

# Is Proton Pump Inhibitor Testing an Effective Approach to Diagnose Gastroesophageal Reflux Disease in Patients With Noncardiac Chest Pain?

## A Meta-analysis

Wei Hong Wang, MD; Jia Qing Huang, PhD; Ge Fan Zheng, MD; Wai Man Wong, MD; Shiu Kum Lam, MD; Johan Karlberg, MD; Harry H. X. Xia, PhD; Ronnie Fass, MD; Benjamin C. Y. Wong, MD

**Background:** Gastroesophageal reflux disease (GERD) is common in patients with noncardiac chest pain (NCCP). Results of studies evaluating the accuracy of a proton pump inhibitor (PPI) treatment as a diagnostic test for GERD-related NCCP have varied. We evaluated the overall accuracy of this modality.

**Methods:** We searched the PubMed, MEDLINE, EMBASE, CINAHL, and Cochrane databases to May 2004 and included randomized, placebo-controlled studies evaluating the accuracy of findings from PPI testing in the diagnosis of GERD in patients with NCCP. The GERD diagnosis was confirmed by results of endoscopy and/or 24-hour esophageal pH monitoring. A summary diagnostic odds ratio and summary receiver operating characteristic curve analysis were used to estimate the overall accuracy and to explore any contributing factors.

**Results:** Six studies met the inclusion criteria. The over-

all sensitivity and specificity of a PPI test were 80% (95% confidence interval [CI], 71%-87%) and 74% (95% CI, 64%-83%), respectively, compared with 19% (95% CI, 12%-29%) and 77% (95% CI, 62%-87%), respectively, in the placebo group. The PPI test showed a significant higher discriminative power, with a summary diagnostic odds ratio of 19.35 (95% CI, 8.54-43.84) compared with 0.61 (95% CI, 0.20-1.86) in the placebo group. The impact of the prevalence of GERD and treatment duration on the accuracy of the test could not be determined because of the lack of an adequate number of studies.

**Conclusion:** The use of PPI treatment as a diagnostic test for detecting GERD in patients with NCCP has an acceptable sensitivity and specificity and could be used as an initial approach by primary care physicians to detect GERD in selected patients with NCCP.

*Arch Intern Med.* 2005;165:1222-1228

### Author Affiliations:

Department of Gastroenterology, Peking University First Hospital, Beijing, China (Dr Wang); Department of Medicine (Drs Huang, Zheng, W. M. Wong, Lam, Xia, and B. C. Y. Wong) and Clinical Trials Centre (Drs Huang and Karlberg), Faculty of Medicine, University of Hong Kong, Hong Kong; and Neuroenteric Clinical Research Group, Section of Gastroenterology, Southern Arizona Veterans Affairs Health Care System and University of Arizona Health Sciences Center, Tucson (Dr Fass).

**Financial Disclosure:** Dr Huang has received research funding and honoraria from AstraZeneca and Merck.

**R**ECURRENT EPISODES OF retrosternal pain lacking documented cardiac abnormalities are defined as noncardiac chest pain (NCCP).<sup>1</sup> The annual prevalence of NCCP in the general population of the western world ranges from approximately 25% to 35%.<sup>2,3</sup> Patients with NCCP have a poor quality of life<sup>4</sup> and consume a large proportion of health care resources.<sup>5</sup> Although NCCP may have a number of causes, gastroesophageal reflux disease (GERD) is considered to be the most common.<sup>1,6-12</sup>

Several esophageal tests are used to evaluate NCCP, including endoscopy, 24-hour esophageal pH monitoring, esophageal manometry, or provocative tests.<sup>1,8,13</sup> However, the sensitivity of endoscopy is limited because most patients with NCCP have nonerosive GERD.<sup>14</sup> Although 24-hour esophageal pH monitoring has been considered the best modality for diagnosing GERD in patients with NCCP,<sup>1</sup> it is invasive,

and often unavailable in the primary care setting.

The knowledge that GERD is the most common cause of NCCP has led investigators to treat patients empirically with proton pump inhibitors (PPIs).<sup>15-18</sup> The successful management of NCCP with PPIs has proven the causal relationship between GERD and NCCP.<sup>15-18</sup> Treatment with PPIs is now used by investigators as a diagnostic test for identifying GERD in patients with NCCP.<sup>19</sup> Studies assessing the accuracy of the PPI test result demonstrated sensitivity ranging from 78% to 92% and specificity from 67% to 86% for diagnosis of GERD in patients with NCCP.<sup>20-22</sup> However, these studies generally have a small number of patients,<sup>20-22</sup> and the potential impact of demographics of patients has not been evaluated. Thus, the clinical value of the PPI test in evaluating NCCP remains to be thoroughly assessed.<sup>23</sup>

We undertook this systematic review and meta-analysis to evaluate the overall

sensitivity and specificity of the PPI test for detecting GERD in patients with NCCP. We also examined the impact of the characteristics of the study population or the study design on study findings.

## METHODS

### LITERATURE SEARCH

A computerized literature search was performed in the PubMed, MEDLINE, EMBASE, CINAHL, and Cochrane Controlled Trials Register databases for relevant articles published in any language between 1966 and May 2004 with the following medical subject heading terms and/or text words: *chest pain, noncardiac, or noncardiac* in combination with *omeprazole, lansoprazole, pantoprazole, rabeprazole, or esomeprazole*. Meeting abstracts were searched from CD-ROMs of major international gastroenterological meetings held from 1995 through 2003 (American Digestive Disease Week, American College of Gastroenterology, World Congress of Gastroenterology, and United European Gastroenterology Week) using the same terms. Finally, we manually searched the reference list of all relevant review articles and original studies that we retrieved.

The title and abstract of all potentially relevant studies were screened for relevance before the retrieval of the full articles. Full articles were also scrutinized for relevance if the title and abstract were ambiguous. The literature search was conducted independently by 3 reviewers (W.H.W., J.Q.H., and G.F.Z.).

### INCLUSION AND EXCLUSION CRITERIA

The following criteria were used to include studies: (1) adult patients with recurrent episodes of chest pain without documented cardiac abnormalities; (2) GERD diagnosed by results of endoscopy and/or 24-hour esophageal pH monitoring; (3) only randomized, placebo-controlled trials because a symptomatic response to PPI treatment in patients was evaluated as a diagnostic test; and (4) the number of true-positive, false-positive, true-negative, and false-negative findings were described explicitly, or such numbers could be derived from studies.

We excluded (1) therapeutic trials evaluating the efficacy of PPI treatment in patients with GERD-related NCCP, (2) studies without raw data for re-

trieval, and (3) duplicate publications. When duplications were found, we only included the publication that reported the most extensive information.

### DATA EXTRACTION

Data were extracted independently from each study by 3 researchers (W.H.W., J.Q.H., and G.F.Z.) using a predefined review spreadsheet. Any disagreements between the reviewers were resolved by discussion to reach consensus.

### ASSESSMENT OF STUDY QUALITY

Criteria modified by Irwig et al<sup>24</sup> on behalf of the Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests were applied to assess the quality of each study. These criteria include an explicit statement of the spectrum of disease, a clear definition of NCCP, a reference test used, a blinded measurement of the PPI and reference tests, execution of the test, explicit definition of the improvement of symptoms and reporting of the cutoff point of the test, and a description of the demographic information and sampling strategy.

### STATISTICAL ANALYSIS AND ASSESSMENT OF HOMOGENEITY

For each study, we calculated sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and their 95% confidence intervals (CIs). We also reported the diagnostic odds ratio (DOR) as a measure for the discriminative power of a diagnostic test for individual studies.

We used 2 methods to summarize the data. First, statistical pooling of the sensitivities and specificities was performed and summary DORs were calculated under a random-effects model.<sup>25</sup> As a complementary method, summary receiver operating characteristic (sROC) curves were plotted with sensitivity (true-positive rate) on the y-axis and 1-specificity (false-positive rate) on the x-axis according to the method proposed by Moses and Shapiro<sup>26</sup> and refined by Littenberg and Moses.<sup>27</sup> We used this approach because sensitivity and specificity are measures of diagnostic accuracy that rely on a single threshold for classifying a test result as positive or negative. For the PPI test, the threshold effect may find its origin in the variation of setting, study design, definition of NCCP, medications

and the dosages and duration used, washout period, and the definition of symptom improvement. Studies were combined by using the sROC method if the definition of symptom improvement and reference test used were comparable, and the natural logarithms (Ln) of DOR of the included studies were homogeneous. The use of the sROC curve that combines results from different studies allows simultaneous evaluation of sensitivity and specificity and facilitates the comparison of the accuracy of the results from PPI and placebo tests. The regression curves were extended only over the range of the data for the few studies included in the analysis. Areas under the sROC curves were also calculated directly under the curve where data existed. We performed analyses that were unweighted and weighted by the inverse of the variance. The significance of the difference between the PPI and placebo was statistically analyzed by applying a Wilcoxon paired-sample test to the parameter D ( $D = \ln OR$ , where the OR equals  $[Sensitivity / (1 - Sensitivity)] / [(1 - Specificity) / Specificity]$ ). The influence of covariates on the accuracy of the test was determined by the ROC regression analysis.<sup>26</sup>

We first used a  $\chi^2$  test to assess the statistical homogeneity between studies.<sup>28</sup> Then we plotted the sensitivity, specificity, and DOR of individual studies and their 95% CIs to evaluate study variations.<sup>28</sup> If a visual heterogeneity was identified, we searched for the sources of any possible clinically important heterogeneity.

Data were reported according to the guidelines for meta-analysis evaluating diagnostic tests.<sup>28,29</sup> Analyses were performed using MetaTest software (version 0.6) written by Joseph Lau, MD, and specially designed for meta-analysis of diagnostic tests.

## RESULTS

### LITERATURE SEARCH AND STUDY SELECTION

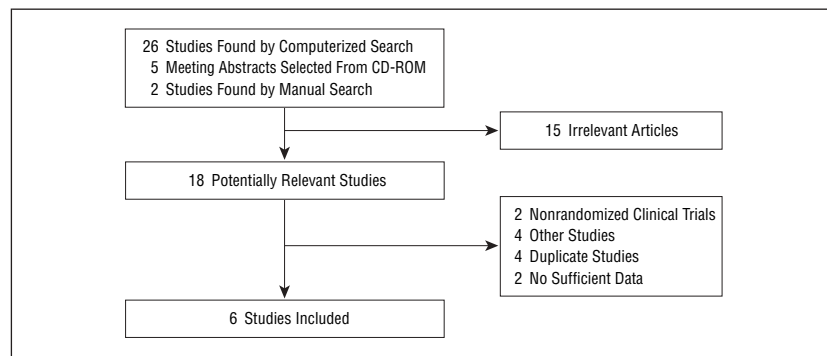
We identified a total of 33 reports, of which only 6 studies met the inclusion criteria<sup>20-22,30-32</sup> (**Table 1**). Fifteen irrelevant articles were deleted after screening the titles and abstracts. Full articles of the remaining potentially relevant articles were further scrutinized. Of these, 12 were excluded for the following reasons: 2 were nonrandomized, placebo-controlled clinical trials<sup>17,33</sup>; 2 were cost-effective analyses of the PPI

**Table 1. Characteristics of Studies Included in Meta-analysis\***

Source	Study Design	Definition of NCCP	Diagnosis of GERD	Exclusion Criteria	Comments
Xia et al, <sup>20</sup> 2003	Randomized, single-blind, placebo-controlled	>12 wk in preceding 12 mo by means of angiography	24-h pH monitoring (any of the 6 variables)	PUD, esophagitis, GERD/dyspepsia symptom	Evaluate negative endoscopy findings in NCCP
Bautista et al, <sup>31</sup> 2004	Randomized, double-blind, placebo-controlled crossover	>3 Episodes/wk for >3 mo, by means of angiography or treadmill test, thallium chloride, technetium Tc 99m, or MIBI testing	EGD esophagitis or 24-h pH monitoring (1 variable)	Patients with other diseases, previous antireflux therapy, PUD, gastric surgery	Washout time, 14 d
Fass et al, <sup>21</sup> 1998	Crossover randomized, double-blind, placebo-controlled	>3 Episodes/wk for >3 mo, by means of angiography or treadmill, thallium, technetium Tc 99m, or MIBI testing	EGD esophagitis or 24-h pH monitoring (1 variable)	PUD, gastric surgery after treatment	Washout time, 14 d
Pandak et al, <sup>22</sup> 2002	Crossover randomized, double-blind, placebo-controlled	>3 Episodes/wk for > 6 mo, by means of technetium Tc 99m MIBI testing	EGD esophagitis or 24-h pH monitoring (1 variable)	PUD, surgery after treatment, or abnormal findings on radiographs or from the physical examination	Washout time, 21 d
Fass et al, <sup>30</sup> 2002	Randomized, double-blind, placebo-controlled crossover	>3 Episodes/wk by means of comprehensive evaluation	EGD esophagitis or 24-h pH monitoring (1 variable)	Unknown	Washout time, 7 d
Squillace et al, <sup>32</sup> 1993	Crossover randomized, double-blind, placebo-controlled	Episode unknown by means of angiographic findings, thallium stress test	EGD esophagitis or 24-h pH monitoring with symptom index	Unknown	Washout time, 5 d

Abbreviations: EGD, endoscopy; GERD, gastroesophageal reflux disease; MIBI, methoxy isobutyl isonitrile; NCCP, noncardiac chest pain; PUD, peptic ulcer disease.

\*Improvement was defined as greater than 50% improvement in all 6 studies.



**Figure 1.** Flow diagram of meta-analysis.

test<sup>34,35</sup>; 2 were comments on a single study<sup>36,37</sup>; 4 presented duplicated data<sup>18,38-40</sup>; 1 included only patients with NCCP and GERD<sup>15</sup>; and 1 did not compare the outcomes with a reference test<sup>41</sup> (**Figure 1**).

### STUDY QUALITY

Of the 6 studies, 5 were double-blind<sup>21,22,30-32</sup> and 1 was single-blind.<sup>20</sup> Four were full-text articles,<sup>20-22,31</sup> and 2 were abstracts.<sup>30,32</sup>

A total of 220 study subjects with NCCP were included in these studies. The episodes and the duration of chest pain of patients at baseline were clearly described in 4 studies.<sup>21,22,30,31</sup> Diagnosis of NCCP was based on a normal finding on a cardiac angiogram or in other comprehensive cardiac evaluations in 4 studies<sup>20,21,31,32</sup> and on a negative result on technetium Tc 99m methoxy isobutyl isonitrile testing in 1 study.<sup>22</sup> The diagnosis of NCCP was

not clearly described in 1 study.<sup>32</sup> Four studies claimed to exclude patients with peptic ulcer, history of gastric surgery, or recent treatment with antireflux medications.<sup>20-22,31</sup> The NCCP patients with endoscopic esophagitis were excluded in 1 study.<sup>20</sup> Five studies provided a clear description of the demographic information,<sup>20-22,30,31</sup> with the mean  $\pm$  SD age being 54.4  $\pm$  6.1 years and the percentage of men, 60.4%. Five studies were performed in a crossover fashion with a washout period of 5 to 21 days.<sup>21,22,30-32</sup>

Three studies evaluated omeprazole (60-80 mg/d)<sup>21,22,32</sup>; 2, lansoprazole (30-90 mg/d)<sup>20,31</sup>; and 1, rabeprazole (40 mg/d)<sup>30</sup> as a diagnostic test for detecting GERD in patients with NCCP. Of the 6 studies, 5 assessed the value of a short course (1-2 weeks) of high-dosage PPI treatment,<sup>21,22,30-32</sup> whereas 1 used a standard dosage (30 mg/d) of lansoprazole for 4 weeks.<sup>20</sup> A positive test result was defined as an improve-

**Table 2. Main Results of PPI Test in the Individual Studies**

Source	No. of Patients	Drug Therapy (Dosage/Duration)	Sensitivity (95% CI), %	Specificity (95% CI), %	Prior Probability, %	PPV (95% CI)	NPV (95% CI)	DOR (95% CI)
Xia et al, <sup>20</sup> 2003	36	Lansoprazole (30 mg OD/4 wk)	92 (60-100)	67 (45-83)	33	58 (36-77)	94 (73-99)	17.73 (2.33-134.89)
Fass et al, <sup>21</sup> 1998	37	Omeprazole (40 mg OD and 20 QN/1 wk)	78 (56-92)	86 (56-97)	62	90 (70-97)	71 (47-87)	18.93 (3.38-105.87)
Pandak et al, <sup>22</sup> 2002	38	Omeprazole (40 mg BD/2 wk)	90 (67-98)	67 (41-86)	53	75 (55-88)	86 (60-96)	15.90 (2.94-85.93)
Fass et al, <sup>30</sup> 2002	20	Rabeprazole (20 mg BD/7 d)	83 (51-97)	75 (36-95)	60	83 (55-95)	75 (41-93)	12.65 (1.55-103.23)
Bautista et al, <sup>31</sup> 2004	40	Lansoprazole (60 mg OD and 30 mg QN/7 d)	78 (52-93)	91 (70-98)	45	88 (61-98)	83 (62-94)	35.00 (5.62-218.11)
Squillace et al, <sup>32</sup> 1993	17	Omeprazole (80 mg OD/unknown)	69 (39-90)	75 (22-99)	76	90 (60-98)	43 (16-75)	6.75 (0.53-86.61)

Abbreviations: BD, twice daily; CI, confidence interval; DOR, diagnostic odds ratio; NPV, negative predictive value; OD, once in the morning; QN, once at night; PPI, proton pump inhibitor; PPV, positive predictive value.

**Table 3. Main Results of Placebo as a Diagnostic Test in Individual Studies**

Source	No. of Patients	Duration	Sensitivity (95% CI), %	Specificity (95% CI), %	Prior Probability	PPV (95% CI)	NPV (95% CI)	DOR (95% CI)
Xia et al, <sup>20</sup> 2003	32	4 wk	33 (11-64)	65 (41-84)	38	36 (15-65)	62 (41-79)	0.94 (0.21-4.14)
Fass et al, <sup>21</sup> 1998	37	1 wk	22 (8-44)	93 (64-100)	62	83 (44-97)	42 (26-59)	3.05 (0.38-24.30)
Pandak et al, <sup>22</sup> 2002	38	2 wk	5 (0-27)	83 (58-96)	53	25 (0-70)	44 (29-61)	0.31 (0.04-2.67)
Fass et al, <sup>30</sup> 2002	20	7 d	8 (0-40)	100 (63-100)	60	100 (21-100)	42 (23-64)	3.67 (0.05-295.09)
Bautista et al, <sup>31</sup> 2004	40	7 d	22 (7-48)	64 (41-81)	45	33 (11-64)	50 (31-69)	0.50 (0.12-2.05)
Squillace et al, <sup>32</sup> 1993	17	Unknown	8 (0-38)	100 (40-100)	76	100 (21-100)	25 (10-50)	1.74 (0.02-146.25)

Abbreviations: CI, confidence interval; DOR, diagnostic odds ratio; NPV, negative predictive value; PPV, positive predictive value.

**Table 4. Results of Meta-analysis Under a Random-Effects Model**

Diagnostic Test	Sensitivity (95% CI), %	Specificity (95% CI), %	sDOR (95% CI)	AUC	
				Unweighted	Weighted
PPI	80 (71-87)	74 (64-83)	19.35 (8.54-43.84)	0.197	0.200
Placebo	19 (12-29)	77 (62-87)	0.61 (0.20-1.86)	0.062	0.059

Abbreviations: AUC, area under the summary receiver operating characteristic curve; CI, confidence interval; PPI, proton pump inhibitor; sDOR, summary diagnostic odds ratio.

ment of chest pain by more than 50% after the treatment with a PPI. Three studies stated that the assessment of symptom improvement was independent of the results of the reference test.<sup>20-22</sup> The main characteristics of these studies are listed in Table 1.

#### RESULTS OF INDIVIDUAL STUDIES

The main results of the 6 studies are summarized in **Table 2** and **Table 3**. The sensitivities of PPI test results in all 6 studies were significantly higher than that of placebo.

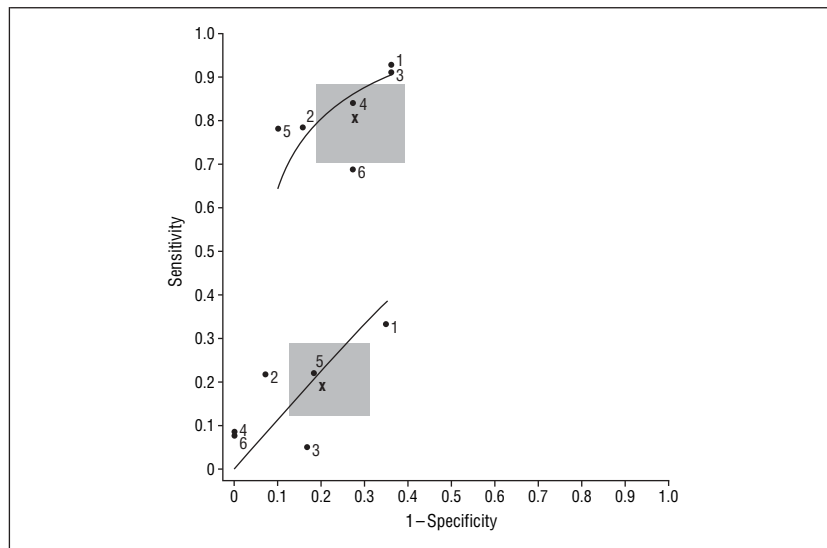
However, the specificities were comparable or higher in 2 studies.<sup>20,31</sup> In the remaining 4 studies, the specificities in the PPI group were lower than that in the placebo group.<sup>21,22,30,32</sup> The PPV and NPV were higher in the PPI-treated group than in the placebo group in 4 studies,<sup>20-22,31</sup> except for 2 studies that included a small number of patients.<sup>30,32</sup> According to the DORs of the individual studies, 5 studies demonstrated that the PPI treatment had a significantly high discriminative power for diagnosing GERD in patients with NCCP.<sup>20-22,30,31</sup> However, this was not observed in the placebo group (Tables 2 and 3).

#### STATISTICAL POOLING

The overall sensitivity for a PPI diagnostic test was 80% (95% CI, 71%-87%) compared with 19% (95% CI, 12%-29%) in the placebo group. The summary specificity for the PPI test was 74% (95% CI, 64%-83%) compared with 77% (95% CI, 62%-87%) in the placebo group (**Table 4**). The PPI test had a significantly higher discriminative power for diagnosing GERD in patients with NCCP, with an estimated DOR of 19.35 (95% CI, 8.54-43.84) compared with 0.61 (95% CI, 0.20-1.86) for the placebo group ( $P = .03$ ).

**Figure 2** shows the sROC curves plotted by using the results of accuracy from the individual studies. The shape of the sROC curve for the PPI group is different from that for the placebo group, with the PPI having a sharper increase in the sensitivity for a given increase in the false-positive rate (1 - specificity).

Incorporation of the additional covariates into the ROC regression



**Figure 2.** Summary receiver operating characteristic curves of proton pump inhibitors (PPIs) (A) and placebo (B) as a diagnostic test in all studies. The curves suggest the superiority of PPI compared with placebo in detecting gastroesophageal reflux disease in patients with noncardiac chest pain. Numbers alongside the plots indicate individual studies (1-6 represent Xia et al,<sup>20</sup> Fass et al,<sup>21</sup> Pandak et al,<sup>22</sup> Fass et al,<sup>30</sup> Bautista et al,<sup>31</sup> and Squillante et al,<sup>32</sup> respectively). The X's mark the pooled estimate of the true-positive and false-positive rates. The shaded regions mark the zone of 95% confidence intervals of the pooled sensitivity and specificity.

analysis indicated that variables such as the type and year of publication, study design, prevalence of NCCP, reference test used, and treatment duration had no significant effect on the test accuracy estimated.

### EVALUATION OF HOMOGENEITY

No statistical heterogeneities between studies were found ( $P = .95$  for PPI and  $P = .17$  for placebo). This was confirmed using a graphic presentation where 95% CIs of sensitivity, specificity, and DOR of individual studies overlap considerably (**Figure 3**).

### COMMENT

In the present meta-analysis, we found that the overall sensitivity was 80% (95% CI, 71%-87%) and the specificity was 74% (95% CI, 64%-83%) for the PPI test. We also found a significantly higher discriminative power associated with the PPI test, with an estimated DOR of 19.35 (95% CI, 8.54-43.84). The PPV and NPV should give the most useful information of a diagnostic test because they mimic the situation in which the test is used. However, the predictive value suffers a disadvantage because the calculation is closely

related to the prevalence of the disease in the study population.<sup>42</sup> In the present analysis, the prevalence of GERD in patients with NCCP ranges from 33% to 76%.<sup>20-22,30-32</sup> The large variability in the prevalence has made the estimate of the overall PPV and NPV unreliable. Therefore, the data should be explained with caution.

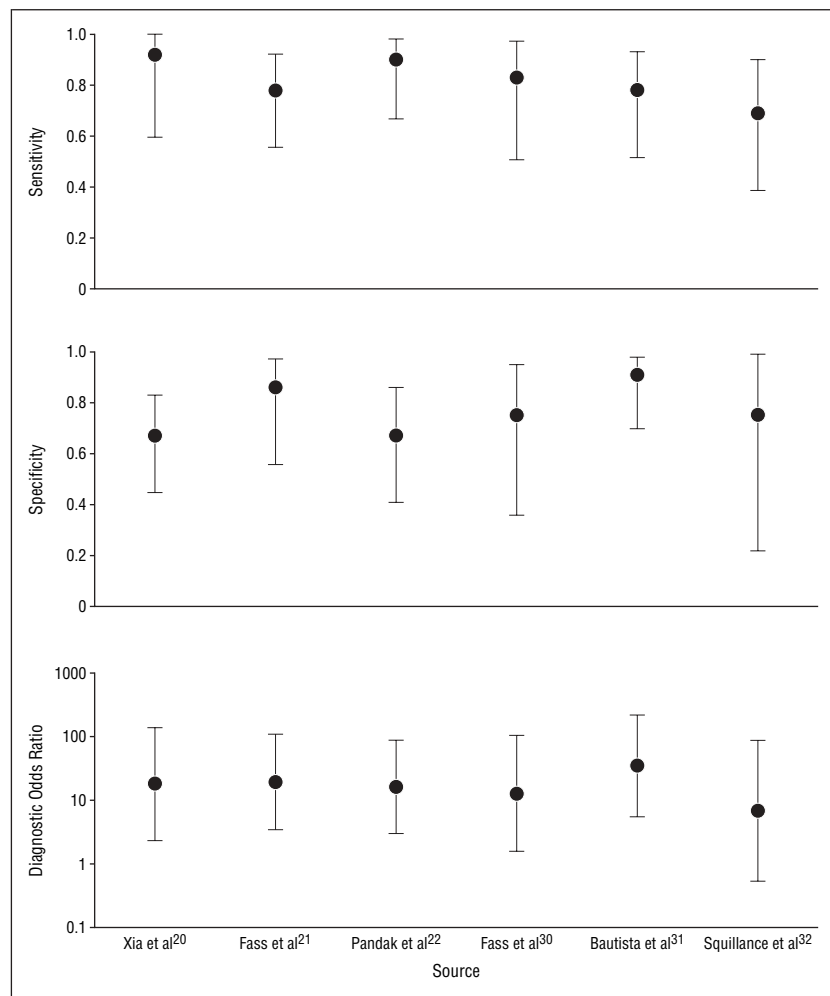
Our results show that the sensitivity of the PPI test was significantly higher than that for placebo, whereas the specificity was almost the same between both groups. Treatment with PPIs and placebo showed similar and better effects on improving NCCP symptoms in patients without GERD, indicating a possible placebo effect. The considerably high placebo effect is not uncommon in patients with functional bowel disorders and not surprising in patients with non-GERD-related NCCP. Nevertheless, interpretation of the study results must be cautioned because of the observed high placebo effect.

The accuracy of a diagnostic test should be evaluated by comparing its results with a gold (reference) standard that has been validated. However, this is not available for the diagnosis of GERD. Endoscopy results are frequently normal in patients with symptoms of GERD and

abnormal esophageal acid exposure.<sup>43</sup> The sensitivity of symptom evaluation falls short of a gold standard.<sup>44</sup> Ambulatory 24-hour esophageal pH monitoring is generally considered to provide the most objective measurement of pathologic reflux. However, the sensitivity is reported to range from 85% to 90%.<sup>45</sup> The sensitivity and specificity can be increased if reflux symptoms are also evaluated.<sup>46</sup> Nevertheless, 24-hour esophageal pH monitoring alone is insufficient to be considered a gold standard. In most of the studies we have included, a combination of endoscopy and pH testing was used, which is the closest to the accepted reference test for GERD. Thus, findings in some patients with GERD may have been classified as negative for GERD. In addition, physiological acid reflux can also induce chest pain in individuals in whom the esophagus is hypersensitive to gastric acid.<sup>47-49</sup> These patients, although considered to have false-positive findings, may respond to the PPI test.

The sensitivity of the PPI test seems to be related to the duration of the treatment. Extending the duration of treatment from 1 to 2 or to 4 weeks increases the sensitivity by approximately 10% (Table 2). However, extending treatment duration beyond 4 weeks was unnecessary because 80% of patients who were likely to respond to PPI would respond within 4 weeks (Table 4). Therefore, we propose that initial treatment with a PPI can be given up to 4 weeks at least twice a day, based on the patient's frequency of symptoms. The degree of relief expected would be at least 50%.

Although no statistical heterogeneity between studies was found, we cannot rule out the possibility of absence of any between-study heterogeneities. For example, there are differences in the definition of NCCP, type of PPIs and dosages used, wash-out period, degree of blinding, execution of test, and reference standard for diagnosing GERD. Therefore, certain biases may exist and could threaten the validity of our conclusions. For example, although all studies used endoscopic esophagitis and/or abnormal 24-hour esophageal pH monitoring as



**Figure 3.** Point estimates with confidence intervals of sensitivity, specificity, and diagnostic odds ratios of 6 studies on the validity of the proton pump inhibitor test for the diagnosis of gastroesophageal reflux disease in patients with noncardiac chest pain. No significant outliers were found.

the reference test, different variables among studies were used to diagnose GERD. One study excluded patients with esophagitis and used abnormal findings of 24-hour esophageal pH monitoring as the reference standard.<sup>20</sup> Another study used endoscopic esophagitis or abnormal findings of 24-hour pH monitoring with symptoms index as the reference standard.<sup>32</sup> Therefore, a differential verification bias among studies may exist.<sup>50</sup>

There are several limitations to our study. First, only 6 randomized, placebo-controlled trials were included in the final analysis, with a total of 220 patients. Therefore, the results are susceptible to possible selection bias.<sup>28</sup> The small number of patients may also limit the ease of generalizability of the study findings to other populations. Second, subgroup analysis was not feasible

to perform in the present study because the number of studies was too small to obtain reliable estimates. We therefore cannot determine any possible influence of factors such as the type and year of publication, prevalence of NCCP, study design, reference test, and dose and duration of PPI treatment on the results of this analysis. Third, the quality of the meeting abstracts was a concern because the information available from these studies was limited.<sup>30,32</sup> As a consequence, a possible verification bias may exist.<sup>24,28</sup>

## CONCLUSIONS

Our meta-analysis suggests that testing with the high-dosage PPI treatment up to 4 weeks has an acceptable sensitivity and specificity and could be considered as an initial use-

ful and possibly cost-saving strategy by primary care physicians in managing patients with NCCP suspected of esophageal disorders and with no alarming symptoms. If more than 50% reduction in symptom scores can be achieved, the chance of having GERD-related NCCP is significantly increased, and the PPI treatment should be continued. Future well-designed, adequately powered studies are needed to ascertain the findings of this analysis.

**Accepted for Publication:** January 31, 2005.

**Correspondence:** Benjamin C. Y. Wong, MD, Department of Medicine, Faculty of Medicine, University of Hong Kong, Hong Kong (bcywong@hku.hk).

**Funding/Support:** The study was supported by the Gastroenterological Research Fund, University of Hong Kong, Hong Kong. Dr Wang is a visiting professor under the Croucher Foundation Chinese Visitorship Scheme, Hong Kong, at the Department of Medicine, University of Hong Kong.

## REFERENCES

1. Fang J, Bjorkman D. A critical approach to noncardiac chest pain: pathophysiology, diagnosis, and treatment. *Am J Gastroenterol.* 2001;96:958-968.
2. Locke GR III, Talley NJ, Fett SL, Zinsmeister AR, Melton LJ III. Prevalence and clinical spectrum of gastroesophageal reflux: a population-based study in Olmsted County, Minnesota. *Gastroenterology.* 1997;112:1448-1456.
3. Eslick GD, Talley NJ, Young LJ, Jones MP. Noncardiac chest pain in the population: prevalence and impact [abstract]. *J Gastroenterol Hepatol.* 1999;14:A179.
4. Rose S, Achkar E, Easley KA. Follow-up of patients with noncardiac chest pain: value of esophageal testing. *Dig Dis Sci.* 1994;39:2063-2068.
5. Ockene IS, Shay MJ, Alpert JS, Weiner BH, Dalen JE. Unexplained chest pain in patients with normal coronary arteriograms: a follow-up study of functional status. *N Engl J Med.* 1980;303:1249-1252.
6. Fass R. Chest pain of esophageal origin. *Curr Opin Gastroenterol.* 2002;18:464-470.
7. Lam HG, Dekker W, Kan G, Breddijk M, Smout AJ. Acute noncardiac chest pain in coronary care unit. *Gastroenterology.* 1992;102:453-460.
8. Fass R, Winters GF. Evaluation of the patients with noncardiac chest pain: is gastroesophageal reflux disease or an esophageal motility disorder the cause? *Medscape Gastroenterol.* 2001;3:1-10.
9. Ergun GA, Kahrilas PJ. Clinical applications of esophageal manometry and pH monitoring. *Am J Gastroenterol.* 1996;91:1077-1089.
10. Soffer EE, Scalabrini P, Wingate DL. Spontaneous noncardiac chest pain: value of ambulatory

- esophageal pH and motility monitoring. *Dig Dis Sci*. 1989;34:1651-1655.
11. DeMeester TR, O'Sullivan GC, Bermudez G, Midell AI, Cimochoowski GE, O'Drobinak J. Esophageal function in patients with angina-type chest pain and normal coronary angiograms. *Ann Surg*. 1982; 196:488-498.
  12. DeVault KR, Achem SR. Noncardiac chest pain of esophageal origin. *Curr Treat Options Gastroenterol*. 1998;1:49-55.
  13. Stahl WG, Beton RR, Johnson CS, Brown CL, Waring JP. Diagnosis and treatment of patients with gastroesophageal reflux and noncardiac chest pain. *South Med J*. 1994;87:739-742.
  14. Hsia PC, Maher KA, Lewis JH, Cattau EL Jr, Fleischer DE, Benjamin SB. Utility of upper endoscopy in the evaluation of noncardiac chest pain. *Gastrointest Endosc*. 1991;37:22-26.
  15. Achem SR, Kolts BE, MacMath T, et al. Effects of omeprazole versus placebo in treatment of noncardiac chest pain and gastroesophageal reflux. *Dig Dis Sci*. 1997;42:2138-2145.
  16. Singh S, Richter JE, Hewson EG, Sinclair JW, Hackshaw BT. The contribution of gastroesophageal reflux to chest pain in patients with coronary artery disease. *Ann Intern Med*. 1992;117:824-830.
  17. Chambers J, Cooke R, Anggiansah A, Owen W. Effect of omeprazole in patients with chest pain and normal coronary anatomy: initial experience. *Int J Cardiol*. 1998;65:51-55.
  18. Fass R, Fennerty MB, Johnson C, Camargo L, Sampliner RE. Correlation of ambulatory 24-hour esophageal pH monitoring results with symptom improvement in patients with noncardiac chest pain due to gastroesophageal reflux disease. *J Clin Gastroenterol*. 1999;28:36-39.
  19. Richter JE. Chest pain and gastroesophageal reflux disease. *J Clin Gastroenterol*. 2000;30 (3 suppl):S39-S41.
  20. Xia HH, Lai KC, Lam SK, et al. Symptomatic response to lansoprazole predicts abnormal acid reflux in endoscopy-negative patients with noncardiac chest pain. *Aliment Pharmacol Ther*. 2003; 17:369-377.
  21. 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. *Gastroenterology*. 1998;115:42-49.
  22. Pandak WM, Arezo S, Everette S, et al. Short course of omeprazole: a better first diagnostic approach to noncardiac chest pain than endoscopy, manometry, or 24-hour esophageal pH monitoring. *J Clin Gastroenterol*. 2002;35:307-314.
  23. Botoman VA. Noncardiac chest pain. *J Clin Gastroenterol*. 2002;34:6-14.
  24. Irwig L, Tosteson ANA, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med*. 1994;120:667-676.
  25. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7:177-188.
  26. Moses LE, Shapiro D. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med*. 1993;12:1293-1316.
  27. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. *Med Decis Making*. 1993; 13:313-321.
  28. Deville WL, Buntinx F, Bouter LM, et al. Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol*. 2002; 2:9-21.
  29. Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests. Recommended methods. Updated December 21, 1998. Available at: <http://www.cochrane.org/cochrane/sadt.htm>. Accessed May 27, 2003.
  30. Fass R, Fullerton H, Hayden CW, Garewal HS. Patients with noncardiac chest pain (NCCP) receiving an empirical trial of high dose rabeprazole, demonstrate early symptom response: a double blind, placebo-controlled trial [abstract]. *Gastroenterology*. 2002;122:W1175.
  31. Bautista J, Fullerton H, Briseno M, Cui H, Fass R. The effect of an empirical trial of high-dose lansoprazole on symptom response of patients with non-cardiac chest pain: a randomized, double-blind, placebo-controlled, crossover trial. *Aliment Pharmacol Ther*. 2004;19:1123-1130.
  32. Squillace SJ, Young MF, Sanowski RA. Single dose omeprazole as a test for noncardiac chest pain [abstract]. *Gastroenterology*. 1993;104:A197.
  33. Achem SR, Kolts BE, Wears R, Burton L, Richter JE. Chest pain associated with nutcracker esophagus: a preliminary study of the role of gastroesophageal reflux. *Am J Gastroenterol*. 1993; 88:187-192.
  34. Ofman JJ, Gralnek IM, Udani J, Fennerty MB, Fass R. The cost-effectiveness of the omeprazole test in patients with noncardiac chest pain. *Am J Med*. 1999;107:219-227.
  35. Borzecki AM, Pedrosa MC, Prashker MJ. Should noncardiac chest pain be treated empirically? a cost-effectiveness analysis. *Arch Intern Med*. 2000; 160:844-852.
  36. Castell DO, Katz PO. The acid suppression test for unexplained chest pain. *Gastroenterology*. 1998; 115:222-224.
  37. Friedel D, Ozick LA. High-dose omeprazole in patients with noncardiac chest pain. *Gastroenterology*. 1999;116:226-227.
  38. Arezo S, Pandak WM, Everette S, Decosta G, Zfass AM. A better diagnostic appreciation of noncardiac chest pain [abstract]. *Gastroenterology*. 2001; 120:A2245.
  39. Fass R, Fennerty MB, Yalam JM, Camargo L, Garewal H, Sampliner RE. Evaluation of the "omeprazole test" in patients with noncardiac chest pain [abstract]. *Gastroenterology*. 1997;112:A114.
  40. Pandak WM, Everette S, Arezo S, et al. Short course of omeprazole: better first diagnostic and economic approach to non-cardiac chest pain (NCCP) than endoscopy, manometry, or 24-hour esophageal pH monitoring [abstract]. *Gastroenterology*. 2000;118:A2629.
  41. Zierer ST, Sanowski RA, Young MF, Squillace SJ. Can a single dose of omeprazole be used to identify gastroesophageal reflux and acid-related chest pain? *Am J Gastroenterol*. 1993;88:1501. Abstract 82.
  42. Diamond GA. Clinical epistemology of sensitivity and specificity. *J Clin Epidemiol*. 1992;45:9-13.
  43. Kahrilas PJ, Quigley EM. Clinical esophageal pH recording: a technical review for practice guideline development. *Gastroenterology*. 1996;110: 1982-1996.
  44. Klausner AG, Schindlbeck NE, Muller-Lissner SA. Symptoms in gastro-oesophageal reflux disease. *Lancet*. 1990;335:205-208.
  45. Johnsson F, Joelsson B, Isberg PE. Ambulatory 24-hour oesophageal pH-monitoring in the diagnosis of gastroesophageal reflux disease. *Gut*. 1987;28:1145-1150.
  46. Smout AJPM. Evaluation and significance of reflux from the duodenum and stomach. *Aliment Pharmacol Ther*. 1997;11(suppl 2):27-32.
  47. Shi G, Bruley des Uarannes S, Scarpignato C, Le Rhun M, Galmich JP. Reflux related symptoms in patients with normal esophageal exposure to acid. *Gut*. 1995;37:457-464.
  48. Rao SS, Gregersen H, Hayek B, Summers RW, Christensen J. Unexplained chest pain: the hypersensitive, hyperactive, and poorly compliant esophagus. *Ann Intern Med*. 1996;124:950-958.
  49. Galmiche JP, Scarpignato C. Oesophageal sensitivity to acid in patients with non-cardiac chest pain: is the oesophagus hypersensitive? *Eur J Gastroenterol Hepatol*. 1995;7:1152-1159.
  50. Whiting P, Rutjes AWS, Reitsma JB, Glas AS, Bossuyt PMM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med*. 2004;140:189-202.