

Diagnostic utility of reflux disease symptoms

M Shaw

Gut 2004;53(Suppl IV):iv25–iv27. doi: 10.1136/gut.2003.034280

Symptom assessment, by structured interview or questionnaire, is central to the diagnosis of gastro-oesophageal reflux disease (GORD). However, empiric support for the diagnostic utility of reflux symptom measures is lacking. Reliable reflux symptom questionnaires have been developed with content validity. These questionnaires need to be evaluated in terms of diagnostic accuracy, support for application of specific treatment, and improved outcomes resulting from their use. The impact on clinical outcomes of GORD diagnosis by valid questionnaires or structured interview has not been studied.

GORD, will inevitably show strong support for selection of specific treatments. Although of use to the practitioner, this would not be applicable for most clinical trials.

The impact on clinical outcomes of GORD diagnosis by valid questionnaires or structured interview has not been studied.

INTRODUCTION

Symptoms play a central role in the diagnosis of disease, especially gastro-oesophageal reflux disease (GORD), given the limitations of objective medical testing. Symptom assessment can be achieved by structured interview or questionnaire. Although a structured interview more closely approximates clinical practice, questionnaires possess a number of inherent advantages, especially if automated.¹ Questionnaires minimise interobserver variability and facilitate quantitative assessment of subject responses. Easily scored questionnaires also lend themselves to efficient, inexpensive symptom assessment. Research on the diagnostic utility of reflux disease symptoms will most likely use questionnaires in the future given these and other advantages.

A useful diagnostic test needs to satisfy three criteria: (1) provide an accurate diagnosis; (2) support application of specific, efficacious treatment; (3) lead to improved outcomes.² Satisfaction of these criteria by the systematic gathering of reflux symptom data will be examined.

ACCURACY OF A SYMPTOM DIAGNOSIS

The accuracy of symptom diagnosis has been examined with interviews³ and questionnaires^{4,5} in GORD enriched populations. Increasing the prevalence of affected individuals improves the positive predictive value of symptom assessment, but impairs the negative predictive value. The diagnostic utility of reflux disease symptoms should ideally be assessed in an unselected population of symptomatic individuals. In some cases, however, it may be possible to extrapolate the data from selected populations to estimate the accuracy of symptom assessment in an unselected primary care population (see Moayyedi and colleagues⁶ in this supplement (*page iv55–iv57*)).

Interviews by experienced gastroenterologists were used in a study using ambulatory pH monitoring as the gold standard.³ The interviewers categorised the patients as having equivocal or unequivocal histories of GORD.

Abbreviations: GORD, gastro-oesophageal reflux disease; OTC, over the counter; RDQ, reflux disease questionnaire

SUMMARY

Symptom assessment, by structured interview or questionnaire, is central to the diagnosis of gastro-oesophageal reflux disease (GORD). To be of use, it needs to provide an accurate diagnosis, support specific treatment selection and lead to improved outcomes.

Interviewer and questionnaire methods have shown relatively high sensitivities of 70% or more for the diagnosis of GORD, but low specificity. However, a gold standard comparative measure is lacking, and comparison—for example with pH monitoring—risks underestimating the diagnostic accuracy of symptom assessment. Use of a word picture to describe heartburn enhances diagnostic accuracy, and the same may apply for regurgitation and epigastric pain, although this requires detailed scrutiny. Rigorous development and validation of questionnaires and interviewer methods is underway to enhance diagnostic accuracy. The Reflux Disease Questionnaire is a 12 item questionnaire of three scales (heartburn, regurgitation, and dyspepsia) that is reliable and valid, although diagnostic validity remains to be established. Selection of the predominant symptom has been suggested to improve diagnostic accuracy, but is limited by the inability of most patients to identify a predominant symptom when asked to do so directly.

Selection of the predominant symptom heartburn has been shown to predict response to proton pump inhibitor therapy. Also, endorsement of the individual questionnaire items “burning rising feeling” or “relief from antacids” identifies those more likely to respond. Use of questionnaires, coupled with a trial of treatment in patients thus diagnosed as likely to have

Correspondence to:
Dr M Shaw, Park Nicollet
Institute, 3800 Park
Nicollet Boulevard,
Minneapolis, MN 55416-
2699, USA; shawmj@
parknicollet.com

Table 1 Diagnostic accuracy of reflux symptoms³

| | Sensitivity (%) | Specificity (%) | Positive predictive value (%) |
|---------------------|-----------------|-----------------|-------------------------------|
| Unequivocal history | | | |
| Heartburn | 73 | 53 | 70 |
| Regurgitation | 66 | 58 | 70 |
| Equivocal history | | | |
| Heartburn | 52 | 52 | 46 |
| Regurgitation | 46 | 45 | 40 |

This selection process had an impact on the diagnostic utility with positive predictive values for the symptoms of heartburn and regurgitation of 70% each in the unequivocal group and 46% for heartburn and 40% for regurgitation in the equivocal group (table 1).

The first questionnaire diagnosis study also used a gold standard for diagnosis, of ambulatory pH monitoring and oesophagogastroduodenoscopy (OGD).⁴ The questionnaire consisted of four items with word pictures for the quality, location, and movement of the discomfort along with symptom frequency and response to antacids. Response options were dichotomised and positive responses were required to all four questions for a diagnosis of GORD. Only 33% of subjects referred with presumed GORD had a positive questionnaire diagnosis with these stringent criteria and the positive predictive value was 85%.

This was followed by another study using a different questionnaire, but the same gold standards for diagnosis.⁵ The word picture described above was incorporated along with other questions that either supported or detracted from a diagnosis of GORD. Final questionnaire scoring used item weighting assigned by the investigators. A cut off for diagnosis relied on computer modelling of different options. Using the computer derived cut off of 4, a sensitivity of 70% and specificity of 46% was determined in subjects referred for presumed GORD. The authors correctly concluded that it was not appropriate to calculate predictive values in a population enriched for the disease of interest.

Highly accurate diagnosis of GORD has not been demonstrated in any of these studies. Critical to any study of diagnostic accuracy is the comparative measure—the gold standard—used to classify patients as having GORD. Although initially proposed as a “gold standard” for the diagnosis of GORD, 24 hour ambulatory pH monitoring is probably not a gold standard, as sensitivity probably does not exceed 75%.^{7, 8} Comparing a new diagnostic method with an inferior standard risks underestimating the accuracy of the new test, as it will always be inferior to the existing standard.¹ The questionnaire and interviewer methods have not been developed and validated following established, rigorous methods.

Enhancing diagnostic accuracy

Multiple approaches have been suggested to enhance the suboptimal diagnostic accuracy of GORD symptom surveys, and rigorous development and validation of questionnaires and interviewer methods is underway.⁹ Diagnostic treatment trials may increase specificity. Similar considerations have promoted interest in systematic exploration of symptom weighting, including predominant symptom identification and selection of symptoms supporting and detracting from the diagnosis.

Rigorous development of interview methods or questionnaires requires identifying accurate descriptive wording for these symptoms. Traditionally, experts define symptoms with limited input from patients about the content and clarity of these definitions. The descriptive language rarely receives

formal testing. Patient interpretation of the word “heartburn” highlights the need to do this. Representative patient responses during cognitive interviews for a description of the meaning of the word “heartburn” included “...can feel acid rising...”, “...burning sensation in the stomach...”, “...a bad taste in my mouth...”, and “...painful burning in stomach and whole GI tract...”.⁹ Two of these four responses are not consistent with the definition of heartburn: “epigastric/retrosternal burning feeling that rises into the chest”. One is possibly due to regurgitation and the other non-specific abdominal pain.

Patient interpretation of the word “heartburn” has also been evaluated quantitatively.^{5, 10} The sensitivity of the word “heartburn” was compared with the word picture “a burning feeling rising from your stomach or lower chest up towards your neck” in a Swedish population.⁵ Only 13% of a population with upper gastrointestinal complaints endorsed “heartburn” as their predominant symptom, compared with 40% who acknowledged that the word picture for reflux accurately identified the predominant symptom when responding to a self report questionnaire.

Inaccurate interpretation of the word “heartburn” was seen in a US population, although the results are not as remarkable.¹⁰ Patients’ questionnaire responses to the word “heartburn” were compared with a gastroenterologist interview where the interviewer had no knowledge of responses to the questionnaire.¹⁰ Overall agreement was 82%. When patients selected “heartburn”, the physician did not agree that the symptom was reflux 12% of the time. In 6% of patients, the physician interpreted the patient’s symptoms as heartburn although the patient did not endorse “heartburn” on the survey. The kappa value of 0.57 indicated fair agreement. Patient interpretation of “heartburn” in the USA showed a less remarkable decline in sensitivity with reduction in specificity being more significant.

Regurgitation and epigastric pain have not received as detailed scrutiny. Excellent agreement was seen when using a description for regurgitation instead of the words “acid regurgitation” compared with a gastroenterologist interview with overall agreement of 88% and kappa of 0.76.¹⁰

The reflux disease questionnaire (RDQ) was developed according to established principles of questionnaire design and validation to overcome the limitations of previous questionnaires.⁹ Content and wording were developed with multiple iterations of expert opinion and patient interviewing. On completion of item wording, standard psychometric analyses—including multitrait scaling—resulted in a 12 item questionnaire of three scales (heartburn, regurgitation, and dyspepsia) that was reliable and valid. The diagnostic validity of the RDQ remains to be established, however.

An intriguing observation during the development of the RDQ was that the symptom response to self directed, over the counter (OTC) antisecretory medications was the strongest predictor of GORD diagnosis (table 2).⁹ The predictive power of this single question surpassed that of either RDQ reflux scale. However the OTC response question showed poor internal validity and internal consistency resulting in its deletion from the final questionnaire. The work of Fass and colleagues also suggests that a structured approach to assessing treatment response may be the preferred diagnostic method.⁸ A diagnostic treatment trial using the RDQ to select patients and assess treatment response to standardised treatment with superior acid suppression may retain the observed strength of the OTC response question, without compromising reliability and validity. Response to antisecretory medications, however, would be inappropriate for a reflux questionnaire being used to select GORD patients for trials that compared acid suppressive therapy with an alternative. Selecting patients on the basis that they respond

Table 2 Enhancing diagnostic accuracy: potential role for treatment trials

| Predictive validity | p Value |
|-------------------------|---------|
| Positive OTC response* | 0.0076 |
| RDQ heartburn scale | 0.01 |
| RDQ regurgitation scale | 0.01 |

*Response to over the counter antisecretory therapy.
RDQ, Reflux Disease Questionnaire.

to antisecretory therapy before entering the trial would give an overly optimistic assessment of the efficacy of these treatments.

Selection of the predominant symptom has been evaluated as a method to enhance diagnostic accuracy.³ This approach is severely limited, as the majority of patients do not have a predominant symptom, with only 124 of 304 patients able to endorse a predominant symptom in one study.³ Although specificity of heartburn (89%) and regurgitation (95%) were quite high in this subset, sensitivity suffered significantly (heartburn 38%; regurgitation 6%). The inability of the majority to identify a predominant symptom eliminates the potential of this approach to enhance diagnostic accuracy because of the degree of compromise in sensitivity. Predominant symptom identification has been shown to be effective in selection of patients who will respond to proton pump inhibitors (PPIs) in clinical trials.¹⁰ Although treatment response and diagnostic accuracy are related, they are not equivalent.

Questionnaire score weighting has been used in one study.⁵ The questionnaire included questions felt not only to contribute to the diagnosis, but also to detract from the diagnosis. The investigators supplied positive and negative weightings without empiric data. To establish valid question weights requires prohibitively large datasets.² It is feasible to do this only for questionnaires receiving large volume application throughout a population.

SUPPORT FOR SPECIFIC TREATMENT SELECTION

GORD symptom questionnaires or structured interviews have received limited study in the application of specific and efficacious treatment. Individual questions and questionnaire total score were examined for the capability to predict symptom response to PPI therapy.⁵ The questionnaire total score did not predict response although endorsement of the individual items "burning rising feeling" or "relief from antacids" did identify those who were significantly more likely to respond to treatment. The failure of the questionnaire total score to successfully identify treatment responders limits the diagnostic utility of the questionnaire.

Predominant symptom selection has been shown to predict response to PPI therapy if the predominant symptom is heartburn.¹¹ Given the compromise in sensitivity as discussed above, such a question adversely affects diagnostic accuracy. In addition, the predominant symptom is dynamic and may change from week to week.

Implementation of treatment trials as part of the diagnostic strategy using questionnaires will inevitably show strong support for selection of specific treatments. This would be useful for the practitioner, but would not be applicable for most clinical treatment trials.

OUTCOMES IMPROVEMENT

Multiple outcomes are of interest to patients and physicians including absence of symptoms, improved health related quality of life, healing of oesophagitis, and maintenance of remission. The impact on these outcomes when GORD diagnosis is facilitated by valid questionnaires or structured interview has not been studied. As studies of endoscopy negative reflux disease show high efficacy when patients are selected by less rigorous measures,¹² it can be reasonably anticipated that validated symptom measures would perform similarly. Such studies will require that the symptom measure is the ultimate criterion leading to subject entry into the trial.

SUMMARY

Despite the importance of symptoms in identifying GORD patients, empiric support for the diagnostic utility of reflux symptom measures is lacking. Reliable reflux symptom questionnaires have been developed with content validity. These questionnaires need to be evaluated in terms of diagnostic accuracy, support for application of specific treatment, and improved outcomes resulting from their use.

REFERENCES

- 1 Streiner DL, Norman GR. *Health measurement scales*. 2nd ed. New York: Oxford Medical Publications, 1995.
- 2 Sackett DL, Haynes RB, Guyatt GH, et al. *Clinical epidemiology*. 2nd ed. Boston, MA: Little, Brown, 1991.
- 3 Klausner AG, Schindlbeck NE, Muller-Lissner SA. Symptoms on gastroesophageal reflux disease. *Lancet* 1990;**335**:205-8.
- 4 Johnsson F, Roth Y, Damgaard NE, et al. Cimetidine improves GERD symptoms in patients selected by a validated GERD questionnaire. *Aliment Pharmacol Ther* 1993;**7**:81-6.
- 5 Carlsson R, Dent J, Bolling-Sternevald E, et al. The usefulness of a structured questionnaire in the assessment of symptomatic gastroesophageal reflux disease. *Scand J Gastroenterol* 1998;**33**:1023-9.
- 6 Moayyedi P, Duffy J, Delaney J. New approaches to enhance the accuracy of the diagnosis of reflux disease. *Gut* 2004;**53**(Suppl IV):iv55-7.
- 7 Fass R, Ofman JJ, Gralnek IM, et al. Clinical and economic assessment of the omeprazole test in patients with symptoms suggestive of gastroesophageal reflux disease. *Arch Int Med* 1999;**159**:2161-8.
- 8 Fass R, Fennerty MB, Ofman JJ, et al. The clinical and economic value of a short course of omeprazole in patients with noncardiac chest pain. *Gastroenterol* 1998;**115**:42-9.
- 9 Shaw MJ, Talley NJ, Beebe TJ, et al. Initial validation of a diagnostic questionnaire for gastroesophageal reflux disease. *Am J Gastroenterol* 2001;**96**:52-7.
- 10 Locke GR, Talley NJ, Weaver AL, et al. A new questionnaire for gastroesophageal reflux disease. *Mayo Clinic Proc* 1994;**69**:539-47.
- 11 Talley NJ, Meineche-Schmidt V, Pare P, et al. Efficacy of omeprazole in functional dyspepsia: double-blind, randomized, placebo-controlled study (the BOND and OPERA studies). *Aliment Pharmacol Ther* 1998;**12**:1055-65.
- 12 Lind T, Havelund T, Carlsson O, et al. Heartburn without esophagitis: efficacy of omeprazole therapy and features determining therapeutic response. *Scand J Gastroenterol* 1997;**32**:974-9.