Comparative Accuracy of 3 Monoclonal Stool Tests for Diagnosis of Helicobacter pylori Infection among Patients with Dyspepsia

Xavier Calvet, Sergio Lario, Maria José Ramírez-Lázaro, Antònia Montserrat, Mariela Quesada, Lynsey Reeves, Helen Masters, David Suárez-Lamas, Marta Gallach, Jordi Sánchez-Delgado, Eva Martínez-Bauer, Mireia Miquel, Félix Junquera, Isabel Sanfeliú, and Ferran Segura

Background. Well-devised studies comparing new but different monoclonal fecal tests for diagnosing Helicobacter pylori infection are scarce. The objective of this study was to compare the diagnostic accuracy of 3 monoclonal stool tests: 2 rapid in-office tools—RAPID Hp StAR and ImmunoCard STAT! HpSA—and an enzyme immunoassay test—Amplified IDEIA Hp StAR—for diagnosing H. pylori infection prior to eradication treatment.

Methods. Diagnostic reliability was evaluated in 199 untreated consecutive patients with dyspeptic symptoms. The gold standard for diagnosing H. pylori infection was defined as the concordance of the rapid urease test, histopathology, and urea breath test. Readings of immunochromatographic tests were performed by 2 different observers. Sensitivity, specificity, positive and negative predictive values, and 95% confidence intervals were calculated. Sensitivity and specificity were compared using the McNemar test.

Results. The sensitivity and specificity of Amplified IDEIA Hp StAR were 90% and 89%, respectively. This enzyme immunoassay test was significantly more sensitive than ImmunoCard STAT! HpSA and more specific than RAPID Hp StAR. The sensitivity and specificity of RAPID Hp StAR were 91% and 80%, respectively, according to observer 1, and 92% and 76%, respectively, according to observer 2. It was significantly more sensitive and less specific than ImmunoCard STAT! HpSA. The sensitivity and specificity of ImmunoCard STAT! HpSA were 69% and 90%, respectively, according to observer 1, and 74% and 89%, respectively, according to observer 2.

Conclusions. Amplified IDEIA Hp StAR seems to be the most accurate stool test for diagnosing H. pylori for patients with dyspeptic symptoms. The currently available in-office tests obtain slightly less reliable results.
The second possible reason for the variability is the technical characteristics of the tests. Previous studies have suggested that EIAs offer greater accuracy than does immunochromatography. Most studies with the Amplified IDEIA Hp StAR have generally shown a sensitivity and specificity of ~90% [9–21]. Evaluations of ImmunoCard STAT! HpSA (Meridian Diagnostics)—the first widely available immunochromatographic test—have also been generally positive [8–10, 12, 15, 16, 20–26]. Direct comparisons between the 2 tests tend to be favor the former [9, 10, 12, 15, 16, 18, 20, 21].

A few years ago, a new immunochromatographic test (RAPID Hp StAR; Oxoid) was developed. It is manufactured using the same monoclonal antibody as is used in Amplified IDEIA Hp StAR to detect H. pylori, but it allows an immediate in-office reading. Initial evaluations, both for diagnosing infection for patients with dyspeptic symptoms and for determining a cure after eradication, showed suboptimal accuracy [9, 12, 19]. For this reason, this device has recently been modified by the manufacturer to increase its reliability. To date, the new test has not been evaluated. Therefore, the present study aimed to compare the accuracy of 3 last-generation H. pylori stool tests—2 rapid immunochromatographic monoclonal tests (ie, RAPID Hp StAR and ImmunoCard STAT! HpSA) and a monoclonal EIA (ie, Amplified IDEIA Hp StAR)—for the diagnosis of H. pylori infection in a large series of consecutive patients with dyspeptic symptoms.

**METHODS**

Patients. Outpatients sent to the endoscopy unit of the Hospital de Sabadell for evaluation of dyspeptic symptoms during the period from February 2006 to January 2008 were recruited for the study. Patients were contacted prior to the endoscopy and were asked to participate. Those who agreed were instructed to avoid using antisecretory drugs 2 weeks before the test. Patients unable to stop antisecretory drugs, those who had received antibiotics 4 weeks before the endoscopy, and those with previous treatment for H. pylori infection were excluded. Patients were asked to bring a fecal sample on the day the endoscopy was to be performed. Before the endoscopy, each patient had to sign an informed consent form, and a 13C-urea breath test (UBiT est 100 mg; Otsuka Pharmaceutical Europe) was administered. During endoscopy, 2 antral biopsies for histopathology and 1 for rapid urease test (JATROX HP test; CHR Heim Arzneimittel) were obtained. Aliquots of the feces were frozen and stored at ~80°C for further analysis.

Two hundred and nine patients were included in the study. For a variety of technical reasons, 10 of these patients were excluded because of the unavailability of the urea breath test, fecal samples, or histology. The remaining 199 patients were available for the study. Patients’ clinical and demographic data are shown in Table 1.

**Table 1. Demographic and Clinical Characteristics of the Patients with Dyspepsia in Our Study**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Patients (n = 199)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>92 (46)</td>
</tr>
<tr>
<td>Female</td>
<td>107 (54)</td>
</tr>
<tr>
<td><strong>Age, mean years ± SD</strong></td>
<td>48.2 ± 14.2</td>
</tr>
<tr>
<td><strong>Endoscopic diagnosis</strong></td>
<td></td>
</tr>
<tr>
<td>Duodenal ulcer and/or erosive duodenitis</td>
<td>34 (17)</td>
</tr>
<tr>
<td>Gastric ulcer</td>
<td>4 (2)</td>
</tr>
<tr>
<td>Erosive gastritis</td>
<td>12 (6)</td>
</tr>
<tr>
<td>Oesophagitis</td>
<td>26 (13)</td>
</tr>
<tr>
<td>Normal</td>
<td>123 (62)</td>
</tr>
</tbody>
</table>

**NOTE.** Data are no. (%) of patients, unless otherwise indicated.

SD, standard deviation.

The gold standard for diagnosing H. pylori infection was defined by the concordance of the results from the rapid urease test, histopathology, and the urea breath test, in accordance with the recommendations of the European H. pylori Study Group [27]. Patients with ≥2 positive test results were considered infected; otherwise, they were considered uninfected.

Fecal tests. After the stool samples were unrozen, all the tests were performed simultaneously. All 3 are commercial tests and were performed according to the specifications of the manufacturers.

Two independent observers (S.L. and M.J.R.-L.) performed the readings of the immunochromatographic fecal tests. They were unaware of the results of the reference techniques, the results of the other tests, or the readings of the other evaluator. Results obtained with ImmunoCard HpSA and RAPID Hp StAR were graded as follows: A strong clear test line equal to or stronger than the control line was given a score of 3; a clear test line that was slightly weaker than the control line was given a score of 2; a clear pale test line was given a score of 1; a trace test line was given a score of 0.5; and no test line (control line only) was given a score of 0.

Tests were considered positive when a score of ≥0.5 was assigned by the reader. In cases in which one observer rated the test as positive—usually a trace line reading of 0.5—and the other rated it as negative, the test was repeated and blindly read again by the 2 observers.

**Statistical methods.** Results of the immunochromatographic tests were analyzed by each observer. No attempt was made to achieve consensus on the discordant results during the initial evaluation. Sensitivity, specificity, positive and negative predictive values (and their 95% confidence intervals), and positive and negative likelihood ratios were calculated by standard methods. The McNemar test was used to compare the sensitivity and specificity of the different tests [28]. To correct for multiple com-
parisons, only $P$ values of $<.01$ were considered to be statistically significant. The Pearson correlation and the Spearman test were used to assess the concordance between the readings of the 2 observers of the immunochromatographic tests. Quantitative variables are given as mean values ($\pm$ standard deviation).

Assuming a prevalence of $H. pylori$ infection of 75% in the population of patients with dyspepsia whose tests were evaluated [29, 30], a sample size of 203 patients was calculated to obtain an estimation of sensitivity with a confidence interval of 0.05 and a confidence level of 0.9. All calculations were performed using SPSS for Windows, version 15.0 (SPSS). The study was performed in compliance with the recommendations of the STAndards for the Reporting of Diagnostic (STARD) accuracy studies [31].

RESULTS

According to the gold standard selected, 109 patients were positive for $H. pylori$ infection (with 16 patients having 2 positive reference test results and 93 patients having 3 positive reference test results), and 90 patients were negative for $H. pylori$ infection (with 14 patients having 1 positive test result [1 histology finding, 1 rapid urease test result, and 12 urea breath test results] and 76 patients having all negative reference test results). The possible reasons for the unexpectedly high rate of false-positive results from the commercial urea breath test used in our study, and the approach taken to correct it, have been described in a previous study [32]. A STARD flow diagram [31] of the study is shown in Figure 1.

The EIA had the best diagnostic reliability, with sensitivity and specificity close to 90% (Table 2). Among the immunochromatographic tests, the new version of RAPID Hp StAR had a sensitivity ranging from 91% to 92% and a specificity ranging from 76% to 80%. Its sensitivity was significantly better than that of ImmunoCard STAT! HpSA ($P<.001$), but its specificity was lower. This last difference, however, reached statistical significance ($P<.01$) in some but not all of the comparisons (Table 2). Correlation between observers was good for both RAPID Hp StAR (Pearson correlation, 0.949; $P<.001$) and ImmunoCard STAT! HpSA (Pearson correlation, 0.89; $P<.001$).

In post hoc analyses, neither the use of only histology finding and rapid urease test as the gold standard, nor the exclusion of patients with only 1 positive reference test result, nor the exclusion of trace line readings led to appreciable changes in the sensitivity and specificity of the fecal tests. Considering trace line readings as negative decreased the reliability of the immunochromatographic test considerably. Considering patients with only 1 positive reference test result as infected also clearly reduced the reliability of all stool tests (data not shown).

Discordant readings in immunochromatographic tests. Discrepancies in trace lines were quite frequent. Observer 1 and observer 2, respectively, rated 23 (12%) and 30 (15%) of the 199 RAPID Hp StAR readings as 0.5. The corresponding values for 199 ImmunoCard Stat! HpSA readings were 26 (13%) and 34 (17%).

In addition, some tests were interpreted as positive by one observer and as negative by the other. Five readings for RAPID Hp StAR and 11 readings for ImmunoCard Stat! HpSA were interpreted as positive by one of the readers and negative by
the other. In these cases, the test was repeated. The 2 observers agreed on the interpretation of the new test in 4 of the 5 discordant results for RAPID Hp StAR and in 9 of the 11 discordant results for ImmunoCard STAT! HpSA. Repeating the tests with discordant readings, however, did not appreciably improve the diagnostic performance (data not shown).

**DISCUSSION**

Probably the most important message of the present study is that there are major differences in the diagnostic accuracy of the different fecal tests, even when monoclonal antibody-based tests are compared. As stated, there is clear, accumulated evidence showing that polyclonal *H. pylori* stool tests are far less reliable than monoclonal ones and that these devices should be avoided in clinical practice [2, 3].

What is new in our study is that the technical characteristics of monoclonal tests are also of paramount importance. In this regard, immunochromatographic tests are potentially very good alternatives for immediate in-office *H. pylori* diagnosis; the results are available within minutes and do not require the use of laboratory equipment. These features represent clear advantages over EIAs, which must be performed in the laboratory and give delayed results. Unfortunately, the present study shows that the diagnostic reliability of monoclonal immunochromatographic tests is lower than that of monoclonal EIAs. Specifically, the diagnostic performance of Amplified IDEIA Hp StAR was better than both the new RAPID Hp StAR and ImmunoCard STAT! HpSA for diagnosing *H. pylori* infection. Therefore, immunochromatographic tests need further improvement before they can be compared with laboratory EIAs. Among other aspects, the problem of trace line readings clearly should be addressed.

Regarding the comparison of the 2 rapid tests, the present study also shows that the new RAPID Hp StAR has a sensitivity of >90% and a specificity of 76%–80%, depending on the observer. ImmunoCard STAT! HpSA has 69%–74% sensitivity and 89%–90% specificity. The comparison between tests showed that RAPID Hp StAR was significantly more sensitive than ImmunoCard STAT! HpSA, although it shows lower specificity. A further important point highlighted by our study is that, because accuracy seems to depend on both the test’s technical characteristics and the antigen detected, local validation seems advisable before a stool test can be recommended for a given population.

In general, our results for the 2 stool tests that have been evaluated previously corroborate those of other researchers. As stated, the performance of Amplified IDEIA Hp StAR has been consistently good [9–21]. ImmunoCard STAT! HpSA has achieved good results in a post eradication setting and more irregular results among patients with dyspepsia [8–10, 12, 15, 16, 18, 20–26]. The few direct comparisons of the 2 tests performed to date tend to favor the EIA [9, 10, 12, 15, 16, 18, 20, 21].

The major strengths of our study are the large number of patients and the strict methodology followed to establish the gold standard and to obtain the samples. A clear limitation is that, unexpectedly, the commercial urea breath test that was used provided a high rate of false-positive results, thus requiring a recalculation of the urea breath test cutoff [32]. This point was addressed both a priori and a posteriori, a priori by using the stringent criteria of the European *H. pylori* Study Group [27] for considering that a patient was infected, and a posteriori by performing multiple post hoc analyses. These post hoc analyses evaluated the possible effect of the suboptimal performance of the urea breath test in the study results. With this purpose in mind, the preestablished gold standard was modified in 2 different ways: by using only biopsy-based tests as a gold standard or by excluding from the analysis all patients with only a positive reference test result. None of these post hoc analyses showed appreciable improvements in stool test accuracy, thus confirming the reliability of the findings of our study. The results of these post hoc analyses are available on request.

### Table 2. Data on Comparison of the 3 Fecal Tests for Diagnosing *Helicobacter pylori* Infection

<table>
<thead>
<tr>
<th>Fecal test, observer&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>PPV, % (95% CI)</th>
<th>NPV, % (95% CI)</th>
<th>+LR</th>
<th>−LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplified IDEIA Hp StAR</td>
<td>89.9&lt;sup&gt;b&lt;/sup&gt; (82–94)</td>
<td>88.9&lt;sup&gt;c&lt;/sup&gt; (80–94)</td>
<td>90.7 (83–95)</td>
<td>87.9 (79–94)</td>
<td>8.1</td>
<td>0.11</td>
</tr>
<tr>
<td>RAPID Hp StAR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observer 1</td>
<td>90.8&lt;sup&gt;b&lt;/sup&gt; (83–95)</td>
<td>80.0 (70–87)</td>
<td>84.6 (76–90)</td>
<td>87.8 (78–94)</td>
<td>4.5</td>
<td>0.11</td>
</tr>
<tr>
<td>Observer 2</td>
<td>91.7&lt;sup&gt;b&lt;/sup&gt; (84–96)</td>
<td>75.6 (65–84)</td>
<td>82.0 (74–88)</td>
<td>88.3 (78–94)</td>
<td>3.8</td>
<td>0.11</td>
</tr>
<tr>
<td>ImmunoCard STAT! HpSA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observer 1</td>
<td>68.8 (59–77)</td>
<td>90.0&lt;sup&gt;c&lt;/sup&gt; (81–95)</td>
<td>89.3 (80–95)</td>
<td>70.4 (61–78)</td>
<td>6.9</td>
<td>0.35</td>
</tr>
<tr>
<td>Observer 2</td>
<td>74.3 (65–82)</td>
<td>88.9&lt;sup&gt;c&lt;/sup&gt; (80–94)</td>
<td>89.0 (80–94)</td>
<td>74.1 (65–82)</td>
<td>6.7</td>
<td>0.29</td>
</tr>
</tbody>
</table>

**NOTE.** CI, confidence interval; +LR, positive likelihood ratio; −LR, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value.<br>

<sup>a</sup> Readings of immunochromatographic tests were performed by 2 different observers.<br>

<sup>b</sup> *P*<sub>.01</sub>, compared with ImmunoCard STAT! HpSA (observers 1 and 2).<br>

<sup>c</sup> *P*<sub>.01</sub>, compared with RAPID Hp StAR (observer 2).
A second problem experienced during the development of the study was the interpretation of the trace line and discordant readings when analyzing the immunochromatographic tests. Thus, we attempted to repeat the analysis of the samples when readings were discordant. Readings of these repeated tests were coincidental in most cases. Unfortunately, because the results of this second test did not correlate especially well with the gold standard, repeated readings did not appreciably change the sensitivity and specificity of the tests. With regard to the significance of trace reading lines, considering trace line results as negative markedly reduced the accuracy of the immunochromatographic tests.

Further validation of the tests in different populations is needed to confirm the present results. This is especially true for the new RAPID Hp StAR, because, to our knowledge, this is the first study to evaluate its reliability. In addition, it is quite possible that further refinements in the immunochromatographic tests could improve their reliability, thus making these practical tests suitable for rapid in-office diagnosis of the H. pylori infection. Such improvements are eagerly awaited.

In conclusion, we found that the new modified in-office RAPID Hp StAR test shows an acceptable sensitivity for detecting H. pylori infection among patients with dyspepsia, although specificity was suboptimal. By contrast, ImmunoCard STAT! HpSA shows a low sensitivity but a better specificity. Though specificity was suboptimal. By contrast, ImmunoCard STAT! HpSA shows a low sensitivity but a better specificity. Hp StAR test shows an acceptable sensitivity for detecting H. pylori infection in stools. Such improvements are eagerly awaited.

In conclusion, we found that the new modified in-office RAPID Hp StAR test shows an acceptable sensitivity for detecting H. pylori infection among patients with dyspepsia, although specificity was suboptimal. By contrast, ImmunoCard STAT! HpSA shows a low sensitivity but a better specificity. None of the immunochromatographic tests, however, performed as well as Amplified IDEIA Hp StAR, which, at present, seems to be the best available stool test for diagnosing H. pylori infection.

Acknowledgments

We thank Neus Mateo, Alicia Sánchez, and Laura Moreno for their help in collecting samples and performing the breath test. We are also indebted to Michael Maudsley for his help with the English.

Financial support. Oxoid; Instituto de Salud Carlos III (grants PI 05/1157 and PI 05/0664); and Societat Catalana de Digestologia.

Potential conflicts of interest. I.R. and H.M. report that they are employees of Oxoid. X.C., S.L., and M.J.R.-L. report that they have received recent research funding from Oxoid. All other authors: no conflicts.

References


