Reviewer's report

Title: Cost-effectiveness of six strategies for Helicobacter pylori diagnosis and management in uninvestigated dyspepsia

Version: 1 Date: 26 July 2010

Reviewer: Xavier Calvet

Reviewer's report:

The study of Holmes et al. attempts to determine the relative cost-effectiveness of different diagnostic and treatment strategies for uninvestigated dyspepsia. The study is methodologically reasonably well performed. I wonder, however, whether the conclusions are either reliable or applicable to other settings. As usual, the risk of bias came from both the model and the assumptions. I believe that additional work is needed to make the results both convincing and useful to the reader. After reading the article I have the following major concerns:

1. Despite the study deals with a low-risk for cancer population (less than 55 y old, no alarm symptoms), the authors assumed that all patients whose symptoms relapsed either after triple therapy or a PPI trial will receive an expensive endoscopy. By contrast, as suggested by Spiegel et al. (Gastro 2002), it seems more reasonable to test for H. pylori and treat the infection in patients failing empirical PPI therapy before endoscopy. It would also be preferable to test patients who had received eradication treatment for cure of H. pylori infection with and UBT before endoscopy. Did these plausible different approaches change the study conclusions?

2. Prevalence of H. pylori infection is given for the US general population. However, this results are applicable neither to other countries (for example, Central and South-American, Mediterranean, African or Asian countries) nor to specific populations in the US (as Afro-American, Hispanic or Asian), that have an Hp prevalence around 60%. Are the conclusions of the study the same with such high prevalence?

3. The same applies for costs, which could markedly change from country to country. How did major differences in the cost of the tests influence the study conclusions?

4. Values of sensitivity and specificity are generous, especially for serology and stool tests. It should be remembered that the different kits show marked differences in sensitivity and specificity. Sensitivity and specificity may be very low depending on the stool kit or even the methodology of the urea breath test (Calvet et al., Clinical Infectious Diseases). In addition, as variability of serology accuracy is even larger, its use is not recommended provided local validation was performed (Maastricht guidelines). What happens in the model when values of sensitivity and specificity for serology are decreased to, for example, 70%?

5. Obtaining the values for assumptions of different studies instead from direct
comparative trials leads to bizarre results. Values coming from different studies heavily depend on the differences in the baseline characteristics of the populations evaluated instead of reflecting true differences of the efficacy of the treatments. So, for example, the study assumes that, after eradication, the probability of symptom relapse at one year is higher in patients with peptic ulcer than in those without. By contrast, in the few available randomized trials, the one-year probability of symptoms relapse in patients with uninvestigated dyspepsia after a PPI trial is near 100% (Rabeneck, AJG, Marmo BMJ) whereas probability of relapse is far lower after eradication treatment in H pylori positive patients (Marmo BMJ). What happens in the model when these different values are incorporated?

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I haveno compting interests