Short-term treatment with proton-pump inhibitors as a test for gastroesophageal reflux disease: a meta-analysis of diagnostic test characteristics
Numans M E, Lau J, de Wit N J, Bonis P A

CRD summary
This review assessed the accuracy of symptomatic response to proton-pump inhibitor (PPI) treatment in the diagnosis of gastroesophageal reflux disease (GERD). The authors concluded that a response to short-term treatment with PPIs does not accurately diagnose GERD as defined by traditional objective criteria. The authors' conclusions are likely to be reliable.

Authors' objectives
To determine the diagnostic accuracy of symptomatic response to proton-pump inhibitor (PPI) treatment as an indicator for gastroesophageal reflux disease (GERD).

Searching
MEDLINE (from January 1980 to July 2003) and the Cochrane Controlled Trials Register were searched for studies published in English; the key terms were stated. In addition, the reference lists from relevant reports were checked. The studies had to be published in peer-reviewed journals.

Study selection
Study designs of evaluations included in the review
The included studies were open-label trials, randomised controlled trials (RCTs) and crossover RCTs.

Specific interventions included in the review
Studies that assessed a symptomatic response to a short (1 to 4 weeks) course of normal- or high-dose PPI were eligible for inclusion. The included studies used omeprazole (20 or 40 mg), lansoprazole (30 or 60 mg) or pantoprazole (40 mg). The authors' definition of response was accepted for the review. The included studies used different definitions of treatment response (further details were reported). The included studies assessed treatment response at 5 days to 4 weeks.

Reference standard test against which the new test was compared
The studies had to compare clinical response to PPI with an objective test for GERD (i.e. 24-hour monitoring and/or endoscopy, or a structured symptom scoring system). The included studies used ambulatory 24-hour pH monitoring, upper gastrointestinal endoscopy and symptom scores. An abnormal pH results was defined as a pH of less than 4.0 recorded for more than 4% of a 24-hour period. For endoscopy, abnormal was defined as at least grade I oesophagitis on a commonly used classification system. For symptom scores, the review accepted a diagnosis of GERD when the authors used predetermined clinical criteria or when the symptoms reached an adequate score for GERD. The reviewers stated that symptom scores had not been validated extensively.

Participants included in the review
Studies of adults (aged 18 years or over) with presumptive diagnosis of GERD, based on symptoms and history, were eligible for inclusion. Studies that focused on children or patients with complications caused by GERD, alarm symptoms, extraesophageal symptoms of GERD or suspected cardiovascular disease, were excluded. The studies had to include patients regardless of their symptom characteristics. The included studies were conducted in primary care and specialist care settings, and included patients with the full spectrum of GERD severity. Most of the studies had more male than female participants, and most patients had symptoms for more than 3 months.

Outcomes assessed in the review
The inclusion criteria for the outcomes were not specified.
How were decisions on the relevance of primary studies made?
One reviewer screened studies for inclusion. The final decision on study exclusions was made by consensus with two other reviewers.

Assessment of study quality
The studies were assessed following published criteria (see Other Publications of Related Interest). These criteria focused on: clarity of definition for normal and abnormal results according to the reference test; the inclusion of patients with and without GERD according to the reference standard; whether the method used to perform the reference test was described adequately; whether the interpretation of the test results was adequate; whether the results were reported sufficiently to allow the study to be replicated; and the consistency of outcome measures and objective measures of GERD with accepted standards. The authors did not state who performed the validity assessment.

Data extraction
Two reviewers independently extracted the data and reached consensus through discussion. Data were extracted separately for relevant treatment arms in RCTs and crossover studies (data were only extracted for the period before the crossover). For each study, the sensitivity, specificity, positive and negative predictive values, and likelihood ratios were calculated from the original data.

Methods of synthesis
How were the studies combined?
Studies that used comparable definitions for symptomatic response and the reference test, and also provided adequate data, were combined using a meta-analysis. Separate summary receiver-operating characteristic curves were created for studies using 24-hour monitoring and studies using endoscopy as the reference test. Pooled sensitivity and specificity were calculated, together with their respective 95% confidence intervals (CIs), using a random-effects model (further details of the methods used were given). Studies that used a symptom score for selecting patients with GERD were not included in the meta-analyses.

How were differences between studies investigated?
Diagnostic accuracy values for individual studies were tabulated. The studies were pooled separately according to the reference test (see above). The authors stated that there were insufficient data to assess the effect on diagnostic accuracy of different doses and duration of treatment with PPI.

Results of the review
Fifteen studies (n=2,793) were included.

Most studies adequately described the reference standard test and presented adequate data (11 studies).

Standard reference test 24-hour pH monitoring.

The positive likelihood ratios ranged from 1.64 to 1.84, as reported in the tables (1.63 to 1.87 in the abstract). The sensitivity ranged from 0.69 (corresponding specificity 0.58) to 1.00 (corresponding specificity 0.43) based on 5 studies. The pooled sensitivity was 0.78 (95% CI: 0.66, 0.86) and the pooled specificity was 0.54 (95% CI: 0.44, 0.65).

Standard reference test upper endoscopy.

Values for diagnostic accuracy were lower than for studies using 24-hour pH monitoring as the reference test. The positive likelihood ratios ranged from 1.01 to 1.66. The sensitivity ranged from 0.48 (corresponding specificity 0.71) to 0.79 (corresponding specificity 0.48), based on 6 studies. The pooled sensitivity was 0.68 (95% CI: 0.56, 0.79) and the pooled specificity was 0.46 (95% CI: 0.34, 0.59).

Authors' conclusions
The response to short-term treatment with PPI did not accurately diagnose GERD as defined by traditional objective criteria.

**CRD commentary**

The review question was clear in terms of the intervention, participants, test being evaluated and reference test. The inclusion criteria were not explicit in terms of the study design or outcome. Only two databases were searched and this might have resulted in the omission of other relevant studies. No attempts were made to avoid language bias and unpublished studies were excluded. Two reviewers made final decisions on study exclusions and two reviewers independently extracted the data, thus reducing the potential for bias and errors. Validity was assessed using defined criteria but the blinding of assessors of tests was not reported.

Some relevant information on the included studies was tabulated, and the authors calculated diagnostic accuracy values from primary studies. Studies using comparable tests were combined in a meta-analysis. Diagnostic accuracy appeared to vary among studies, but two potential sources of differences (PPI dose and treatment duration) could not be explored because of insufficient data. The influence of other potential differences on diagnostic accuracy (such as GERD severity, clinical setting) were not explored. The authors' conclusions about the lack of usefulness of response to PPI in diagnosing GERD are likely to be reliable.

Four of the authors had either held shares, given expert testimony, or received grants from at least one of the following companies: Merck, Pfizer, Janssen Cilag, AstraZeneca, Altana Pharma and Novartis.

**Implications of the review for practice and research**

Practice: The authors stated that until better methods of diagnosing GERD are available, treatment for patients should be based on the individual clinical setting, the response to therapy and careful diagnostic testing.

Research: The authors did not state any implications for further research.

**Bibliographic details**


PubMedID 15068979

Original Paper URL http://www.annals.org/cgi/content/full/140/7/518

**Other publications of related interest**


**Indexing Status**

Subject indexing assigned by NLM

MeSH

Adult; Endoscopy, Gastrointestinal; Esophagitis, Peptic /etiology; Female; Gastroesophageal Reflux /complications /diagnosis /drug therapy; Heartburn /etiology; Humans; Hydrogen-Ion Concentration; Male; Middle Aged; Monitoring, Physiologic /methods; Predictive Value of Tests; Proton Pump Inhibitors; ROC Curve

AccessionNumber
Record Status
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.