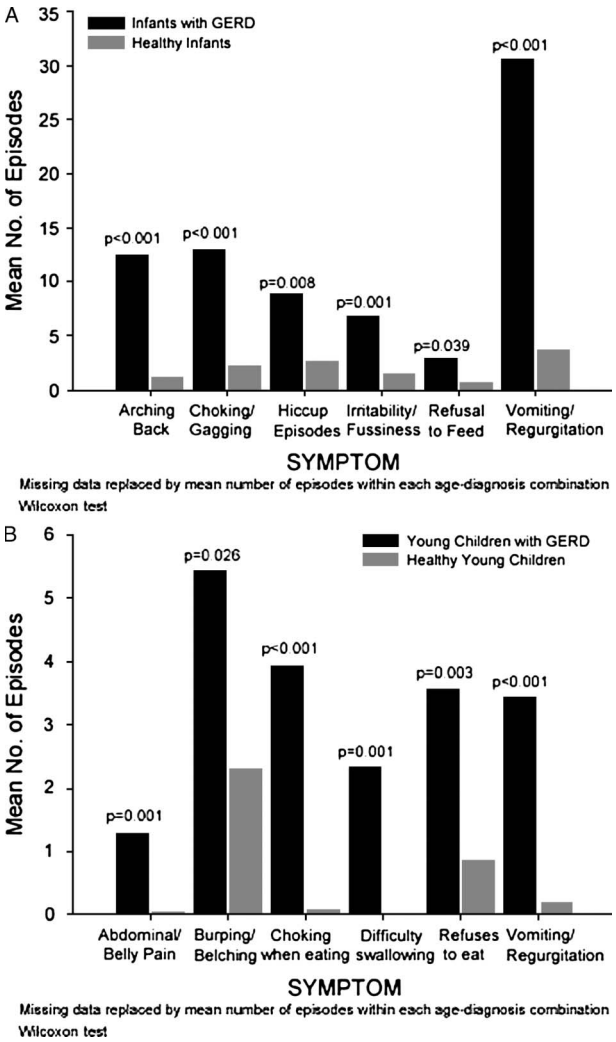

FIG. 2a. Frequency of symptom episodes among GERD infants and age-matched controls 1 through 11 months of age. **FIG. 2b.** Frequency of symptom episodes among GERD children and age-matched controls 1 through 4 years of age.

Statistical Methods

The study was powered to detect a difference in the mean CSS at the 0.05 level with a projected sample size of 40 patients with GERD and 20 subjects in the control group for each age cohort. It was assumed that the mean CSS would be 60 for the patients with GERD and 20 for the control group with a common standard deviation of 50.

The proportion of infants or young children with each symptom (ie, the prevalence of the symptom) was compared to the controls by the Fisher exact test. Frequency and severity differences between the two groups were evaluated for each age cohort using the Wilcoxon test. In comparing symptom severity, only data from children experiencing the symptom were used. The proportion of male and female infants or young children with each symptom was compared. The CSSs and ISSs were compared using the Wilcoxon test.

In both groups, missing frequency or severity scores were replaced by the mean frequency or severity score for that symptom from the other subjects in the corresponding diagnostic group and age cohort. For example, if the severity score for vomiting was missing for a healthy infant, then the mean severity score for vomiting from the other healthy infants was assigned when calculating the ISS and CSS for vomiting.

RESULTS

Ease of Use

Parents/guardians of the two GERD cohorts spent a mean of 7 minutes completing the questionnaire. The majority of time was spent on the prior medications section. Study site coordinators reported that belching/burping was the only symptom needing additional clarification. At one site, the parent/guardians of 5 infants used text in a numeric field (eg, hiccups: ''too many to count''). In the

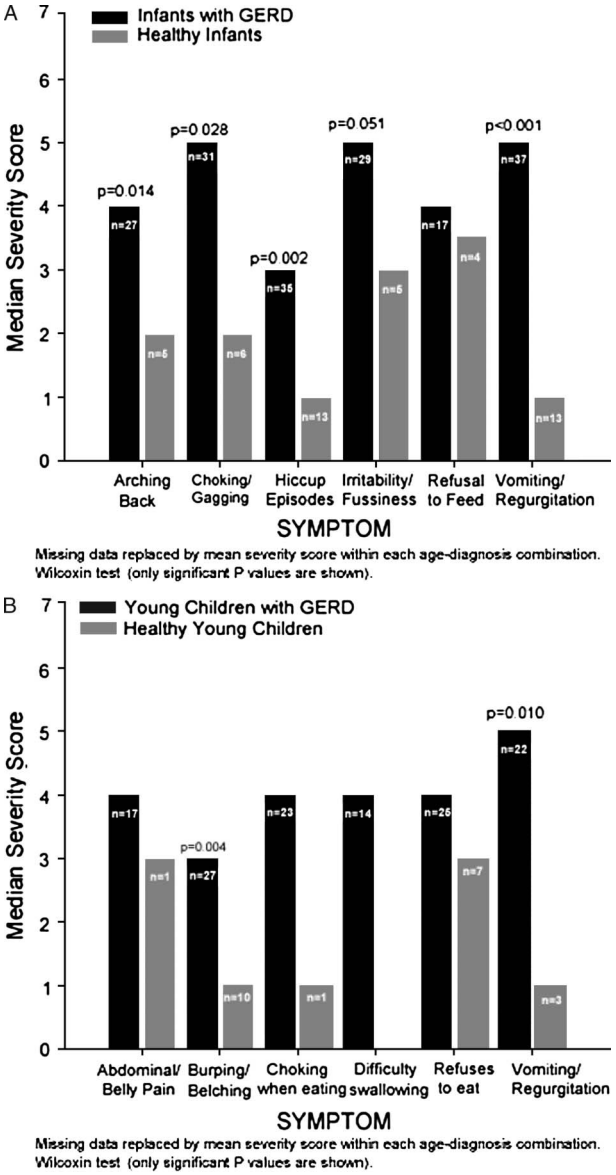

FIG. 3a. Severity of symptoms among GERD infants and controls. **Fig. 3b.** Severity of symptoms among GERD children and controls.

young group, some respondents reported such a high frequency of burping/belching episodes we assumed they may have been reporting the actual number of burps or belches instead of the number of episodes.

Infants

The parents/guardians of 64 infants aged 1 through 11 months (median age, 4.8 months) participated. Of these, 41 (64%) had GERD and 23 were age-matched controls. The majority (98%) of infants with GERD was diagnosed by clinical symptoms, but one or more diagnostic tests were also used in 52.2%. The GERD infants and controls were similar in age and although the GERD group had 16% more boys than girls and the control group had 22% more girls than boys, the difference was not significant. Sixty-six percent of the infants were white (Table 1). Thirty-one of the GERD infants (76%) had taken one or more medications for GERD in the 7 days before the questionnaire: proton pump inhibitor (n = 10), histamine-2 receptor antagonist (24), prokinetic (n = 11), antigas (n = 2), and antacid (n = 1).

All 64 questionnaires were analyzed. GERD infants had a significantly higher ($P < 0.05$) prevalence of symptoms than controls with the exception of ''refusal to feed,'' which approached but did not reach statistical significance ($P = 0.058$). The most prevalent symptom among GERD infants was ''vomiting/regurgitation'' (90%), which was significantly higher than in control subjects ($P = 0.003$). The least prevalent symptom among GERD infants was ''refusal to feed'' (42%) (Fig. 1a). All symptoms occurred with significantly greater frequency in GERD infants than in controls. The symptom with the highest frequency in GERD infants was ''vomiting/regurgitation'' (mean, 30.6 episodes/7 days). The symptom with lowest frequency was ''refusal to feed'' (mean, 2.8 episodes/7 days) (Fig. 2a). Additional symptoms reported manually in GERD patients were constipation in 4, cough in 3 and poor sleeping or night-time waking in 2.

GERD infants were reported to have more severe symptoms ($P < 0.05$) than controls for all symptoms except ''refusal to feed'' ($P = 0.56$) and ''irritability/fussiness'' ($P = 0.051$). Three of the six symptoms had a median score of 5 on the 7-point severity scale compared to a median of 3 or less for the same symptoms in controls (Fig. 3a).

The mean ISS and CSS were significantly higher for GERD infants than for controls. ''Vomiting/regurgitation'' (ISS = 151.9) contributed most (~40%) to the overall CSS (373.1). Arching back, choking/gagging, hiccups, and irritability/fussiness each contributed at least 10% to the overall CSS with ''refusal to feed'' contributing only ~4% (Table 2). There were no statistical differences between males and females in the number of episodes, CSS ($P = 0.91$), or ISS ($P \geq 0.43$).

A subgroup analysis of symptom frequency, ISS and CSS in which data for infants with GERD diagnosed by objective testing (n = 15) were compared with all controls showed that all measures except ''refusal to feed'' remained significantly different. We found that CSS >27 had a sensitivity of 90% for detecting patients with GERD (37/41 patients with GERD had scores >27) and a specificity of 83% (19/23 control group subjects had scores ≤27). Lowering the threshold CSS to >16 raised the sensitivity to 95% but lowered specificity to 74%.

Young Children

The parents/guardians of 67 subjects 1 through 4 years (median age, 29 months) participated. Of these 40 were

TABLE 2. *Symptom frequency, individual (ISS) and composite symptom score (CSS): infants 1–11 months*

Symptom	Measurement	GERD (mean ± SD)	Healthy (mean ± SD)	*P* values	
				t-test	Wilcoxon
Arching back	Frequency	12.3 ± 19.3	1.1 ± 3.1	0.007	0.000
	ISS	58.3 ± 101.6	1.9 ± 5.9	0.010	0.000
Choking/gagging	Frequency	12.9 ± 24.2	2.2 ± 6.0	0.042	0.000
	ISS	71.0 ± 162.4	6.1 ± 18.0	0.062	0.000
Episodes of hiccups	Frequency	8.8 ± 13.2	2.6 ± 3.5	0.032	0.008
	ISS	42.4 ± 87.3	5.0 ± 7.6	0.045	0.001
Irritability/fussiness	Frequency	6.7 ± 9.6	1.4 ± 3.4	0.013	0.001
	ISS	35.5 ± 56.1	4.8 ± 12.7	0.012	0.000
Refusal to feed	Frequency	2.8 ± 5.6	0.6 ± 1.7	0.069	0.039
	ISS	14.1 ± 31.8	2.7 ± 10.0	0.099	0.039
Vomiting/regurgitation	Frequency	30.6 ± 43.9	3.7 ± 9.0	0.005	0.000
	ISS	151.9 ± 222.9	11.5 ± 43.4	0.004	0.000
Composite score	CSS	373.1 ± 510.0	32.1 ± 61.6	0.002	0.000

GERD, gastroesophageal reflux disease.

diagnosed as GERD children. Diagnosis was based on clinical symptoms in 37, but one or more diagnostic tests had been performed in 29. The GERD and controls were similar in age. The GERD group had slightly more boys than girls. The majority of children in both groups were white (Table 3). Twenty-six GERD children (65%) had taken medication for GERD in the 7 days before the questionnaire: proton pump inhibitor (n = 14), histamine-2 receptor antagonist (n = 12), prokinetic (n = 8), and antacid (n = 1).

All 67 questionnaires were analyzed. The GERD children had a significantly higher prevalence of symptoms ($P < 0.05$) than controls. The most prevalent symptom in GERD children was burping/belching (68%) and the least prevalent was difficulty swallowing (35%) (Fig. 1b). GERD children had significantly more frequent symptoms ($P < 0.05$) than controls for all symptoms evaluated. ''Burping/belching'' was the symptom with highest frequency, more than twice the frequency reported in controls. Abdominal/belly pain was the symptom with lowest frequency in both groups but the frequency was still twice as prevalent ($P = 0.001$) in GERD than in controls (Fig. 2b). Additional symptoms reported manually in GERD children included ''wakes up in the middle of the night crying,'' ''night time apnea-crying episodes,'' ''screaming/inconsolable after naps'' (3 each), ''belly-ache/stomach pain,'' ''hiccups,'' ''refusal to swallow,'' ''coughing,'' ''irritability'' and ''irritability/crying when eating'' (2 each).

GERD children had significantly more severe symptoms than controls for two of the six symptoms (burping/belching and vomiting/regurgitation; $P = 0.004$ and

TABLE 3. *Demographic and baseline characteristics: young children*

	Healthy	GERD	Total
	(n = 27)	(n = 40)	(n = 67)
Age (months)			
Mean + SD	31.3 ± 14.8	30 ± 12.1	30.5 ± 13.1
Median (range)	28.3 (12.2–59.3)	29.2 (13–59.7)	29.0 (12.2–59.7)
Sex, n (%)			
Boys	14 (52)	24 (60)	38 (57)
Girls	13 (48)	16 (40)	29 (43)
Ethnic Origin, n (%)			
Asian/Asian American	3 (11)	1 (3)	4 (6)
Black	8 (30)	8 (20)	16 (24)
Hispanic	2 (7)	1 (3)	3 (4)
White	14 (52)	30 (75)	44 (67)
Method of Diagnosis, n (%)			
Symptoms	N/A	37 (92.5)	
Endoscopy +/− histology	N/A	15 (37.5)	
Upper GI*	N/A	9 (22.5)	
pH	N/A	4 (10)	
Radionuclide study	N/A	1 (2.2)	

GI, gastrointestinal study.
*Including barium swallow and contrast radiography.
Not all percentages add up to 100% because some patients were diagnosed by more than one criterion.

$P = 0.010$, respectively) (Fig. 1b). In GERD children, five of the six symptoms had a median severity score ≥ 4 on the 7-point severity scale versus controls whose highest median rating was 3 for the same symptoms.

The mean ISS and CSS were statistically higher for GERD children than controls for every symptom (Table 4). Burping/belching (ISS = 19.6), choking when eating (ISS = 18.4), refuses to eat (ISS = 19.4) and vomiting/regurgitation (ISS = 18.0) each contributed approximately 20% to the overall CSS. There was no significant difference between boys and girls in number of episodes (Fig. 4), CSS ($P = 0.79$) or ISS ($P \geq 0.26$).

Subgroup analysis of symptom frequency, ISS, CSS comparing data only from GERD children who were diagnosed by objective tests (n = 21) with all age-matched controls showed a general increase in the differences between GERD and control which did not reach the same significance level because of the smaller sample size.

We found that a CSS > 8 had a sensitivity of 85% for detecting children with GERD (34 of 40 GERD children had scores >8) and a specificity of 81.5% (22 of 27 controls had scores ≤ 8).

DISCUSSION

This study supports the appropriateness and relevance of the symptom set employed on each questionnaire for the differentiation of GERD subjects from healthy subjects of the same age group. All of the symptoms were more prevalent in GERD children than in their healthy age-matched controls, and all but one of the symptoms were more prevalent in GERD infants than in healthy infants. In both age cohorts, all symptoms were more frequent in GERD subjects than in controls. It is of interest that although the majority of GERD subjects were receiving pharmacotherapy, the instruments were still able to distinguish these patients from controls.

The frequency and severity scales in this instrument suggest no obvious ceiling or floor effects in patients with GERD symptoms. Therefore, the range of measurement of these scales appears to be adequate and remained unchanged in the final versions. In particular, our decision not to restrict the responses with a maximum score on the frequency scale allowed us to accommodate the wide range in both groups, and suggests that no further revision of this scale is needed.

Supplemental information supplied by parents/guardians indicated that the questionnaires sufficiently captured the pertinent symptoms of GERD in both age groups. Although additional symptoms were recorded in the ''other'' category in both age cohorts, their prevalence was not significant, they were similar to formally listed symptoms, or they were not typical of the usual presentation of GERD. To assist parents in relating their observations to the formally listed symptoms and to avoid confusion over the definition of symptoms occurring in episodes such as belching/burping, additional time was needed with the parents/guardians to review the symptom definitions and explain the use of the word ''episodes'' versus individual events. The ''other'' category will remain in the questionnaire until the symptom sets have been validated in a clinical trial.

The significantly higher mean ISS and CSS for both age cohorts with GERD compared to age matched controls provides evidence that the questionnaires were useful in differentiating infants and young children with GERD from healthy subjects of the same age. Initial CSS threshold values ≥ 27 for infants and ≥ 8 for young children are suggested from this pilot study.

The results of this pilot study of the GSQ-I and GSQ-YC questionnaires for infants and young children with GERD suggest that 1) the selected symptoms for each questionnaire were appropriate and relevant; 2) the measurement range for frequency and severity scales was adequate for GERD evaluation; 3) the questionnaires were easy to use with extra time needed to explain the meaning of ''episodes''; and 4) the CSS of the two age cohorts were higher for clinically diagnosed GERD subjects than for controls.

One major limitation of this pilot study is that the GSQ-I and GSQ-YC, intended as a measure of outcome in clinical trials, were assessed on their ability to differentiate between a sample of clinically diagnosed GERD subjects and controls who were not clinical trial participants. It is difficult to know if there are fundamental differences between the participants in this study and those who might participate in a clinical trial. Use of the questionnaires in a clinical trial setting will provide the data necessary to answer this question. In addition, although many of the GERD subjects had additional diagnostic tests to confirm the clinical diagnosis of GERD made by experienced pediatric gastroenterologists, this pilot study did not require such tests for inclusion.

The choice of symptoms in the questionnaires was appropriate to distinguish GERD patients from controls because of the significant differences in prevalence, frequency, and severity of the symptoms. In a clinical trial with only GERD subjects, however, the ability of the symptom frequency and severity scores to reflect response to treatment is an important issue untested in this study.

It was difficult to fully assess the ability of the severity scales to distinguish between GERD patients and controls because of the low prevalence of symptoms in controls of both ages. However, we did not see a floor or ceiling effect in GERD subjects, suggesting that the severity scale is appropriate for assessing GERD. In addition, the lack of symptoms in the control groups further supports the symptom selection as appropriate for GERD patients.

Although the results of this pilot study did not lead to any changes in the questionnaires themselves, additional investigation using longitudinal data from an actual

clinical trial is necessary to establish the validity of these instruments, including test-retest reliability, inter-rater reliability, concurrent validity, and internal consistency in the clinical trial setting.

The GSQ-I and GSQ-YC questionnaires are easy to use and clearly differentiate patients with GERD from control infants and young children. The questionnaires may be useful in clinical trials to assess symptom response to treatment. We emphasize that our purpose was to validate the two questionnaires for use in clinical trials. In light of this objective, it was not appropriate to subject participants to invasive procedures. Despite this decision, a subset analysis comparing only patients diagnosed using objective criteria (endoscopy, pH metry, upper GL, oxypneumogram, or radionuclide study) with controls demonstrated results similar to the full group comparisons. Prospective studies are recommended to establish the criterion validity of the questionnaires as diagnostic tools by comparisons with objective diagnostic tests such as endoscopy and biopsy.

Acknowledgment: The authors express their gratitude to Arthur R. Euler, MD for his assistance in refining the original list of symptoms and to Gail M. Comer, MD of Wyeth Research for assistance with data analysis and manuscript preparation. Ms Maria Vinall of Wyeth Research assisted in writing this manuscript.

REFERENCES

1. Rudolph CD, Mazur LJ, Liptak GS, et al. Guidelines for evaluation and treatment of gastroesophageal reflux in infants and children: recommendations of the North American Society for Pediatric Gastroenterology and Nutrition. *J Pediatr Gastroenterol Nutr* 2001; 32:S1–S31.
2. Colletti RB, Di Lorenzo C. Overview of pediatric gastroesophageal reflux disease and proton pump inhibitor therapy. *J Pediatr Gastroenterol Nutr* 2003;37:S7–11.
3. Gold BD. Outcomes of pediatric gastroesophageal reflux disease: In the first year of life, in childhood, and in adults...oh, and should we really leave *Helicobacter pylori* alone? *J Pediatr Gastroenterol Nutr* 2003;37:S33–9.
4. Nelson SP, Chen EH, Syniar GM, Christoffel KK. Prevalence of symptoms of gastroesophageal reflux during infancy. A pediatric practice-based survey. *Arch Pediatr Adolesc Med* 1997;151:569–72.
5. Nelson SP, Chen EH, Syniar GM, Christoffel KK. One-year follow-up of symptoms of gastroesophageal reflux during infancy. Pediatric Practice Research Group. *Pediatrics* 1998;102:E67.
6. Nelson SP, Chen EH, Syniar GM, Christoffel KK. Prevalence of symptoms of gastroesophageal reflux during childhood: a pediatric practice-based survey. Pediatric Practice Research Group. *Arch Pediatr Adolesc Med* 2000;154:150–4.
7. Orenstein S. Tests to assess symptoms of gastroesophageal reflux in infants and children. *J Pediatr Gastroenterol Nutr* 2003;37:S29–32.
8. Badriul H, Vandenplas Y. Gastro-esophageal reflux in infancy. *J Gastroenterol Hepatol* 1999;14:13–9.
9. Gold BD. Review article: epidemiology and management of gastro-esophageal reflux in children. *Aliment Pharmacol Ther* 2004;19 (Suppl 1):22–7.
10. Orenstein SR, Cohn JF, Shalaby TM, Kartan R. Reliability and validity of an infant gastroesophageal reflux questionnaire. *Clin Pediatr* 1993;32:472–84.
11. Orenstein SR, Shalaby TM, Cohn JF. Reflux symptoms in 100 normal infants: diagnostic validity of the infant gastroesophageal reflux questionnaire. *Clin Pediatr* 1996;35:607–14.
12. Rothman M, Orenstein S, Kleinman L, et al. *Development of a revised infant gastroesophageal reflux questionnaire. Presented at the Second European Pediatric Gastrointestinal Motility Meeting, 2003, Bruges, Belgium. Acta Gastroenterologica Belgica* 2003:73.
13. Rothman M, Farup C, Stewart W, Helbers L, Zeldis J. Symptoms associated with gastroesophageal reflux disease: development of a questionnaire for use in clinical trials. *Dig Dis Sci* 2001;46:1540–9.
14. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. 2nd ed. Oxford, New York: Oxford University Press; 1995.
15. Ware JE Jr. Standards for validating health measures: definition and content. *J Chronic Dis* 1987;40:473–80.