Accuracy of the diagnosis of GORD by questionnaire, physicians and a trial of proton pump inhibitor treatment: the Diamond Study

John Dent,[1] Nimish Vakil,[2] Roger Jones,[3] Peter Bytzer,[4] Uwe Schöning,[5] Katarina Halling,[6] Ola Junghard,[7] Tore Lind[8]

► Additional material published online only. To view these files please visit the journal online (http://gut.bmj.com).

For numbered affiliations see end of article.

Correspondence to
Professor John Dent, Department of Gastroenterology & Hepatology, Royal Adelaide Hospital, Adelaide, SA 5000, Australia;
john.dent@health.sa.gov.au

Dedication
This paper is dedicated to the memory of one of the authors, Dr Ola Junghard, who died on 11 November 2009 after a very long illness. Ola Junghard's statistical skills were vital to the success of this study, from its protracted design phase to repeated interrogation of the study database. His lively interest in the gathering and interpretation of data on gastrointestinal symptoms represents a major contribution to knowledge in this area.

Contributors
Peter Kahrilas and John Pandolfino, Northwestern University, Feinberg School of Medicine, Chicago, USA, served as paid consultants to monitor the quality of oesophageal pH recordings and to analyse the pH and symptom association data.
The authors were all involved in the development of the study protocol, interaction with investigators from study sites, interpretation of the full data set and preparation of the manuscript for publication. US and TL are currently employed by AstraZeneca. OJ and KH are past employees of AstraZeneca. JD, NV, RJ and PB have provided consultancy services to AstraZeneca. All statistical analyses were performed by OJ, within AstraZeneca, including exploratory analyses requested by other authors. All authors external to AstraZeneca had free access to the entire database.

Revised 3 January 2010
Accepted 26 January 2010

ABSTRACT

Objective The aim of this study was to determine the accuracy of the diagnosis of gastro-oesophageal reflux disease (GORD) by the Reflux Disease Questionnaire (RDQ), family practitioners, gastroenterologists and a test of esomeprazole therapy.

Methods This was a single-blind, single-arm study over 3–4 weeks from September 2005 to November 2006. Each symptom-based diagnostic assessment was made blinded to prior diagnoses. Patients were those presenting to their family practitioner with troublesome upper gastrointestinal symptoms (n=308). The RDQ was completed and a symptom-based diagnosis was made by the family practitioner. Placebo esomeprazole was started. Gastroenterologists made a symptom-based diagnosis and then performed endoscopy with 48 h oesophageal pH and symptom association monitoring to determine the presence/absence of GORD. Symptoms were recorded during treatment with 40 mg of esomeprazole for 2 weeks. The main outcome measure was RDQ scoring for the presence of GORD compared with symptom-based diagnosis by family physicians and gastroenterologists.

Results GORD was present in 203/308 (66%) patients. Only 49% of the patients with GORD selected either heartburn or regurgitation as the most troublesome symptom. Sensitivity and specificity, respectively, of the symptom-based diagnosis of GORD, were 62% and 67% for the RDQ, 63% and 63% for family practitioners, and 67% and 70% for gastroenterologists. Symptom response to esomeprazole was neither sensitive nor specific for the diagnosis of GORD.

Conclusions The RDQ, family practitioners and gastroenterologists have moderate and similar accuracy for diagnosis of GORD. Symptom response to a 2 week course of 40 mg of esomeprazole does not add diagnostic precision.

Clinical trial number NCT00291746.

Significance of this study

What is known already about this subject?
- Heartburn and regurgitation are considered to be distinctive symptoms of GORD.
- In a significant, but poorly defined proportion of patients with GORD, other symptoms predominate.
- Symptom evaluation is an important approach to diagnosis of GORD, given the limitations of endoscopy and other investigations.

What are the new findings?
- There was a wide spectrum of symptoms in patients with GORD, with only 49% reporting heartburn or regurgitation as their most troublesome symptom.
- Family practitioners have only modest accuracy for the symptom-based diagnosis of GORD, virtually identical to the Reflux Disease Questionnaire (RDQ).
- The symptom-based diagnosis of GORD by gastroenterologists was only slightly better than that of family practitioners and the RDQ
- A 2 week trial of esomeprazole treatment did not enhance the accuracy of symptom-based diagnosis.

How might it impact on clinical practice in the foreseeable future?
- Development of better-informed stepwise approaches to diagnosis of GORD which maximise the value, but also recognise the limitations of diagnostic symptom evaluation.
- Provide an impetus to formal evaluation of brief diagnostic questionnaires as a diagnostic and decision support tool for use in family practice.
- Reduce apparently unwarranted confidence in the value of a test of proton pump inhibitor treatment as a support to the diagnosis of reflux disease.

INTRODUCTION

Gastro-oesophageal reflux disease (GORD, reflux disease) is a major healthcare issue because of its high prevalence, impact on quality of life and significant cost.[1 2] Because of the limited accuracy and the costs of diagnostic tests, current guidelines recommend making a clinical diagnosis of GORD based on the cardinal symptoms of heartburn and acid regurgitation.[3 4] Despite this, there has been relatively little rigorous research on the spectrum of symptoms in reflux disease and the accuracy and best processes for making a symptom-based diagnosis. Gastroenterologists and family practitioners acquire this skill predominantly by experience in practice, using analysis of symptom patterns and assessment of how much the reflux-induced symptoms are troubling the patient. This latter evaluation is important as reflux-induced symptoms are highly prevalent in the general population, but in only a minority are these symptoms

sufficiently severe to justify them being regarded as causing 'disease'.[5]

Short patient self-report (written completion without assistance from another person) diagnostic questionnaires are used in family practice as standardised instruments to aid the symptom-based diagnosis of several highly prevalent disorders. The Reflux Disease Questionnaire (RDQ) (online appendix 1) is a short patient self-report diagnostic questionnaire which has been formally developed through an iterative process.[6] The accuracy of a short diagnostic questionnaire for reflux disease, such as the RDQ, has not been adequately assessed in the setting of family practice. Also, there are no data that evaluate and compare the accuracy of the symptom-based diagnosis of reflux disease by questionnaire and by physicians.

The main aims of this study, known as the Diamond Study in reflux disease, were to determine the accuracy of the symptom-based diagnosis of reflux disease by the RDQ[6] and to compare this with the symptom-based diagnosis of GORD by family practitioners and gastroenterologists. The diagnostic value of a 2 week trial of proton pump inhibior (PPI) treatment[7] was also tested. Since there is no single 'gold standard' test for GORD,[3 4 7 8] endoscopy and wireless 48 h pH recording[9] with symptom association monitoring[10] were used in all patients to provide an independent and objective reference standard for the diagnosis of reflux disease.

METHODS

Study population

Patients aged 18–79 years with symptoms considered by their family practitioner to be of upper gastrointestinal origin were assessed at 73 family practitioner clinics in Germany, Sweden, Canada, Denmark, Norway and the UK. For study entry, symptoms had to have occurred at least twice a week for ≥4 weeks prior to presentation, and on at least 3 days during the week prior to inclusion.

Major exclusion criteria were: upper gastrointestinal endoscopy during the year prior to the study; previous antireflux surgery or surgery for peptic ulcer or other gastrointestinal resections; use of prescription or over-the-counter PPIs within 2 months prior to Visit 1; daily aspirin intake >165 mg/day or daily non-steroidal anti-inflammatory drugs (NSAIDs), including cyclo-oxygenase-2 (COX-2) inhibitors, at any dose; and alert (or 'alarm') features, such as unintentional weight loss in the previous 3 months, haematemesis, melaena or rectal bleeding in the previous year, progressive dysphagia, anaemia or any other symptom suggestive of malignancy. Patients with Los Angeles (LA) grade D oesophagitis were allowed to remain in the study but were excluded from pH monitoring for safety reasons. Other patients with contraindications to attachment of the pH probe[9] were excluded.

Study schedule

The protocol (figure 1) was approved by ethics committees at the individual centres. The study was supported in full by AstraZeneca (ClinicalTrials.gov: number NCT00291746).

Screening visit

Potentially eligible patients consented to screening and completed the RDQ (online appendix 1), unaware that it would be used to make a diagnosis of reflux disease. From a protocol-specified list, the family practitioner recorded a 'most likely' symptom-based diagnosis, which was not revealed to the patient.

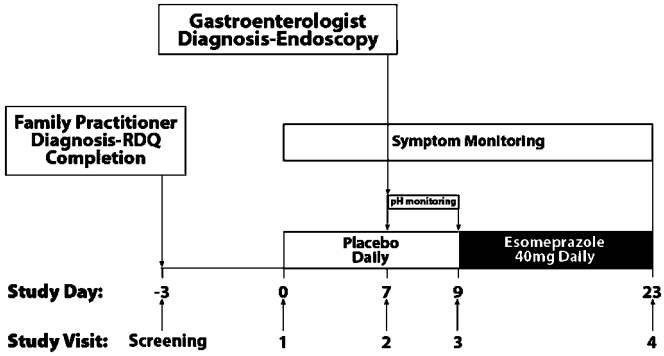

Figure 1 Outline of the main study interventions and their timings. RDQ, Reflux Disease Questionnaire.

At Visit 1, within 3 days of initial screening, patient eligibility was confirmed and informed consent obtained for the entire study. Patients were aware that they would receive both placebo and active drug during the course of the study but not when and for how long. Patients identified their most and second most troublesome symptoms from a list of 19 symptoms described in lay language, based on validated word pictures from the Gastrointestinal Symptom Rating Scale (GSRS).[11] A paper symptom diary card was completed daily at home throughout the study. Single-blinded placebo esomeprazole was started.

Visit 2 occurred at one of the 22 specialist centres, no more than 10 days after the screening visit. Prior to endoscopy and blind to the family practitioner diagnosis, the gastroenterologist also made a symptom-based diagnosis that was not revealed to the patient. Upper gastrointestinal endoscopy was then performed; any abnormality that placed the patient at risk by remaining in the study required patient withdrawal. Otherwise, endoscopic findings were not disclosed. *Helicobacter pylori* screening was performed on gastric biopsies (HUT test).[12] Directly after endoscopy, a wireless pH-monitoring capsule[9] (Bravo) was anchored in the distal oesophagus and oesophageal pH recorded for 48 h (see below).

At Visit 3, 48 h later, patients returned the pH monitor and the special pH monitoring diary card. Remaining placebo capsules were returned and counted, and replaced with a 17 day supply of esomeprazole 40 mg daily.

At Visit 4, 14±3 days after the start of esomeprazole treatment, therapeutic adherence was recorded and the symptom diary card retrieved. Outcomes of diagnostic assessments were then explained to the patient and further management was planned according to normal clinical practice.

Study investigations

The RDQ

The RDQ is a self-administered, 12-item diagnostic questionnaire that was formally developed by an iterative process that involved patient focus groups and an expert working party.[6] Six symptoms covering heartburn, regurgitation and upper abdominal pain are evaluated separately for their frequency and intensity by a 6-point modified Likert scale (online appendix 1).

Endoscopy

Reflux oesophagitis was graded according to the LA classification.[13] Other endoscopic abnormalities were noted.

Oesophageal pH monitoring

Intraoesophageal pH was monitored in all patients, other than those with LA grade D oesophagitis in whom this is

contraindicated.[9] Acidic food or beverages and alcohol were not allowed during pH monitoring. Oesophageal acid exposure was analysed for 24 h from the first midnight on the first day after pH capsule placement, in order to standardise the time window for pH data analysis, since pH capsules were placed during both morning and afternoon. This approach also allowed time for patients to recover from the effects on reflux patterns of fasting and sedation associated with endoscopy and pH capsule placement.

Symptom event monitoring during oesophageal pH measurement

Patients were given written and verbal instructions on symptom event marking during the 48 h pH monitoring time and used a special diary card to record the occurrence and severity (mild, moderate or severe) of the two most troublesome symptoms that they had selected at enrolment. The timing of these symptoms was recorded by event markers. The symptom association probability (SAP)[10] was determined over 24 h from midnight on the day the pH capsule was placed.

Investigation-based criteria for reference diagnosis of GORD

For the primary predetermined analysis, GORD was diagnosed when at least one of the following four protocol-defined criteria was present:

1. Reflux oesophagitis (LA grades A–D).
2. Oesophageal pH <4 for >5.5% of the time.
3. Positive SAP (≥95%) for association of symptoms with acid reflux.
4. Borderline high oesophageal acid exposure (oesophageal pH <4 for 3.5–5.5% of time) and positive response of reflux-related symptoms to esomeprazole treatment (absence in the last 3 days of esomeprazole treatment of heartburn, regurgitation, central chest pain or dysphagia). This analysis applied only to patients who had identified one or two of these symptoms as the first and/or second most troublesome symptom. These evaluation criteria differed from those used for the analysis of the overall outcome of the trial of esomeprazole treatment (see below).

Statistical analyses

Sample size

The prevalence (ie, diagnosis) of GORD was assumed to be 50% in the study cohort. Thus, with 315 evaluable subjects, the prevalence of GORD could be estimated with a 95% CI of ±6%, and the sensitivity and specificity (both assumed to be 70%) with a 95% CI of ±7%.

Main population analysed

The fully evaluable population included the 308 patients who had completed the screening RDQ and the diagnostic procedures according to the protocol (figures 1 and 2).

RDQ evaluation

Receiver operating characteristic (ROC) curves were constructed for two prespecified methods for scoring items from the RDQ (online appendix 1). The first method took the sum of the scores for frequency and severity of 'Burning behind the breastbone' and 'Unpleasant movement of material upwards from the stomach', with subtraction from this of the pain or discomfort in the upper stomach (dyspepsia)[14] score. This scoring approach was developed because of unpublished exploratory data from an earlier study,[15] that suggested that pain in the upper abdomen may be a negative predictor for GORD. The second approach[6] simply added the item scores for the RDQ subscale 'GORD' (frequency and intensity of the two heartburn and the two regurgitation items).

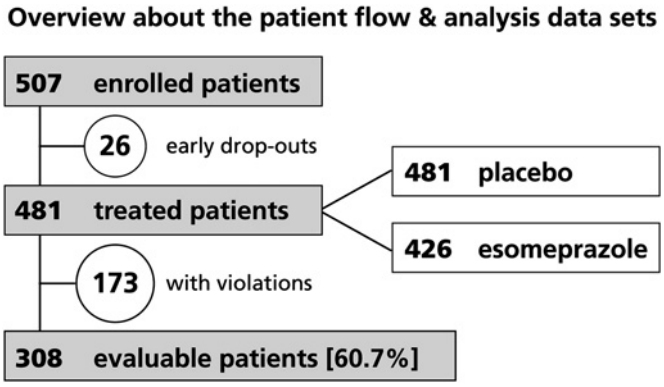

Figure 2 Patient flows in the study. Most common violations in the 199 excluded patients were: pH monitoring not done (n=78); unsatisfactory quality of pH recording (n=32); some visits out of time range (n=39); and intake of disallowed medication (n=20).

Acid inhibitory treatment test

This was defined as positive for GORD if the daily symptom diary recorded that the originally selected most troublesome symptom, regardless of symptom type, was absent during the last 3 days of esomeprazole treatment.

RESULTS

Patient population

Of the 706 patients that were formally screened, 507 were enrolled; of these, 308 (61%) were fully evaluable and are the focus of this report. Patient flows are summarised in figure 2. The excluded patient population had a similar distribution of symptoms (table 1), demographics and endoscopic findings (table 2) to the 308 fully evaluable patients. No malignancies were discovered endoscopically in the entire patient population, and no patient was withdrawn from the study because of endoscopic findings that made it clinically inappropriate for continuation in the study.

Investigations in fully evaluable GORD and non-GORD populations

Patients with GORD

Two hundred and three patients were diagnosed with GORD using the investigation-based criteria (see the Methods section). Of these, 116 had erosive oesophagitis (93% LA grades A and B, table 2, figure 3). Oesophageal pH monitoring was abnormal in 66% (n=77) of patients with oesophagitis; only 63% had abnormal acid exposure (figure 3). The small contribution of the SAP is also shown in figure 3. GORD was diagnosed in another 87 patients without erosive oesophagitis on the basis of either abnormal acid exposure or a positive SAP (figure 3). In nine patients, GORD was diagnosed on the basis of a borderline oesophageal acid exposure and response to esomeprazole of heartburn, regurgitation, dysphagia or central chest pain (see the Methods section). Peptic ulcer was present in 8 of the 203 patients with GORD, and hiatal hernia (judged by the personal criteria of the endoscopist) in 92 (45%).

Patients without GORD

The only endoscopic abnormalities noted in the 105 patients without GORD were peptic ulcer in 1 and hiatal hernia in 28 (27%).

Gut 2010;**59**:714–721. doi:10.1136/gut.2009.200063

Table 1 Prevalence of first and second most troublesome symptoms at entry to the study

Symptom	Most troublesome			Second most troublesome		
	Excluded n=192* %	Non-GORD n=105 %	GORD n=203 %	Excluded n=192* %	Non-GORD n=105 %	GORD n=203 %
Heartburn	33.3	21.0	40.4	18.2	14.3	25.1
Dyspepsia	17.7	22.9	21.2	20.8	19.0	17.2
Bloating	12.0	16.2	9.4	10.9	14.3	15.3
Regurgitation	6.3	4.8	8.9	11.5	6.7	16.7
Abdominal pain/discomfort, not dyspepsia	14.6	16.2	9.9	16.7	14.3	8.9
Belching	5.2	4.8	4.4	5.2	8.6	3.9
Early satiety/postprandial fullness	3.6	2.9	1.5	5.2	6.7	4.9
Dysphagia	0.5	0	2.0	2.6	1.0	1.0
Nausea	3.6	10.5	1.5	5.7	12.4	3.9
Pancreatic pain	1.0	1.0	0.5	2.1	2.9	2.5
Vomiting	2.1	0	0.5	1.0	0	0.5

*For seven excluded patients, there was no recording of the most and second most troublesome symptoms.
GORD, gastro-oesophageal reflux disease.

Symptom profiles in the fully evaluable patient population

The prevalences of the most and second most troublesome symptoms selected by patients at enrolment from the list of 19 lay-language descriptors are shown in table 1. Consistent with the Rome II criteria,[14] these symptom descriptors were collapsed down to the 11 symptoms shown as follows: 'heartburn' also included central chest pain; 'dyspepsia' included pain or discomfort in the centre of the upper abdomen[16]; 'abdominal pain/discomfort, not dyspepsia' included generalised, left- or right-sided abdominal pain or discomfort; and 'early satiety' and 'postprandial fullness' were combined. Heartburn or regurgitation was the most troublesome symptom in 49% of patients with GORD and the second most troublesome in 42%. Notably, dyspepsia (see above) was identified as the most and second most troublesome symptom in 21% and 17% of patients with GORD, respectively.

For the patients without GORD, dyspepsia was most common (23% and 19%, the most and second most troublesome symptom, respectively), but heartburn was selected almost as frequently (21% and 14%, the most and second most troublesome, respectively). Nausea was the most or second most troublesome symptom in 23% of patients without GORD, in contrast to 5% of the patients with GORD.

Predefined investigation-based versus the symptom-based diagnosis of GORD

Sensitivities/specificities for diagnosis of GORD were 63%/63% for family practitioners and 67%/70% for gastroenterologists.

ROC curves (figure 4) plot the results of the two prespecified methods for scoring of RDQ items and the diagnosis of GORD by physicians against the investigation-based diagnosis of GORD. The first prespecified RDQ scoring criteria, which used pain or discomfort in the centre of the upper stomach as a negative predictor of GORD, gave the best RDQ-based diagnosis of GORD, with a sensitivity and specificity equal to family practitioners and slightly inferior to gastroenterologists (figure 4a). Figure 5 shows the proportions of patients with GORD for four ranges of RDQ score, determined with the best performing scoring method (figure 4a). Fifty-eight per cent of the GORD population had RDQ scores in the two highest ranges.

Forty-nine per cent of patients with GORD and 26% of patients without GORD identified either heartburn or regurgitation as most troublesome, giving a sensitivity and specificity for GORD of 49% and 74%, respectively, for this patient subgroup. If either most or second most troublesome symptoms are used, heartburn or regurgitation were chosen by 69% (140/203) of patients with GORD and 38% (40/105) of patients

Table 2 Demography and disease details for patients included or excluded from the analyses

Demography	Total population		Fully evaluable population		
	Excluded n=199	Fully evaluable n=308	Non-GORD n=105	GORD Erosive oesophagitis n=116	GORD Non-erosive n=87
Mean age, years (SD)	47 (15)	47 (14)	45 (16)	48 (13)	47 (13)
Gender M/F	79/120	143/165	27/78	75/41	41/46
Mean BMI (kg/m^2)	27.2 (5.1)	26.6 (4.5)	25.1 (4.4)	27.5 (4.3)	27.2 (4.6)
Mean symptom duration (years)	4.1	4.1	3.0	4.4	5.0
H pylori positive	38 (19%)	71 (23%)	24 (23%)	22 (19%)	25 (29%)
No reflux oesophagitis	80%	62%	100%	0%	100%
LA classification A	21 (11%)	59 (19%)	0	59 (51%)	0
LA classification B	15 (8%)	49 (16%)	0	49 (42%)	0
LA classification C	3 (1.5%)	6 (2%)	0	6 (5%)	0
LA classification D	0	2 (0.6%)	0	2 (2%)	0
Hiatal hernia	43 (22%)	120 (39%)	28 (27%)	62 (53%)	30 (34%)
Gastric ulcer	4	5	1	1	3
Duodenal ulcer	5	4	0	1	3

BMI, body mass index; F, female; GORD, gastro-oesophageal reflux disease; LA, Los Angeles; M, male.

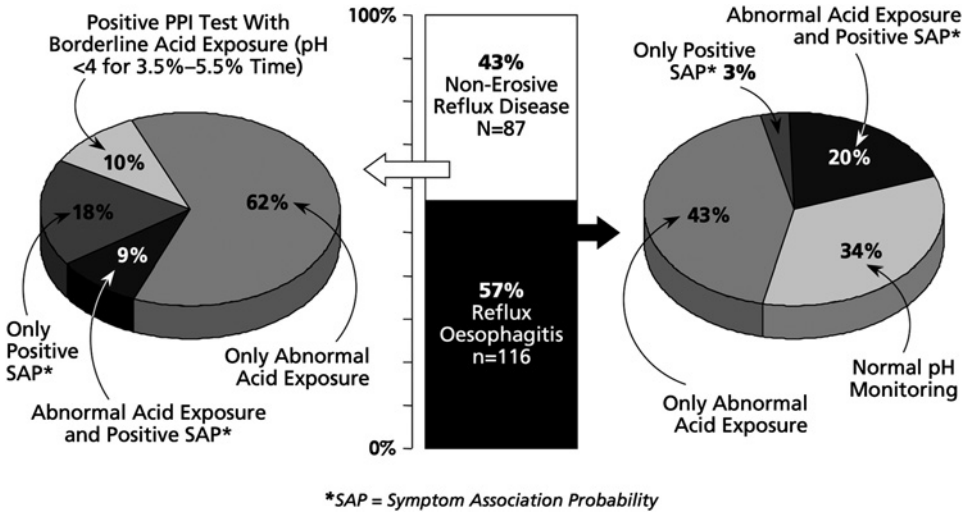

Figure 3 Patterns of oesophageal pH monitoring findings in the 203 patients with gastro-oesophageal reflux disease (GORD). PPI, proton pump inhibitor.

without GORD (sensitivity and specificity for GORD of 69% and 62%, respectively).

Contribution of a trial of acid inhibitory treatment to diagnosis of GORD

A total of 296 patients was evaluable for the outcome of the trial of PPI treatment (table 3). A positive response to esomeprazole was found in 54% (106/197) of patients with GORD compared with 35% (35/99) of patients without GORD. In 43 patients, there was a positive response to placebo, given before active esomeprazole was started (n=28 with reflux disease; n=15 without reflux disease). Placebo responders were defined as patients in whom the most troublesome symptom was absent for the last 2 days of placebo treatment. Restriction of the analysis to the non-responders to placebo (n=253) did not improve test characteristics (table 3). The sensitivity (71%) and positive predictive value (84%) of the test were higher in the subgroup of patients (n=127) who reported a reflux symptom (heartburn, central chest pain, regurgitation or dysphagia) as their most troublesome, but specificity and likelihood ratios did not improve (table 3). Response to treatment was slightly higher in patients with oesophagitis (57% (64/112)) compared with patients with non-erosive GORD (49% (42/85)).

Impacts of reclassifying patients with borderline oesophageal pH monitoring

Thirty-one of the patients had a borderline acid exposure value (oesophageal pH <4 for 3.5–5.5% of time), a negative SAP and no endoscopically evident oesophagitis. As noted above, nine of these patients were defined as having reflux disease according to the predefined criteria of resolution of heartburn, regurgitation, dysphagia or central chest pain during esomeprazole treatment.

Because of reservations about using response of typical reflux disease symptoms to esomeprazole as a diagnostic criterion to resolve the ambiguity of a borderline high acid exposure value and a negative SAP, a posthoc analysis reclassified the nine patients originally diagnosed as having reflux disease on the basis of a response to esomeprazole to the 'not reflux disease' group. There was minimal change in the study outcomes (RDQ sensitivity and specificity 62% and 66%, vs the primary analysis of 62% and 67%).

A further posthoc analysis took no account of response to esomeprazole and censored all 31 patients in whom borderline acid exposure was the only possibly abnormal investigative finding for reflux disease, on the basis that the ambiguity of this result with regard to presence/absence of reflux disease could not be resolved. Again, exclusion of these patients had effectively no impact on the major study findings (RDQ sensitivity and specificity 63% and 67% vs the primary analysis of 62% and 67%). The sensitivity and specificity of the diagnosis of reflux disease by family practitioners and gastroenterologists were

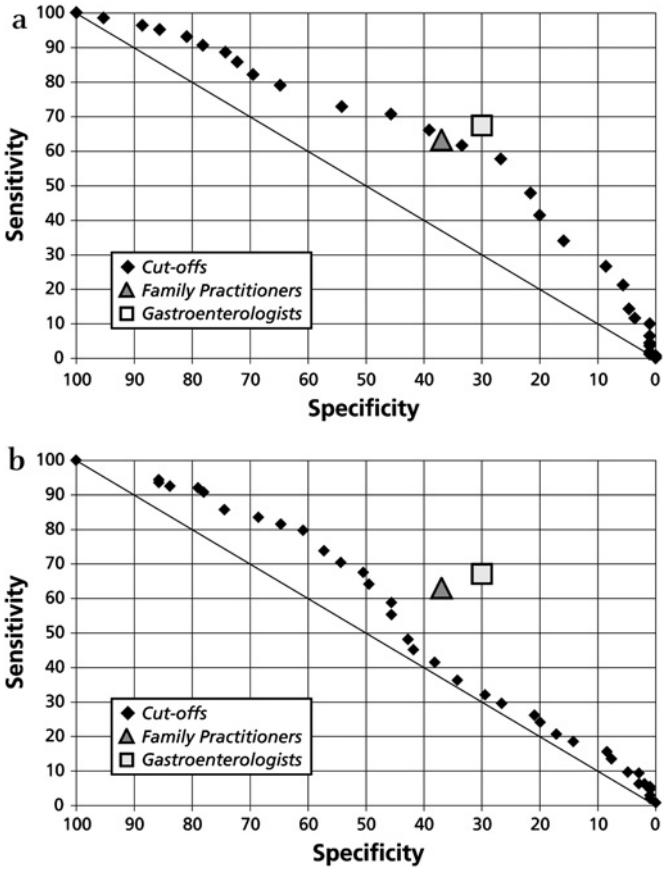

Figure 4 (a) Receiver operating characteristic curve for the best-performing prespecified scoring method for Reflux Disease Questionnaire (RDQ) responses, which assigned a negative value to responses on dyspepsia. (b) Receiver operating characteristic curve for the prespecified scoring method for RDQ responses, which added the scores of the heartburn and regurgitation items.

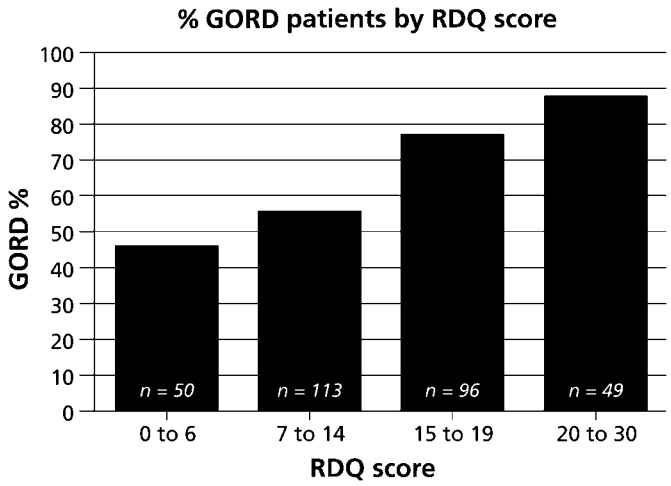

Figure 5 Proportions of patients with gastro-oesophageal reflux disease (GORD) for four ranges of Reflux Disease Questionnaire (RDQ) scores, according to first predetermined scoring method used for analysis shown in figure 4a.

essentially unchanged by exclusion of the patients with only borderline acid exposure values.

DISCUSSION

This study provides novel insights into the performance of different options for the diagnosis of reflux disease in an important patient group—those presenting in primary care with troublesome upper abdominal/lower retrosternal symptoms considered to be of gastrointestinal origin. Notably, the RDQ, a short, 12-item questionnaire for the diagnosis of reflux disease (online appendix 1), had a performance comparable with that of family practitioners and only slightly inferior to that of gastroenterologists. The poor performance of the trial of esomeprazole carried out in this study reinforces the importance of the initial diagnostic symptom assessment. Also, our study provides the best available assessment of the spectrum of symptoms in reflux disease, and shows that in half of reflux disease patients, the most troublesome symptom is neither heartburn nor regurgitation.

This study has several important strengths. It is well powered and generalisable to other primary care populations with a low prevalence of *H pylori* infection, such as in North America and Western Europe. Three separate and independently performed symptom-based diagnostic assessments were compared, and the best possible investigative approaches were used in all patients to determine the presence or absence of GORD. Oesophageal pH monitoring with symptom event marking was an important part of the investigation-based diagnosis; compliance with this was made possible by the excellent tolerance of the catheter-free pH monitoring method that was used.[9] It is possible that performance of 24 h impedance monitoring could have improved the accuracy of the investigation-based diagnosis of reflux disease, but, as yet, this test has not been formally validated for this purpose. Furthermore, a requirement for patients to undergo a 24 h catheter-based monitoring study would have had a major negative impact on patient enrolment, so that the patients from whom data were derived would have been unlikely to be representative of the target patient population.

This RDQ was formally designed for the purpose of separating reflux disease from other causes of upper abdominal/lower retrosternal symptoms in patients presenting in primary care. A multistep data- and expert group-driven process was used.[6] To have a chance of success in this setting, an instrument must be as brief as possible and use wording that has been tested and modified as needed in response to inputs from patient focus groups. To the best of our knowledge, the RDQ is the only diagnostic reflux disease questionnaire which has been formally developed for this purpose and setting. The GORD Impact Scale (GIS)[16] was also developed for use in primary care, but was designed to measure the impacts of reflux-induced symptoms, rather than to diagnose reflux disease. Several other questionnaires devised for use in reflux disease were designed primarily for epidemiological studies.[17] Most of the questionnaires intended to be used for diagnosis were designed on the basis of only expert opinion, and no questionnaire that is intended to diagnose GORD has been subjected to a rigorous validation such as we have now completed for the RDQ.

The less than ideal diagnostic accuracy of symptom-based diagnosis of reflux disease achieved in this study should not be interpreted as an invalidation of this approach. Symptom-based diagnosis is an inexpensive, immediately available option that can be standardised by use of the RDQ (online appendix 1), which is now freely available for use. The RDQ score gives a numerical estimate of the probability of reflux disease being present (figure 5). Such estimates are unlikely to be as reliable when made by physicians because of lack of standardisation. The use of diagnostic score ranges to guide choice of initial management merits further research. This study has shown that the presence of heartburn or regurgitation as the most troublesome symptom has a specificity for reflux disease of 74%. The limitation of the diagnostic utility of these symptoms is that they were present as the most troublesome symptom in only 49% of the patients defined as having reflux disease.

The utility of the symptom-based diagnosis of reflux disease should be judged by comparison with other diagnostic options in terms of accuracy, monetary and social cost, and ease of access. Taken alone, in this study, endoscopy had a sensitivity for reflux disease of 57%, albeit with a presumed high specificity (see below). The substantial limitations of oesophageal pH monitoring, with or without SAP (done relatively infrequently in routine practice), are also evident, since this was negative in 34% of patients with reflux oesophagitis (figure 3). The major conclusions from this study are that there is no perfect single test for reflux disease, that each test adds to diagnostic accuracy when this is needed and that an attempt at symptom-based diagnosis is an essential first step.

Table 3 Characteristics of an acid inhibitory treatment test to diagnose reflux disease

	Sensitivity	Specificity	PPV	NPV	LR+	LR−
All patients, n=296	54%	65%	75%	41%	1.52	0.71
Placebo non-responders, n=253	50%	73%	79%	42%	1.81	0.69
Patients with a most troublesome reflux symptom, n=127	71%	44%	84%	27%	1.26	0.67

LR+, likelihood ratio of a positive test; LR−, likelihood ratio of a negative test; NPV, negative predictive value; PPV, positive predictive value.

The RDQ has not been designed to search for alert (alarm) features that aim to identify patients at risk for major complications of reflux disease or at risk of other more serious upper gastrointestinal disorders. Practice guidelines advise early endoscopy in such patients. The physician must take responsibility for checking for alert (alarm) features, according to criteria that are tailored to regional patterns of upper gastrointestinal disease, as well as the cost and availability of endoscopy. A structured approach to symptom assessment remains an important part of any evaluation even if an early endoscopy is arranged, since major diagnostic uncertainty is, in itself, an alert feature. In developed countries, most patients who have endoscopy because of alert (alarm) features have no major or diagnostic endoscopic findings. Further management in such patients needs to be informed by symptom evaluation, followed by a trial of treatment in those in whom reflux disease is considered a likely cause. It is generally agreed now that oesophageal function testing should be reserved for patients who respond poorly. Though the present study excluded patients with alert features, it seems reasonable to believe that our results would also be applicable to them.

The definition of the significant limitations of the symptom-based diagnosis of reflux disease by this study should help with development of better strategies in routine patient care. Given the results achieved by the RDQ, it is probably time now to research how validated diagnostic questionnaires might be used to improve the practicalities of future management guidelines and decision-making algorithms.[18] The RDQ and an evolution of it, the GerdQ[19] (freely available—from http://www.gastrosource.com or directly from AstraZeneca), also have potential value for use in settings with limited access to physicians and as a tool to aid physician extenders who may become more important in years to come as the availability of physicians working in primary care decreases.[20]

The lack of a 'gold standard' test for reflux disease represents a major challenge for studies which rely on accurate judgement on the presence or absence of reflux disease. It is not possible to determine how well the predefined combination of investigative criteria identified the patients with reflux disease, but we believe that these criteria are at least 'best effort' consistent with effective pursuit of the aims of the study. We had expected the trial of esomeprazole treatment to give better results for diagnosis of reflux disease and, on this basis, used response to esomeprazole as a secondary criterion for diagnosis of reflux disease in patients with a borderline oesophageal acid exposure value. This strategy was driven by a lack of consensus on cut-off values for acid exposure and knowledge of the significant variability of acid exposure values from day to day.[8] Given that the criteria for diagnosis of reflux disease and the end points were predefined in the protocol, we adhered to these for the main data presentation. Posthoc analysis explored two other sets of criteria for the diagnosis of reflux disease; these showed that no matter how the patients with borderline acid exposure values were categorised (see the Results section), the main study findings were not changed. With hindsight, we would probably argue against the use of response to PPI to classify patients in any future study with aims similar to the Diamond Study.

The low utility of the 2 week trial of esomeprazole as a diagnostic test is consistent with other recent studies[21 22] and a meta-analysis.[7] A detailed exploration of the data from our trial of esomeprazole will be published separately.

Oesophagitis was present in what appears to be an unusually high proportion (57%) of patients with GORD, but similar data have been reported from a large Canadian study of patients consulting in primary care with upper gastrointestinal symptoms in whom oesophagitis was present in 42%.[19] It is not feasible to test directly the specificity of an endoscopic finding of erosive reflux oesophagitis for GORD, but the high rate of healing of oesophagitis and resolution of symptoms in placebo-controlled trials of PPI treatment are strong indirect evidence of high specificity.[23]

Our data demonstrate the wide spectrum of presenting troublesome symptoms in a GORD and non-GORD primary care population (table 1). The most important finding is that dyspepsia is common in patients with proven reflux disease and may be the most troublesome symptom. This may account for the observation that some patients with predominant dyspepsia respond to acid inhibition. Other studies have reported that in patients with predominant dyspepsia, pathological gastro-oesophageal reflux is present in ~20% of patients, some of whom have heartburn as a secondary symptom.[24] Our study provides the most definitive assessment of the overlap in symptoms between patients with GORD and patients with dyspepsia and no GORD.

The word 'dyspepsia' poses difficulties for studies such as this one and for clinical practice. 'Dyspepsia' does not describe a single distinctive symptom and the symptoms that are considered to be included under the umbrella of dyspepsia have changed substantially over time, as exemplified by the three published definitions for dyspepsia from the Rome Group. The Rome II definition for dyspepsia, which excludes heartburn from this umbrella descriptor, was current at the time that this study was planned and is the one we used.[14]

The main potential limitation of our study is that we had a significant number of drop-outs. The mixture of reasons for this (see table 2) is indicative of the complexity of the study protocol and, in particular, the need to perform endoscopy and technically successful wireless pH monitoring within a relatively short time window of the patient being seen in primary care. Data that describe the excluded patients (table 2) indicate that the evaluable population used for the RDQ evaluation was acceptably representative of the total enrolled population.

Our study indicates that the symptom-based diagnosis of GORD can be improved by paying attention to certain symptoms, including central upper abdominal pain or discomfort, that have, overall, a negative predictive value for a diagnosis of reflux disease. Although experienced clinicians probably incorporate some questions thought to have useful negative predictive value in their assessment, the focus of symptom-based assessments in upper gastrointestinal disorders has been largely on the positive predictive value of symptoms. Exploratory analyses on our data including negative predictors suggest that the performance of the RDQ itself may be improved; this is the subject of a separate publication.[19]

Author affiliations
[1]Royal Adelaide Hospital, Adelaide, Australia
[2]University of Wisconsin School of Medicine and Public Health, Madison, Wiscosin, USA
[3]Department of General Practice and Primary Care, King's College London, London, UK
[4]Køge University Hospital, Køge, Denmark
[5]Department of Medicine, AstraZeneca, Wedel, Germany
[6]PRO Consulting, Sweden
[7]Biostatistics, AstraZeneca R&D, Mölndal, Sweden
[8]Medicine and Science, AstraZeneca R&D, Mölndal, Sweden

Acknowledgements The authors wish to recognise and thank the investigators who enrolled and assessed patients at the 73 family practitioner clinics and 22 specialist centres involved in the study. The RDQ and the protocol of this study were developed by a working group which included the authors. The other members of this working group who are thanked for their contributions are: Brendan Delaney, Division of Health and Social Care Research, King's College London, UK; Pali Hungin,

Primary Care and General Practice, Durham University, UK; Paul Moayyedi, Division of Gastroenterology, McMaster University, Canada; Mattijs E Numans, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands; Mike Shaw, Park Nicollet Clinic, Institute for Research and Education, Minneapolis, USA; André Smout, Department of Gastroenterology, University Hospital Utrecht, The Netherlands; and Nick Talley, Department of Internal Medicine, Mayo Clinic Florida, USA. Esther Breed is thanked for typing and checking this manuscript.

Funding The study was fully funded by AstraZeneca.

Competing interests None.

Ethics approval This study was conducted with relevant approval from all of the 73 primary care and 22 specialist study centres.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Shaheen NJ,** Hansen RA, Morgan DR, *et al*. The burden of gastrointestinal and liver diseases, 2006. *Am J Gastroenterol* 2006;**101**:2128—38.
2. **Dent J,** El-Serag HB, Wallander M-A, *et al*. Epidemiology of gastro-oesophageal reflux disease: a systematic review. *Gut* 2005;**54**:710—17.
3. **Tytgat GN,** McColl K, Tack J, *et al*. New algorithm for the treatment of gastro-oesophageal reflux disease. *Aliment Pharmacol Ther* 2008;**27**:249—56.
4. **Kahrilas PJ,** Shaheen NJ, Vaezi MF. American Gastroenterological Association Institute technical review on the management of gastroesophageal reflux disease. *Gastroenterology* 2008;**135**:1392—413.
5. **Vakil N,** van Zanten S, Kahrilas P, *et al*. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. *Am J Gastroenterol* 2006;**101**:1900—20.
6. **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.
7. **Numans ME,** Lau J, de Wit NJ, *et al*. Short-term treatment with proton-pump inhibitors as a test for gastroesophageal reflux disease. *Ann Intern Med* 2004;**140**:518—27.
8. **American Gastroenterological Association.** American Gastroenterological Association medical position statement: guidelines on the use of esophageal pH recording. *Gastroenterology* 1996;**110**:1981.
9. **Pandolfino JE,** Richter JE, Ours T, *et al*. Ambulatory esophageal pH monitoring using a wireless system. *Am J Gastroenterol* 2003;**98**:740—9.
10. **Weusten BLAM,** Roelofs JMM, Akkermans LMA, *et al*. The symptom-association probability: an improved method for symptom analysis of 24-hour esophageal pH data. *Gastroenterology* 1994;**107**:1741—5.
11. **Svedlund J,** Sjödin I, Dotevall G. GSRS—a clinical rating scale for gastrointestinal symptoms in patients with irritable bowel syndrome and peptic ulcer disease. *Dig Dis Sci* 1988;**33**:129—34.
12. **Malfertheiner P,** Enrique Domínguez-Muñoz J, Heckenmüller H, *et al*. Modified rapid urease test for detection of *Helicobacter pylori* infection. *Eur J Gastroenterol Hepatol* 1996;**8**:53—6.
13. **Lundell LR,** Dent J, Bennett JR, *et al*. Endoscopic assessment of oesophagitis: clinical and functional correlates and further validation of the Los Angeles classification. *Gut* 1999;**45**:172—80.
14. **Talley NJ,** Stanghellini V, Heading RC, *et al*. Functional gastroduodenal disorders. *Gut* 1999;**45**(Suppl II):II37—42.
15. **Thomson ABR,** Barkun AN, Armstrong D, *et al*. The prevalence of clinically significant endoscopic findings in primary care patients with uninvestigated dyspepsia: the Canadian Adult Dyspepsia Empiric Treatment-Prompt Endoscopy (CADET-PE) study. *Aliment Pharmacol Ther* 2003;**17**:1481—91. Corrigenda in: *Aliment Pharmacol Ther* 2004;**20**:702.
16. **Jones R,** Coyne K, Wiklund I. The gastro-oesophageal Reflux Disease Impact Scale: a patient management tool for primary care. *Aliment Pharmacol Ther* 2007;**25**:1451—59.
17. **Stanghellini V,** Armstrong D, Mönnikes H, *et al*. Systematic review: do we need a new gastro-oesophageal reflux disease questionnaire? *Digestion* 2007;**75**(Suppl 1):3—16.
18. **Holloway RH.** Questionnaires in gastroesophageal reflux disease: what do the answers mean? *J Gastroenterol Hepatol* 2009;**24**:1311—12.
19. **Jones R,** Junghard O, Dent J, *et al*. Development of the GerdQ, a tool for the diagnosis and management of gastro-oesophageal reflux disease in primary care. *Aliment Pharmacol Ther* 2009;**30**:1030—8.
20. **Brotherton SE,** Rockey PH, Etzel SI. US graduate medical education, 2004—2005: trends in primary care specialties. *JAMA* 2005;**294**:1075—82.
21. **van Zanten SV,** Flook N, Talley NJ, *et al*. One-week acid suppression trial in uninvestigated dyspepsia patients with epigastric pain or burning to predict response to 8 weeks' treatment with esomeprazole: a randomized, placebo-controlled study. *Aliment Pharmacol Ther* 2007;**26**:665—72.
22. **Talley NJ,** Vakil N, Lauritsen K, *et al*. Randomized-controlled trial of esomeprazole in functional dyspepsia patients with epigastric pain or burning: does a 1-week trial of acid suppression predict symptom response? *Aliment Pharmacol Ther* 2007;**26**:673—82.
23. **Khan M,** Santana J, Donnellan C, *et al*. Medical treatments in the short term management of reflux oesophagitis. *Cochrane Database Syst Rev* 2007;(2): CD003244.
24. **Tack J,** Caenepeel P, Arts J, *et al*. Prevalence of acid reflux in functional dyspepsia and its association with symptom profile. *Gut* 2005;**54**:1370—6.

Accuracy of the diagnosis of GORD by questionnaire, physicians and a trial of proton pump inhibitor treatment: the Diamond Study

John Dent, Nimish Vakil, Roger Jones, et al.

Gut 2010 59: 714-721
doi: 10.1136/gut.2009.200063