INVITED COMMENTARY

Symptoms and Reflux in Infants: Infant Gastroesophageal Reflux Questionnaire Revised (I-GERQ-R)—Utility for Symptom Tracking and Diagnosis

Susan R. Orenstein

Abstract Answering a need for a thoroughly validated infant gastroesophageal reflux questionnaire, the Infant Gastroesophageal Reflux Questionnaire Revised (I-GERQ-R) was designed, refined, and validated using state-of-the-art psychometric methods. Diagnostic and evaluative (tracking) validity was identified. However, perplexing results of some clinical trials using the I-GERQ-R for diagnosis prompted analysis of possible reasons, including ambiguities in defining symptomatic gastroesophageal reflux disease (GERD) and aspects of the validation process. Symptomatic GERD is defined by "troublesomeness" of symptoms and attribution of their causation to reflux—two crucial issues. Methods of quantifying symptom-reflux associations are described and their limitations identified. The location of "symptomatic esophageal GERD" in the continuum of erosive GERD, histologic GERD, and nonerosive reflux disease is indicated, with the last including "suberosive," "premicroscopic," and "functional heartburn" subcategories. Another category is defined solely by surrogate measures of propensity to GERD (e.g., acid exposure thresholds defined on esophageal pH monitoring). During diagnostic validation of the Infant Gastroesophageal Reflux Questionnaire (I-GERQ) instruments, asymptomatic normals were contrasted with symptomatic GERD infants (who also tested positive with esophageal histology and esophageal pH monitoring). However, the diagnostic validation did not attempt to distinguish symptomatic GERD infants from symptomatic infants without GERD. The I-GERQ-R is thus adequately sensitive to be used diagnostically to screen infants for symptom burden, but should probably be supplemented by other, perhaps invasive, testing to assure appropriate specificity. The I-GERQ-R's validation for evaluative properties, however, supports its use for tracking symptoms within clinical trials.

Keywords Gastroesophageal reflux · Reflux · Infant · Questionnaire · I-GERQ · I-GERQ-R · Validity · Diagnostic validity · Evaluative validity · Symptoms · Symptom Index · Symptom Sensitivity Index · Symptom Association Probability · Erosive · Morphometric · Histology · Nonerosive reflux disease · Clinical trial

Introduction

The exponential growth in prescription of medications for gastroesophageal reflux disease (GERD) in infants [1, 2], combined with the 1997 Food and Drug Administration (FDA) Modernization Act and the 2002 Best Pharmaceuticals for Children Act (http://www.fda.gov/RegulatoryInformation), which encouraged pediatric-specific clinical trials, necessitated development and validation of infant-specific diagnostic and tracking instruments. Since 2006, explicit FDA emphasis on patient-reported outcomes (http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm) heightened the desirability of using noninvasive outcome measures, such as symptom questionnaires, which are particularly advantageous in infants because of infants' greater vulnerability to invasive measures. Infants' verbal immaturity mandates completion of such a symptom questionnaire by a surrogate—a caretaker, usually a parent—who reports the infant's symptoms as they have observed them [3].

S. R. Orenstein (✉)
Pediatric Gastroenterology,
University of Pittsburgh School of Medicine,
303 Church Lane,
Pittsburgh, PA 15238, USA
e-mail: sro.pitt.edu@verizon.net

Infant Gastroesophageal Reflux Questionnaire Revised

To date, the most thoroughly evaluated such questionnaire for infant symptoms is the Infant Gastroesophageal Reflux Questionnaire Revised (I-GERQ-R; Author I-GERQ, Susan Orenstein, MD, copyright 2004, University of Pittsburgh) [4].

The development of the I-GERQ-R began with the original Infant Gastroesophageal Reflux Questionnaire (I-GERQ, Susan Orenstein, MD, copyright 1992, 2002, University of Pittsburgh), a large, 161-item questionnaire designed to improve history taking in infants with suspected GERD by assessing such factors as demographics, symptoms, aspects of the differential diagnosis, and treatable provocative factors. We first published the results of testing of this questionnaire's various types of reliability (test-retest consistency, interobserver consistency, internal consistency, and accuracy) in 69 infants suspected of GERD, and found acceptable reliability for the tested items [5].

Next, interested in determining whether a diagnostic score based on items from the I-GERQ could discriminate infants with GERD from those without, we administered a slightly shorter (138-item) version of the original I-GERQ to 35 GERD infants (defined by gold standards of abnormal esophageal histology or esophageal pH monitoring [EpHM]) and 100 normal infants attending a well-baby clinic [6]. We analyzed differences in responses to the questionnaire items between the two groups via chi-square, identifying 11 items with the highest odds ratios for differentiating GERD from normal infants and constructing a 25-point "I-GERQ Score" from those items. Although normal infants had a high frequency of daily regurgitation (40%) and crying more than an hour a day (17%), nearly 20 items manifested odds ratios of greater than 3, and a cut-point score of 7 (of 25) had a high sensitivity and specificity for differentiating GERD infants from normal infants on the I-GERQ Score.

This I-GERQ Score was further refined into the I-GERQ-R score [4] by an industry-sponsored study, which undertook a large-scale refinement of the items (submitting them to focus groups of parents and experts), resulting in slight differences in wording of some items, and translating them into multiple languages. The I-GERQ-R was then validated as a tracking (evaluative) instrument for use in clinical trials in 185 GERD and 93 non-GERD infants in 16 sites in seven countries. This short questionnaire was also re-validated for diagnostic purposes, using a less-objective gold standard of "clinician diagnosis" than was originally used during the prior validation of the original I-GERQ Score, but confirming the diagnostic validity.

The I-GERQ-R (and unauthorized, unvalidated variants) subsequently was used in several published studies, and it has been widely identified as the most thoroughly validated infant reflux questionnaire in existence. Because it is

copyrighted, licensing from the University of Pittsburgh is required for its use in clinical trials (contact: Carolyn Weber, Technology Marketing Manager, cjweber@otm.tt. pitt.edu, 412-383-7140). By April 2010, 31 academic licenses (from 12 US states and 17 other countries) and five industry licenses had been established with the University of Pittsburgh. However, results of some studies using the questionnaire have been perplexing.

Perplexing Studies Using I-GERQs

An I-GERQ-R–based clinical trial of lansoprazole (comparing two lansoprazole dosing regimens versus a control treatment with hydrolyzed formula alone) showed expected reductions of symptoms in 2 weeks for the proton pump inhibitor (PPI)–treated infants (60% and 67%) contrasted with significantly less improvement in the control group (20%) [7]. However, several other I-GERQ-R–based studies have not shown expected results.

In particular, our large study of 162 symptomatic infants randomly assigned to 4 weeks of lansoprazole or placebo produced identical (54% in each group) proportions of responders to drug and placebo [8]. Although the I-GERQ-R was not used as the primary screen for enrollment, nor as the primary outcome measure, a variant of it was used as a confirmatory diagnostic screen and secondary outcome variable. Other clinical trials of PPIs in infants categorized as having GERD on the basis of symptoms (but not specifically on the basis of the I-GERQs) have been similarly perplexing [9–11].

A crossover clinical trial using omeprazole versus placebo for 2-week treatment periods in crying infants who also had abnormal EpHM (reflux index>5%) and/or abnormal histology resulted in similar decreases in the crying symptom from baseline regardless of the treatment assigned. However, the reflux index decreased significantly more during omeprazole treatment, suggesting that the crying symptom was not solely determined by the esophageal acid exposure, nor specifically treated by suppression of gastric acid secretion [11].

A prospective descriptive study tested 100 infants suspected of having GERD with an unauthorized variant of the I-GERQ Score (but with 30% of the questionnaires containing missing answers) and with EpHM (considering a reflux index>10% as abnormal) [12]. A minority of the infants also underwent esophageal endoscopy and biopsy. The authors found the score they used was unable to predict EpHM (or histologic) abnormalities, but missing data and unvalidated procedures for the histology and the questionnaire make their results difficult to assess.

The PPI drugs used in some of these clinical trials have been shown in rigorous pharmacodynamic studies to

decrease esophageal acid exposure in infants, as well as in older individuals. Why, then, should a questionnaire developed using acid exposure–related objective gold standards for diagnostic validation (EpHM and reflux-associated changes in esophageal histology) fail to show any greater improvement than that produced by placebo during treatment with these drugs?

Definition of Symptomatic GERD

The answer to this conundrum is to be found in the ambiguities of the definition of "symptomatic GERD," particularly in nonverbal infants, and in the limitations of our diagnostic validation processes.

An expert panel recently convened to define GERD in pediatrics asserted that "GERD is present when reflux of gastric contents causes troublesome symptoms and/or complications" [13]. As a member of that panel, I was frustrated by the tautological, circular, and ambiguous nature of the definition: what is "troublesome," and to whom is it troublesome in an infant? In older children, is an episode of heartburn once weekly truly troublesome, and classifiable as "disease"? Is heartburn that is completely obviated by avoiding smoking and overindulgence in alcohol or calories reasonably considered disease? The problem of the degree of "troublesomeness" is not limited to infants: patients of all ages have widely differing tolerances for the vicissitudes of ordinary life. The problem of defining troublesomeness in nonverbal infants is a further complexity: is quantification of crying a reasonable measure of troublesomeness, or should parents define the troublesome nature of, for example, the mess made by regurgitant reflux?

A second problem with the definition relates to causation. How does one assure an accurate attribution of symptoms as being "due to" GERD? Such causality may be indicated by resolution of the symptom during treatment directed at GERD, by close temporal relationship between the symptom and preceding individual reflux episodes, or by epidemiologic associations between the symptom and GERD diagnosis, but problems intrinsic to these methods are circularity, invasiveness, and nonapplicability to individual patients. Resolution of symptoms spontaneously or because of maturation also confounds attribution to therapeutic response, as evidenced by the impressive improvement of infant symptoms during placebo treatment [8].

Inferring that reflux caused a symptom on the basis of symptom resolution during antireflux treatment would be circular in the context of treatment trials, and would suggest causation in studies not using placebo controls even when maturation or placebo response was actually responsible.

Defining causality by documenting temporal linkage of reflux episodes (during EpHM) to specific symptoms also

has several challenges. These include defining the appropriate reflux-symptom *direction* (some studies have not required reflux to precede the symptom [14]) and *interval duration* for "association" (published studies use intervals ranging from 15 sec to 5 min [14, 15]); choosing the method of symptom recording (video, observer-activated key, or handwritten log); and training of the person logging/coding the symptom (who in some studies was simply the parent during an entire 24-hour recording period). The need for 24-hour EpHM recording makes this method somewhat invasive and time-consuming to administer and analyze. Once these 24-hour EpHM studies are complete, several methods have been used to quantify the relationship between reflux episodes and symptoms. The Symptom Index (SI) identifies the percentage of *symptom events* that are temporally related to a reflux event, whereas the Symptom Sensitivity Index (SSI) identifies the percentage of *reflux events* that are temporally related to a symptom [14]. Neither measure takes into consideration the *nonassociated* events, so that, as examples, an infant who refluxes during 95% of study time would have a very high SI for any symptom that occurred, regardless of any causal relationship, and in contrast, an infant who cried during 95% of the study time would have a very high SSI for crying during any reflux that occurred. A third index, the Symptom Association Probability (SAP), attempts to compensate for these shortcomings, by accounting for the nonreflux and nonsymptom periods [14]. When we used a specified direction of association (reflux onset before symptom), 15-second intervals, video recording linked directly to EpHM, and a single trained coder, we found six common behaviors to be closely linked temporally ($P<0.001$ to <0.05) with the onset of reflux events: crying/frowning, regurgitation/belching, yawning, stridor, stretching, and mouthing [15]. Several other less frequent behaviors (hiccupping, sneezing, thumb-sucking, coughing/gagging) were temporally associated with onset of reflux episodes in some individual patients. Although all 10 behaviors followed reflux onsets, some (crying, regurgitation, mouthing, and cough) also preceded reflux and thus may have caused some reflux episodes. In all cases, the symptoms sometimes occurred completely independent of reflux episodes.

Using epidemiologic associations between abnormal amounts of reflux and symptoms to define causality are most useful for chronic symptoms such as asthma, and are beyond the scope of this review.

GERD Definition Framework: Symptomatic, Histologic, Endoscopic, and Surrogate

The definition of *symptomatic* esophageal GERD, just discussed, is the broadest and most ambiguous. On the other

hand, the narrowest and most rigorous definition of esophageal GERD defines it by the presence of endoscopic esophageal erosions. Once other causes of erosions (e.g., infections, medications) have been excluded—which is generally fairly simple—erosive esophagitis is clearly gastroesophageal reflux *disease.* Thus erosive esophagitis is very specific, but very insensitive, in defining GERD in infants.
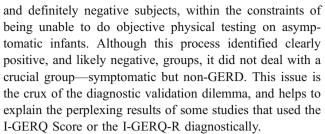
An intermediate characterization of esophageal GERD does not require visible erosions, but relies on changes in histologic morphometric parameters that are associated with excessive acid reflux [16, 17]. This more sensitive measure of pathologic changes in the epithelium resulting from acid exposure depends on adequate size and appropriate orientation of the biopsy specimens, as well as on accurate measurements of papillary height and basal cell thickness, aspects probably rarely assured in infant studies incorporating histology. Dilated intercellular spaces are a smaller scale histopathologic change that has been related to reflux, even in infants [18].

Nonerosive reflux disease (NERD) includes not only the "suberosive" histologic changes just described, but also "premicroscopic" epithelial injury that is not even detectable with light microscopy, and "functional heartburn" symptoms that may be associated with reflux without evident esophageal damage [19]. Such NERD is often identified by the associations between reflux episodes (on EpHM) and symptoms discussed above, but it is often challenging to make these associations with much confidence in individual patients, particularly when the symptoms are so common in normal individuals.

A fourth method of defining GERD is based on surrogates for esophageal damage or pain. This often-used category includes techniques such as using EpHM to quantify the reflux index (total percent time with esophageal intraluminal pH less than 4). The reflux index itself is not pathologic, but only on the basis of associations between particular levels of reflux index and esophageal damage or pain. That is, the damage or pain that ensues depends not only on the duration of exposure of the esophagus to a pH less than 4 during an individual 24-hour period, but also on factors such as the days and months that it has been so exposed, the absolute level of the pH, the esophageal mucosal defense mechanisms, and susceptibility to perception of pain. Although particular reflux index thresholds have been associated with esophageal damage or pain, these associations are imperfect and limited, so that the results of pH probe monitoring can only be taken as suggestive of the possibility of GERD, not as identifying GERD itself.

I-GERQ Diagnostic Validation Issues

During the initial diagnostic validation of the I-GERQ Score [6], we sought to define definitely positive subjects and definitely negative subjects, within the constraints of being unable to do objective physical testing on asymptomatic infants. Although this process identified clearly positive, and likely negative, groups, it did not deal with a crucial group—symptomatic but non-GERD. This issue is the crux of the diagnostic validation dilemma, and helps to explain the perplexing results of some studies that used the I-GERQ Score or the I-GERQ-R diagnostically.

To understand the potential magnitude of this group of confounding subjects, one can examine the epidemiology of infants recruited and screened for a large clinical trial [20]. Infants who are referred because of concerning symptoms seldom have erosions and rarely undergo endoscopy. The proportion of referred symptomatic infants (eliminating those with such exclusion criteria as structural lesions, prior surgery, hematemesis, and apnea) who responded to 2 weeks of nonpharmacologic treatment prior to enrollment was about 25% (96/394) [21]. Of the remaining 298 infants, 208 had esophageal suction biopsies following parental consent, with 180 of them (87%) having abnormal morphometric parameters [22]. Therefore, between 13% and 40% ([298 to 180]/298) of referred symptomatic infants who did not respond to nonpharmacologic treatment might have been biopsy-negative, with an unknown proportion of the biopsy-negative infants also capable of having normal EpHM results. Thus, up to 40% of referred symptomatic infants not responding to nonpharmacologic therapy might have symptoms not due to GERD. This group of infants was not studied during the I-GERQ Score validation process, and this group may constitute a fair proportion of the enrollees in the perplexing studies that used the I-GERQ Score, the I-GERQ-R, or variants as diagnostic criteria for clinical trials of treatments for symptomatic GERD in infants. These symptomatic but non-GERD infants would not be expected to respond to pharmacotherapy for GERD any better than to placebo, and thus would confound any studies enrolling them.

A related consideration is that even symptoms related to reflux in infants may not be from esophageal acid exposure, but may be caused by gastric distension or neutral-pH reflux, and thus would not respond to acid-suppressing medications [23]. We probably need terms to designate infants whose symptoms are caused by nonacid reflux (e.g., symptoms of excessive regurgitation or respiratory manifestations), and those whose symptoms and reflux are co-determined (e.g., those with gastric distension causing both crying and reflux).

All of these issues mean that, although the I-GERQ instruments can provide effective screens for symptom burden, they do not adequately exclude symptomatic infants whose symptoms are not caused by acid reflux. Because the diagnostic validity of the I-GERQ instruments is limited in this way, additional means (e.g., EpHM, biopsy) to cull the

non-GERD infants are needed in clinical trials of GERD medications, and probably in clinical practice, unless future clinical trials indicate that treatment trials are a feasible means to "diagnose" such symptomatic GERD infants.

Evaluative Validity for Tracking Symptoms

In contrast, the evaluative validity of the I-GERQ-R is well established, indicating its ability to follow symptoms effectively during treatment. Responsiveness and effect size analyses confirmed that the I-GERQ-R can detect clinically meaningful changes over time, regardless of whether patients are defined as responders by parents or physicians ($P<0.0001$) [4]. Score reductions between 5 and 6 points can be considered clinically meaningful, and differences of 3 points might represent a minimally important difference. Analyses showed that "on the basis of a 3-point minimally important difference, a clinical trial with the I-GERQ-R total score as the primary end point would need 84 subjects per treatment group (assuming SD=6.9; power=0.80; $P=0.05$)" [4].

Conclusions

In summary, the I-GERQ-R is a useful instrument with robust evaluative properties for tracking symptoms during clinical trials. For initial diagnosis and for inclusion criteria in such trials of symptomatic GERD in infants, the I-GERQ-R can define, with fair sensitivity, a symptom burden threshold necessary for inclusion. However, the I-GERQ-R has not been validated to be adequately specific to differentiate GERD infants from other symptomatic, but non-GERD, infants, particularly if the treatments being tested are directed at acid reflux. For that purpose, additional inclusion criteria, such as EpHM or histology, are currently necessary to assure adequate diagnostic specificity.

References

1. Barron JJ, Tan H, Spalding J, et al.: Proton pump inhibitor utilization patterns in infants. J Pediatr Gastroenterol Nutr 2007, 45:421–427.

2. Orenstein SR, Hassall E: Infants and proton pump inhibitors: tribulations, no trials [invited editorial]. J Pediatr Gastroenterol Nutr 2007, 45:395–398.

3. Kleinman L, Revicki DA, Flood E: Validation issues in questionnaires for diagnosis and monitoring of gastroesophageal reflux disease in children. Curr Gastroenterol Rep 2006, 8:230–236.

4. Kleinman L, Rothman M, Strauss R, et al.: The Infant Gastroesophageal Reflux Questionnaire Revised: development and validation as an evaluative instrument. Clin Gastro Hep 2006, 4:588–596.

5. Orenstein SR, Cohn JF, Shalaby TM, Kartan R: Reliability and validity of an infant gastroesophageal reflux questionnaire. Clin Pediatr 1993, 32:472–484.

6. Orenstein SR, Shalaby TM, Cohn JF: Reflux symptoms in 100 normal infants: diagnostic validity of the Infant Gastroesophageal Reflux Questionnaire. Clin Pediatr 1996, 35:607–614.

7. Khoshoo V, Dhume P: Clinical response to 2 dosing regimens of lansoprazole in infants with gastroesophageal reflux. J Pediatr Gastroenterol Nutr 2008, 46:352–354.

8. Orenstein SR, Hassall E, Furmaga-Jablonska W, et al.: Multicenter, double-blind, randomized, placebo-controlled trial assessing efficacy and safety of proton pump inhibitor lansoprazole in infants with symptoms of gastroesophageal reflux disease. J Pediatr 2009, 154:514–520.

9. Winter H, Kum-Nji P, Mahomedy SH, et al.: Efficacy and safety of pantoprazole delayed-release granules for oral suspension in a placebo-controlled treatment-withdrawal study in infants 1–11 months old with symptomatic GERD. J Pediatr Gastroenterol Nutr 2010, 50:609–618.

10. Orenstein SR, Hassall E: Pantoprazole for symptoms of infant GERD: the emperor has no clothes! J Pediatr Gastroenterol Nutr, in press.

11. Moore DJ, Tao BS-K, Lines DR, et al.: Double-blind placebo-controlled trial of omeprazole in irritable infants with gastroesophageal reflux. J Pediatr 2003, 143:219–223.

12. Salvatore S, Hauser B, Vandemaele K, et al.: Gastroesophageal reflux disease in infants: How much is predictable with questionnaires, pH-metry, endoscopy and histology? J Pediatr Gastroenterol Nutr 2005, 40:210–215.

13. Sherman PM, Hassall E, Fagundes-Neto U, et al.: A global, evidence-based consensus on the definition of gastroesophageal reflux disease in the pediatric population. Am J Gastroenterol 2009, 104:1278–1295.

14. Luthold SC, Rochat MK, Bahler P: Disagreement between symptom-reflux association parameters in gastroesophageal reflux disease investigation. World J Gastroenterol 2010, 16:2401–2406.

15. Feranchak AP, Orenstein SR, Cohn JF: Behaviors associated with onset of gastroesophageal reflux episodes in infants: prospective study using split-screen video and pH probe. Clin Pediatr 1994, 33:654–662.

16. Black DD, Haggitt RC, Orenstein SR, Whitington PF: Esophagitis in infants: morphometric histologic diagnosis and correlation with measures of gastroesophageal reflux. Gastroenterology 1990, 98:1408–1414.

17. Sabri MT, Hussain SZ, Shalaby TM, Orenstein SR: Morphometric histology for infant gastroesophageal reflux disease: evaluation of reliability in 497 esophageal biopsies. J Pediatr Gastroenterol Nutr 2007, 44:27–34.

18. Ravelli AM, Villanacci V, Ruzzenenti N, et al.: Dilated Intercellular spaces: a major morphological feature of esophagitis. J Pediatr Gastroenterol Nutr 2006, 42:510–515.

19. Orlando RC: Current understanding of the mechanisms of gastrooesophageal reflux. Drugs 2006, 66(Suppl 1):1–5.

20. Orenstein SR, Shalaby TM, Frankel EA, Kelsey SF: Cimetidine, cisapride, both, or neither for infantile esophagitis: symptomatic

response to 6-months randomized, double-blind, placebo-controlled therapy in 100 babies. Gastroenterology 2004, 126(Suppl 2):239.

21. Shalaby TM, Orenstein SR: Efficacy of telephone teaching of conservative therapy for infants with symptomatic gastroesophageal reflux referred by pediatricians to pediatric gastroenterologists. J Pediatr 2003, 142:57–61.

22. Orenstein SR, Shalaby TM, Kelsey SF, Frankel E: Natural history of infant reflux esophagitis: symptoms and morphometric histology during one year without pharmacotherapy. Am J Gastroenterol 2006, 101:628–640.

23. Orenstein SR: Crying in infant GERD: acid or volume? Heartburn or dyspepsia? Cur Gastroenterol Rep 2008, 10:433–436.