Do commercial serological kits for Helicobacter pylori infection differ in accuracy: a meta-analysis

Loy C T, Irwig L M, Katelaris P H, Talley N J

Authors' objectives
To compare the accuracy of common serological kits for Helicobacter pylori (H. pylori) and to ascertain factors affecting accuracy.

Searching
MEDLINE was searched from 1983 to 1995 for publications in the English language, using the search terms 'campylobacter', 'helicobacter', 'antibodies-bacteria' and 'sensitivity'. In addition, the reference lists of retrieved studies and abstract proceedings of the IV-VII Workshop on Gastroduodenal Pathology and Helicobacter pylori, were searched manually.

Study selection

Study designs of evaluations included in the review
No inclusion criteria relating to the study design were specified. Only studies that compared pairs of kits were included in the analysis.

Specific interventions included in the review
No criteria relating to the specific tests included were specified. The studies included in the review used 12 commercial serological kits that measure immunoglobulin G response to H. pylori by enzyme-linked immunosorbent assay or latex agglutination. These 12 kits were Biometra ECP, Biorad-GAP, Cobas-1G, Cobas-2G, Helico-G, HELp, Malakit, Premier, Pyloriset Latex, Pylori-Stat, Quickvue and Roche-MTP.

Reference standard test against which the new test was compared
No inclusion criteria relating to the reference standard were specified. The reference standards used by studies in the review included culture alone, culture and histology, and culture, histology and urease testing on biopsy. Three studies included in the review used unspecified 'other' reference standards.

Participants included in the review
No inclusion criteria relating to the characteristics of the participants were specified. No details of the participants in the included studies were reported.

Outcomes assessed in the review
No inclusion criteria relating to the outcome measures were specified. The outcome measures in the review were the sensitivity and specificity of commercial serological kits for the diagnosis of H. pylori infection.

How were decisions on the relevance of primary studies made?
The authors did not state how the papers were selected for the review, or how many reviewers performed the selection.

Assessment of study quality
Validity was assessed using a critical appraisal checklist. This assessed sample characteristics and inclusion criteria (spectrum composition bias), use of an appropriate reference standard, blinding of the person administering the test to the result of the reference standard (test review bias), verification bias and quality of data reporting. All of the papers and abstracts were critically appraised by one of the authors. Another author reviewed papers that were difficult to interpret and any disagreements were resolved by discussion.

Data extraction
A standardised form was used for the critical appraisal data.

**Methods of synthesis**

*How were the studies combined?*

The authors identified studies evaluating pairs of kits and compared the kits only within those studies. Pairs of kits were subjected to analysis only if there were at least 4 studies comparing them. Two methods were used to compare the pairs of kits.

1. Differences in sensitivity and specificity between pairs of kits were calculated for each of the studies collected. These were then averaged, with equal weights, across the group of common studies and tested for significance using the paired t-test. In addition to the t-test, McNemar's test was performed in the 4 studies in which serological results for each reagent kit under comparison were available for every individual in the sample.

2. Summary receiver operator characteristic (sROC) curves were constructed for the same groups of studies, and were used to compare the overall test accuracy between the kits. A sROC curve was also constructed for all the included studies.

*How were differences between studies investigated?*

Factors associated with study design and population, which were considered to be possible predictors of accuracy, were investigated using univariate regression analysis: added as separate variables, one at a time, to the sROC model.

**Results of the review**

Twenty-one studies (11 published papers and 10 abstracts) were included; the total number of participants was not stated.

Comparing test accuracy of different kits: 3 of the 11 pairs of kits showed some significant differences.

**Cobas-IG and Helico-G:** Cobas-IG displayed a significantly higher sensitivity (11%, \( P<0.001 \)), but there was no significant difference in specificities (\( P>0.9 \)) using the paired t-test (McNemar's test was insignificant in the two available papers). The sROC analysis also showed no significant difference.

**Cobas-2G and Biorad-GAP:** there was no significant difference in the sensitivities (\( P<0.1 \)) between these tests; Cobas-2G displayed a significantly higher specificity (\( P<0.05 \)) when using the paired t-test (McNemar's test was insignificant in the only available study). The sROC analysis suggested significantly greater accuracy for Cobas-2G (\( P=0.05 \)). Malakit and Helico-G: Malakit displayed a significantly higher sensitivity (8%, \( P<0.05 \)), but there was no significant difference in specificities (\( P<0.03 \)) between the two tests when using the paired t-test (McNemar's test was insignificant in the only available paper). The sROC analysis showed no significant differences between the two tests.

Overall test accuracy: the average sensitivity and specificity of all tests over all studies were 85 and 79%, respectively.

Potential predictors of test accuracy: prevalence of infection was found to be the only potential predictor of test accuracy; a lower prevalence was associated with a higher test accuracy (\( P=0.03 \)).

**Authors' conclusions**

There was little evidence in the literature to suggest that any one of the common commercial serological kits was more accurate than the other kits. The overall accuracy of these kits may not be adequate for clinical decision-making in all patient groups.

**CRD commentary**

The objective of the study was clearly stated. However, no inclusion criteria were defined and there was no information on the participants, study design, or the selection of studies for their relevance. The lack of information on the included studies makes it very difficult to assess the relevance of the review to particular clinical populations. The description of
review methodology was reasonable, and the validity assessment addressed sources of bias commonly associated with studies of diagnostic accuracy. A single reviewer conducted much of the validity assessment, a practice that can result in bias. The fact that the search strategy was limited to English language publications, and that MEDLINE was the only database searched, may mean that relevant literature has been missed. The authors acknowledged the possibility of publication bias.

The authors discussed the limitations of this review thoroughly. The techniques used to combine the studies, compare accuracy between the tests, and investigate potential sources of heterogeneity were appropriate and were well described. The review provides an interesting comparison of statistical techniques used to test for differences between mean sensitivities and specificities across the studies, and differences in the overall accuracy derived from sROC curves. The authors’ conclusions are appropriately cautious given the data presented.

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

**Practice:** The authors stated that the overall accuracy of thesetestests might not be adequate for clinical decision-making in all patient groups.

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

**Bibliographic details**


**PubMedID**

8651160

**Indexing Status**

Subject indexing assigned by NLM

**MeSH**

Helicobacter Infections /diagnosis; Helicobacter pylori; Humans; ROC Curve; Reagent Kits, Diagnostic /standards /statistics & numerical data; Sensitivity and Specificity; Serologic Tests /instrumentation /standards /statistics & numerical data

**AccessionNumber**

11996001074

**Date bibliographic record published**

30/04/2004

**Date abstract record published**

30/04/2004

**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.