The endoscopic predictors of *Helicobacter pylori* status: a meta-analysis of diagnostic performance

Ben Glover, Julian Teare, Hutan Ashrafian and Nisha Patel

**Abstract**

**Objective:** The endoscopic findings associated with *Helicobacter pylori*-naïve status, current infection or past infection are an area of ongoing interest. Previous studies have investigated parameters with a potential diagnostic value. The aim of this study was to perform meta-analysis of the available literature to validate the diagnostic accuracy of mucosal features proposed in the Kyoto classification.

**Data sources:** The databases of MEDLINE and Embase, clinicalTrials.gov and the Cochrane Library were systematically searched for relevant studies from October 1999 to October 2019.

**Methods:** A bivariate random effects model was used to produce pooled diagnostic accuracy calculations for each of the studied endoscopic findings. Diagnostic odds ratios and sensitivity and specificity characteristics were calculated to identify significant predictors of *H pylori* status.

**Results:** Meta-analysis included 4380 patients in 15 studies. The most significant predictor of an *H pylori*-naïve status was a regular arrangement of collecting venules (diagnostic odds ratio 55.0, sensitivity 78.3%, specificity 93.8%). Predictors of active *H pylori* infection were mucosal oedema (18.1, 63.7%, 91.1%) and diffuse redness (14.4, 66.5%, 89.0%). Map-like redness had high specificity for previous *H pylori* eradication (99.0%), but poor specificity (13.0%).

**Conclusion:** The regular arrangement of collecting venules, mucosal oedema, diffuse redness and map-like redness are important endoscopic findings for determining *H pylori* status. This meta-analysis provides a tentative basis for developing future endoscopic classification systems.

**Keywords:** gastritis, *Helicobacter pylori*, image enhancement, lesion recognition, optical diagnosis

Received: 7 February 2020; revised manuscript accepted: 8 July 2020.

**Introduction**

Since the discovery of *Helicobacter pylori* as an infective agent implicated in the development of peptic ulceration, there has been growing recognition of its role in chronic gastritis, atrophic changes, metaplasia and eventual development of gastric cancers. It is now understood that eradication of *H pylori* by antibiotic treatment can arrest the progression of this pathway and reduce the subsequent risk of cancer. Several diagnostic approaches, including the urea breath test (UBT), stool antigen testing and serological testing for noninvasive diagnosis, as well as endoscopic biopsy for rapid urease test (RUT), histological examination or tissue culture for organisms, are available for assessment of *H pylori* status.

It has long been suspected that the endoscopic appearance of the gastric mucosa may change as a consequence of *H pylori* infection, providing useful diagnostic information to the endoscopist. Early work in this area was characterised by the use of magnifying endoscopy for close examination of the stomach, and demonstrated visible changes in collecting venules of the *H pylori*-infected stomach. The normal appearance is...
Characterised by a ‘regular arrangement of collecting venules’ (RAC), in the gastric corpus, the loss of which is associated with H pylori infection. More recently, the improved resolution and image quality of modern endoscopes has allowed for ever higher levels of mucosal detail to be appreciated, raising the possibility that H pylori predictive mucosal features could be seen even without the use of magnification. The introduction of image enhancement as an adjunct to white light endoscopy (WLE) further improved the level of visual contrast and allowed greater accuracy of assessment for mucosal features.

In the modern era of high-definition endoscopy, the RAC has been confirmed as an important endoscopic predictor of an H pylori-naive stomach, which is visible by careful observation without the aid of magnification. Further mucosal features, including diffuse erythema, linear erythema, gastric erosions, mucosal oedema, swollen gastric folds, mosaic appearance of mucosa, fundic gland polyps, mucosal atrophy, intestinal metaplasia and gastric antral nodularity, have been proposed to predict H pylori status. These features, and others, have been investigated to varying degrees, using a variety of endoscopic imaging modalities and study designs, and the Kyoto classification of gastritis divides patients into three groups: H pylori naïve (nongastritis), patients with current infection (active gastritis) and patients with past H pylori infection (inactive gastritis). Attempts have been made in individual studies to identify and calculate the predictive values of the individual endoscopic findings of the Kyoto classification and to generate predictive models but current studies are within relatively small and homogeneous patient groups. Endoscopic assessment of H pylori status has been identified by the Kyoto global consensus report on H pylori gastritis, as a desirable method for diagnosing H pylori infection for increasing the diagnostic yield of targeted biopsies.

This meta-analysis therefore proposes to further explore the diagnostic performance of commonly recognised endoscopic findings, for the prediction of H pylori status. We aim to identify the strongest and most readily recognisable findings, as the basis of forming a unified diagnostic classification to allow simple and accurate prediction of H pylori status at the point of endoscopy.

**Methods**

**Search strategy and study selection**

The protocol for this meta-analysis was registered with the National Institute for Health Research (NIHR) PROSPERO registry of systematic reviews, with the ID number CRD42019153225. Analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The MEDLINE and Embase databases were systematically searched for studies of diagnostic accuracy of the endoscopic features of H pylori, from October 1999 to October 2019. Studies were identified using the MeSH or Emtree headings for ‘Helicobacter pylori’ and ‘Endoscopy’, and search terms for ‘gastritis’, ‘RAC’ and ‘Regular Arrangement of Collecting Venules’, combined with search terms for ‘high definition’, ‘Narrow Band Imaging’ and ‘i-scan’. The full search strategy is shown in Appendix 1. The database of clinicalTrials.gov was searched for any relevant studies with results, using search terms for ‘Helicobacter Gastritis’, and the Cochrane Library was searched for articles using the search term ‘Helicobacter’.

Following the initial search and removal of duplicate articles, the titles and abstracts were screened for relevance by two investigators independently and excluded if not relevant. The abstracts of the remaining articles were scrutinised, followed by a detailed analysis of the full text of remaining suitable articles. During the screening process, relevant review articles were identified, and after compilation of the list of included studies, the reference lists of all review articles and included studies were further examined to identify any further appropriate studies.

**Inclusion and exclusion criteria**

The inclusion criteria were as follows:

- Studies that attempted to use endoscopic findings for an endoscopist to predict the H pylori status of a patient. This could be classified as either positive versus negative or naïve versus positive versus eradicated.
- Studies using WLE, either with or without image enhancement.
- Studies that attempted to describe or define endoscopic findings to establish H pylori status.
- Studies using an objective reference standard, including histological analysis, H pylori
culture, RUT, UBT, \( H\) pylori serology or \( H\) pylori stool antigen.

- Studies with adequate published data to construct a contingency table and calculate true-positive, false-positive, true-negative and false-negative results.
- Studies published or translated into English.

The exclusion criteria were as follows:

- Review or meta-analysis papers.
- Studies with incomplete data to calculate diagnostic performance characteristics.
- Endoscopic features studied could not be correlated to the Kyoto consensus features.
- Studies with overlapping data or participant cohorts from those already included.
- Studies only including children.

**Data extraction**

Two investigators extracted the diagnostic accuracy data from the studies using a standardised data collection spreadsheet. The primary data obtained were the diagnostic accuracy characteristics, including the accuracy, sensitivity and specificity. These were calculated from the rates of true-positive, false-positive, true-negative and false-negative results. The accuracy characteristics were calculated individually for each endoscopic feature assessed in each study.

Secondary data included the number of patients included, the country and year of publication, the experience level of the endoscopists, the number of endoscopists taking part in the study and the use of any image enhancement techniques.

**Study quality assessment**

All included studies were assessed using the revised Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool\(^{28}\) to produce a structured report of both the risk of bias and the wider applicability of each study. Studies deemed to be at high risk of bias or of low applicability were excluded from the final analysis. All domains of the QUADAS-2 tool were included in the study quality assessments.

For a study to be unbiased in its selection, we preferred prospective recruitment of patients, without unusual exclusion criteria. Retrospective studies were considered, if it was clear that patient or image selection for inclusion followed an appropriate patient cohort without preselection of images. We preferred studies in which the prediction of \( H\) pylori status was made during the endoscopic procedure. Postprocedural image analysis was considered acceptable if the image reviewers were suitably blinded to the \( H\) pylori status. Depending on the study design, we preferred that the endoscopist be blinded to the \( H\) pylori status reference test, unless the study implicitly included this information as part of the endoscopist decision-making process.

**Meta-analysis**

Stata version 15 (StataCorp, College Station, TX, USA) was used for all statistical analyses. We used a bivariate model for diagnostic meta-analysis to calculate weighted pooled sensitivity and specificity data. The sensitivity and specificity characteristics were examined using a summary receiver operating characteristic (SROC) model. Prediction regions within this curve were produced, representing the probability of including the true sensitivity and specificity of a future diagnostic study.

Study heterogeneity was assessed using \( I^2 \) (0–30% was considered a low level of heterogeneity, 31–60% was considered a moderate degree of heterogeneity and >60% was considered a high level of heterogeneity). Heterogeneity was calculated separately for each endoscopic finding.

Trapezoidal integration was performed to calculate the pooled area under the curve (AUC). Under this model, a value of 1.0 indicates a perfectly accurate diagnostic test that will produce the correct diagnosis 100% of the time; a value of 0.5 suggests a test that is equally likely to diagnose a truly positive result as either positive or negative.

We calculated the diagnostic odds ratio (DOR) of each endoscopic finding, as a predictor of \( H\) pylori-naïve, \( H\) pylori-positive or previous infection. This is a measure of the odds of test positivity in the presence of a given condition, relative to the odds of test positivity in the absence of that condition.

**Results**

**Eligible studies**

The PRISMA flow diagram for study selection is reported in Figure 1. Following the initial database
searches and removal of duplicate records, 1337 citations were identified as being of potential interest for inclusion. A screening of the titles and abstracts excluded 1182 papers which were not of relevance, and 155 papers were included for full-text review, of which 28 were selected for qualitative and quantitative analysis.

Application of the exclusion criteria to the identified papers found that three studies used magnified rather than standard endoscopy, two studies included children, two studies included an overlapping patient cohort, one study did not include sufficient diagnostic data for calculation of performance characteristics and one study was not performed at high resolution. These studies are discussed further below but did not contain data for inclusion in the meta-analysis. Fifteen studies were included in the quantitative synthesis.

**Quality assessment**

The quality of the 15 included studies was assessed according to the QUADAS-2 tool. Three of the studies were excluded from analysis because of high risk of bias. These were postprocedural, image-based assessments of endoscopic characteristics rather than real-time, and introduced risk of selection bias at the point of selection of the endoscopic images used in the study. Another image-based study was excluded because of high patient exclusion rates, and selection for patients with existing gastric cancer. Of the 15 studies included for analysis, 14 recruited patients prospectively, 10 of which made real-time endoscopic diagnosis during the procedure. Two studies used retrospective image collection but were deemed only medium risk of bias; these were therefore included in the analysis. Overall, the studies showed low risk of bias and good applicability. The results are displayed in Figure 2.

**Figure 1.** Flowchart of study inclusion during the search and review process.
Study characteristics
A total of 15 studies were included in the meta-analysis, containing adequate data to calculate diagnostic accuracy characteristics. The publication dates spanned from 2009 to 2019, and the mean age of patients included was 63.4 years. The mean prevalence of H pylori was 51.8%. A total of 4380 patients were included for analysis. A summary of the study characteristics is included in Table 1. The mucosal features examined in each study are summarised in Table 2.

Tests of diagnostic accuracy
We first assessed the predictors of an H pylori-naïve status; these included the RAC, fundic gland polyps and streaky erythema. Thirteen studies had investigated the RAC, and four studies had investigated each of the other features. These were all found to have a positive DOR as predictors of H pylori-naïve status; the strongest was RAC, with a DOR of 55.0 (95% confidence interval (CI) 19.8–152.5). The sensitivity was 78.3% (66.6–86.7%) and specificity 93.8% (83.9–97.8%). Full diagnostic data for the predictors of H pylori-negative status are presented in Table 3, and SROC in Figure 3.

We next assessed the features believed to be predictors of active H pylori infection, which were diffuse redness, mucosal oedema, sticky mucus, enlarged gastric folds and antral nodularity. All these features were confirmed to have a positive DOR; the strongest was antral nodularity, with a DOR of 22.5, although 95% confidence intervals were extremely wide (0.5–1040.9), sensitivity was 7.2% (2.4–19.3%) and specificity was 99.7% (88.8–99.9%). The presence of mucosal oedema carried a DOR of 18.1 (8.6–37.8), sensitivity was 63.7% (48.7–76.4%) and specificity was 91.1% (86.9–94.1%). The finding of diffuse redness was associated with H pylori infection with a DOR of 14.4 (6.5–31.9), sensitivity was 66.5 (54.4–76.7%) and specificity was 87.9% (78.5–93.5%). These results are presented fully in Table 4, and SROC in Figure 4.

Predicators of previous H pylori eradication are thought to include map-like redness, gastric atrophy, intestinal metaplasia and xanthomas. There were insufficient data to perform meta-analysis of xanthoma and intestinal metaplasia, as insufficient previous diagnostic data are reported. Of the two remaining features, map-like redness was a strong predictor of previous eradication, with a DOR of 12.2 (5.1–29.7), sensitivity of 0.13% (0.06–0.27%) and specificity of 99% (0.95–1.00%). The presence of gastric atrophy was associated with a DOR of 4.0 for prediction of previous H pylori eradication. Full results for these features are presented in Table 5, and SROC in Figure 5.

Finally, we examined the mucosal features for which the association with H pylori is unclear. These included flat or elevated gastric erosions, ‘white flat elevated lesions’ (WFELs), hyperplastic polyps and haemat/flecks. There were sufficient data to perform meta-analysis of erosions, and haem flecks. The primary studies were heterogenous, and distinction could not always be drawn between flat and elevated erosion; therefore, these have been analysed together. Our results suggested that the presence of gastric erosion was associated with a DOR of 1.3 for diagnosis of active H pylori gastritis (0.4–5.0) and that presence of haem flecks was associated with a DOR of 0.3. Results are presented in Table 6.

Tests of heterogeneity
We found that the majority of the studied mucosal features showed a high degree of heterogeneity in their findings ($I^2 > 60%$). Exceptions...
Table 1. Summary of study characteristics.

<table>
<thead>
<tr>
<th>Author</th>
<th>Recruitment</th>
<th>Diagnosis</th>
<th>Patients taking PPI</th>
<th>Previous <em>Helicobacter pylori</em> eradication</th>
<th>Number of patients</th>
<th>Number of endoscopists</th>
<th>Mean age of participants</th>
<th>Study prevalence of <em>H pylori</em> (%)</th>
<th>Reference standard</th>
<th>Image enhancement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gonen and colleagues</td>
<td>Prospective</td>
<td>Real time</td>
<td>4 weeks off</td>
<td>Excluded</td>
<td>129</td>
<td>1</td>
<td>49.8 (Standard deviation 12.4)</td>
<td>76.0</td>
<td>RUT, histology, UBT</td>
<td>–</td>
</tr>
<tr>
<td>Yan and colleagues</td>
<td>Prospective</td>
<td>Postprocedure</td>
<td>Excluded</td>
<td>Excluded</td>
<td>112</td>
<td>2</td>
<td>47 [17–91]</td>
<td>67.9</td>
<td>RUT, histology</td>
<td>–</td>
</tr>
<tr>
<td>Alaboudy and colleagues</td>
<td>Image database</td>
<td>Postprocedure</td>
<td>Unknown</td>
<td>Excluded</td>
<td>390</td>
<td>N/K</td>
<td>62.9 [13]</td>
<td>58.7</td>
<td>Serology, histology</td>
<td>–</td>
</tr>
<tr>
<td>Cho and colleagues</td>
<td>Prospective</td>
<td>Real time</td>
<td>Excluded</td>
<td>Excluded</td>
<td>617</td>
<td>1</td>
<td>50.0 [10.0]</td>
<td>58.2</td>
<td>RUT, histology</td>
<td>–</td>
</tr>
<tr>
<td>Kato and colleagues</td>
<td>Prospective</td>
<td>Real time</td>
<td>4 weeks off</td>
<td>Excluded</td>
<td>723</td>
<td>2</td>
<td>57.3 [12.4]</td>
<td>70.5</td>
<td>Serology, histology, SA</td>
<td>–</td>
</tr>
<tr>
<td>Kato and colleagues</td>
<td>Prospective</td>
<td>Real time</td>
<td>4 weeks off</td>
<td>Excluded</td>
<td>275</td>
<td>24</td>
<td>64.9 [9.3]</td>
<td>26.0</td>
<td>Histology</td>
<td>IC</td>
</tr>
<tr>
<td>Yagi and colleagues</td>
<td>Prospective</td>
<td>Real time</td>
<td>Included</td>
<td>Included</td>
<td>56</td>
<td>1</td>
<td>N/K</td>
<td>67.9</td>
<td>SA</td>
<td>–</td>
</tr>
<tr>
<td>Mino and colleagues</td>
<td>Prospective</td>
<td>Postprocedure</td>
<td>4 weeks off</td>
<td>Included if &gt;4 weeks</td>
<td>256</td>
<td>2</td>
<td>52.0 [11.7]</td>
<td>44.1</td>
<td>Histology</td>
<td>–</td>
</tr>
<tr>
<td>Mino and colleagues</td>
<td>Prospective</td>
<td>Real time</td>
<td>8 weeks off</td>
<td>Included if &gt;8 weeks</td>
<td>170</td>
<td>1</td>
<td>41.2 [14.8]</td>
<td>69.4</td>
<td>RUT, histology</td>
<td>–</td>
</tr>
<tr>
<td>Chen and colleagues</td>
<td>Prospective</td>
<td>Real time</td>
<td>Unknown</td>
<td>Included</td>
<td>111</td>
<td>1</td>
<td>52.35 [12.90]</td>
<td>29.5</td>
<td>RUT, histology, UBT</td>
<td>LCI</td>
</tr>
<tr>
<td>Gomes-Duran and colleagues</td>
<td>Prospective</td>
<td>Real time</td>
<td>2 weeks off</td>
<td>Included</td>
<td>140</td>
<td>3</td>
<td>49.7 [15.7]</td>
<td>31.0</td>
<td>RUT, histology</td>
<td>–</td>
</tr>
<tr>
<td>Yoshi and colleagues</td>
<td>Prospective</td>
<td>Real time</td>
<td>2 weeks off</td>
<td>Included</td>
<td>498</td>
<td>7</td>
<td>53.1 [16.2]</td>
<td>37.7</td>
<td>Serology, histology</td>
<td>–</td>
</tr>
<tr>
<td>Nishio and colleagues</td>
<td>Prospective</td>
<td>Real time</td>
<td>2 weeks off</td>
<td>Included</td>
<td>127</td>
<td>5 Reviewers</td>
<td>62.4 [14.0]</td>
<td>50.4</td>
<td>Serology, UBT</td>
<td>LCI</td>
</tr>
<tr>
<td>Inui and colleagues</td>
<td>Image database</td>
<td>Postprocedure</td>
<td>Included</td>
<td>Included</td>
<td>576</td>
<td>4 Reviewers</td>
<td>63.4</td>
<td>34.0</td>
<td>RUT, histology, SA</td>
<td>NBI</td>
</tr>
</tbody>
</table>

IC, indigo carmine; LCI, linked colour imaging; NBI, narrow-band imaging; PPI, proton pump inhibitor; RUT, rapid urease test; SA, stool antigen; UBT, urea breath test.
Table 2. Summary of mucosal features assessed in each included study.

<table>
<thead>
<tr>
<th>Author</th>
<th>Mucosal features assessed</th>
<th>Predictors of Helicobacter pylori-naïve status</th>
<th>Predictors of active <em>H pylori</em> infection</th>
<th>Predictors of previous <em>H pylori</em> infection</th>
<th>Other predictors of <em>H pylori</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RAC</td>
<td>FGP</td>
<td>Red streak</td>
<td>Diffuse redness</td>
<td>Mucosal oedema</td>
</tr>
<tr>
<td>Gonen and colleagues</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Yan and colleagues</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Alaboudy and colleagues</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cho and colleagues</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Katake and colleagues</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kato and colleagues</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Yagi and colleagues</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Gomes and colleagues</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mao and colleagues</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Matrakool and colleagues</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chen and colleagues</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garces-Duran and colleagues</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yoshii and colleagues</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Ono and colleagues</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inui and colleagues</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

FGP, fundic gland polyp; RAC, regular arrangement of collecting venules.
**Table 3.** Diagnostic performance for prediction of *Helicobacter pylori*-naïve status.

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>DOR (95% CI)</th>
<th>AUROC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular arrangement of collecting venules</td>
<td>78.3 [66.6–86.7]</td>
<td>93.8 [83.9–97.8]</td>
<td>55.0 [19.8–152.5]</td>
<td>0.92 [0.89–0.94]</td>
</tr>
<tr>
<td>Fundic gland polyps</td>
<td>20.4 [12.9–30.6]</td>
<td>96.9 [93.4–98.5]</td>
<td>7.9 [4.2–15.1]</td>
<td>0.81 [0.77–0.84]</td>
</tr>
<tr>
<td>Red streak</td>
<td>19.5 [12.6–28.9]</td>
<td>95.4 [90.9–97.8]</td>
<td>5.1 [2.9–7.3]</td>
<td></td>
</tr>
</tbody>
</table>

AUROC, area under receiver operating characteristic; CI, confidence interval; DOR, diagnostic odds ratio.

**Figure 3.** SROC curves for the use of RAC presence and FGPs as a predictor of *Helicobacter pylori*-naïve status.

FGPs, fundic gland polyps; RAC, regular arrangement of collecting venules; SROC, summary receiver operating characteristic.

**Table 4.** Diagnostic performance for prediction of active *Helicobacter pylori* infection.

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>DOR (95% CI)</th>
<th>AUROC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodularity</td>
<td>7.2 [2.4–19.3]</td>
<td>99.7 [88.8–99.9]</td>
<td>22.5 [0.5–1040.9]</td>
<td>0.84 [0.81–0.87]</td>
</tr>
<tr>
<td>Mucosal oedema</td>
<td>63.7 [48.7–76.4]</td>
<td>91.1 [86.9–94.1]</td>
<td>18.1 [8.6–37.8]</td>
<td>0.90 [0.87–0.93]</td>
</tr>
<tr>
<td>Diffuse redness</td>
<td>66.5 [54.4–76.7]</td>
<td>87.9 [78.5–93.5]</td>
<td>14.4 [6.5–31.9]</td>
<td>0.84 [0.81–0.87]</td>
</tr>
<tr>
<td>Sticky mucus</td>
<td>0.48 [0.26–0.70]</td>
<td>0.89 [0.52–0.98]</td>
<td>7.0 [2.0–27.0]</td>
<td>0.68 [0.64–0.72]</td>
</tr>
<tr>
<td>Enlarged fold</td>
<td>0.47 [0.29–0.65]</td>
<td>0.87 [0.69–0.96]</td>
<td>6.0 [2.8–12.9]</td>
<td>0.72 [0.68–0.76]</td>
</tr>
</tbody>
</table>

AUROC, area under receiver operating characteristic; CI, confidence interval; DOR, diagnostic odds ratio.

to this were the findings of antral nodularity and of haem flecks, which showed low interstudy heterogeneity, and gastric erosions, which showed moderate interstudy heterogeneity.

**Discussion**

This meta-analysis has attempted to combine the existing studies of diagnostic accuracy of endoscopic findings in the stomach, in the context of
Figure 4. SROC curves for the use of diffuse redness, mucosal oedema, sticky mucus and enlarged folds as predictors of active *Helicobacter pylori* infection.

SROC, summary receiver operating characteristic.

Table 5. Diagnostic performance for prediction of *Helicobacter pylori*–eradicated status.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>DOR (95% CI)</th>
<th>AUROC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Map-like redness</td>
<td>0.13 (0.06–0.27)</td>
<td>0.99 (0.95–1.00)</td>
<td>12.2 (5.1–29.7)</td>
<td>0.67 (0.63–0.71)</td>
</tr>
<tr>
<td>Atrophy</td>
<td>77.6 (47.8–93.0)</td>
<td>53.5 (13.8–89.2)</td>
<td>4.0 (0.4–44.1)</td>
<td>0.74 (0.70–0.78)</td>
</tr>
</tbody>
</table>

AUROC, area under receiver operating characteristic; CI, confidence interval; DOR, diagnostic odds ratio.

prediction of *H pylori* status. Over recent years, the area of investigation has developed from a prediction of ‘positive versus negative’ towards a more nuanced prediction of ‘naïve versus positive versus eradicated’ status, recognising the increased risk of gastric cancer associated with *H pylori* gastritis, as well as chronic gastric atrophy and intestinal metaplasia.
There is increasing interest in the endoscopic prediction of \( H \) \( \text{pylori} \) status, following a growing body of evidence that endoscopic changes may not regress, and risk of progression to gastric cancer may remain elevated even after \( H \) \( \text{pylori} \) eradication.\(^{46}\) Furthermore, in 2013, the Japanese national health insurance system approved the funding of eradication therapy for patients with an endoscopic diagnosis of active \( H \) \( \text{pylori} \) gastritis, with the aim of reducing the mortality associated with gastric cancer.\(^{47}\) This has encouraged the development of classification systems such as the Endoscopic ABC\(^{48}\) which show a high degree of accuracy for prediction of \( H \) \( \text{pylori} \) status.\(^{37}\) The interpretation of endoscopic findings remains complex however, and a simple identification system for the general endoscopist would be of use to allow rapid diagnosis of \( H \) \( \text{pylori} \) status in a less specialist setting.

The studies included in the analysis reflect this change in the literature; those conducted before 2014 had excluded patients with previous \( H \) \( \text{pylori} \) infection, and aimed to differentiate between naïve noninfection and active infection. The arrival of the Kyoto classification in 2015 clarified the importance of also recognising the \( H \) \( \text{pylori} \)-eradicated state, and subsequent studies have tended to include these patients.

Some of the findings described here have previously been extensively investigated; the RAC, for example, identified in 2002,\(^6\) is widely understood to be a predictor of an \( H \) \( \text{pylori} \)-naïve stomach, and is included in most studies analysed here. Likewise, diffuse redness is an established predictor of active infection, and map-like redness is the most extensively evaluated finding to suggest previous infection.

Other findings are less well understood, such as the presence of ‘WFELs’ which are observed in the fundus of an \( H \) \( \text{pylori} \)-naïve and \( H \) \( \text{pylori} \)-eradicated stomach, although are of uncertain clinical

---

**Figure 5.** SROC curves for the use of atrophy and map-like redness as predictors of previous *Helicobacter pylori* infection.  
SROC, summary receiver operating characteristic.

**Table 6.** Diagnostic characteristics for mucosal features of unknown significance.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>DOR (95% CI)</th>
<th>AUROC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gastric erosions</td>
<td>9.0 (4.9–15.9)</td>
<td>93.4 (83.6–97.5)</td>
<td>1.4 (0.4–4.9)</td>
<td>0.36 (0.32–0.4)</td>
</tr>
<tr>
<td>Haem flecks</td>
<td>3.0 (1.3–9.0)</td>
<td>90.7 (86.4–93.6)</td>
<td>0.3 (0.1–0.8)</td>
<td>0.74 (0.70–0.78)</td>
</tr>
</tbody>
</table>

AUROC, area under receiver operating characteristic; CI, confidence interval; DOR, diagnostic odds ratio.
The H pylori-naive stomach
As expected, the presence of RAC in the upper stomach was the strongest predictor of H pylori-naive status. The RAC is a distinctive mucosal appearance, of multiple tiny red starfish-like points, spread throughout the mucosa and readily visible under close endoscopic examination. Previous studies have demonstrated specificity as high as 100% for diagnosis of an H pylori-naive stomach; the pooled meta-analysis showed impressive diagnostic performance, with sensitivity of 78.3%, specificity of 93.8% and a DOR of 55.0. As the RAC is a distinctive finding with a fast learning curve for identification and strong diagnostic accuracy, it would be an appropriate finding for use in a simplified endoscopic assessment.

It must however be remembered that the appearance of the RAC can change and become less prominent with age, even in the absence of H pylori infection, and the diagnostic accuracy of the RAC is optimal in patients younger than 50 years. The appearance of the RAC may also vary throughout the stomach, and is rarely visible within the gastric antrum. We suggest therefore that the diagnostic utility of the RAC should be applied mainly when identified in the gastric corpus.

The H pylori-infected stomach
In this analysis, the finding carrying the highest DOR (22.5) for predicting active H pylori infection was nodularity at the gastric antrum. However, caution should be applied to interpreting this result, in view of the wide confidence interval (0.5–1040), which is likely to be related to the very low incidence of antral nodularity within the analysed studies. As seen in previous studies, antral nodularity does appear to be a very specific finding for the presence of active H pylori gastritis, but its relative rarity would make it less useful as a component of a simple assessment system.

Of greater use could be the presence of diffuse redness (DOR 14.1) or mucosal oedema (DOR 18.1), observed anywhere in the stomach. Identifying these findings could be considered more open to subjective assessment than more focal findings, but this could be standardised somewhat by the development of training resources and education on the expected appearances of the findings.

The H pylori-eradicated stomach
The most extensively studied predictor of previous H pylori infection is ‘map-like redness’, a pattern of red irregular areas of varying size. No standard description of ‘map-like redness’ has been previously proposed, although many studies have investigated ‘patchy redness’, ‘mottled pattern’ and ‘mosaic redness’. For the purposes of this analysis, we included studies which had descriptions or images of mucosa satisfying the description of ‘map-like redness’ as an abnormal, irregular erythematous pattern. It was therefore interesting to note that this homogenised definition produced a strong predictor for H pylori-eradicated status, with a specificity of 99% and DOR of 12.2. Various studies have investigated these appearances and suggested the ‘map-like redness’ may correlate with atrophy or intestinal metaplasia. For the purposes of simplifying the prediction of H pylori status, the distinction may be of lesser importance, but the appearance may also suggest a target region for biopsy, for increasing the yield of further histological assessment.

The other finding predictive of an H pylori-eradicated status was found to be atrophic mucosa (DOR 4.0), although the diagnostic performance was inferior to that of map-like redness. This is likely to reflect the inherent difficulty in predicting gastric atrophy by endoscopic observation, despite attempts to describe atrophic appearances. Prediction of gastric atrophy is complex and requires a high level of endoscopist experience.

Gastric erosions and haem flecks
Kamada and colleagues report that raised or flat erosions, and haem flecks may be seen to a greater or lesser frequency in each of the H pylori states.
Raised erosions on mucosal folds are thought to signify chronic inflammation, and may be associated with 
\( \text{H pylori} \) infection, or drugs such as non-steroidal anti-inflammatory drugs (NSAIDs).\(^5^9\) Flat erosions are generally \(<5 \text{ mm in diameter}\) and \(<1 \text{-mm depth}\), and contain a fibrin exudate, and sometimes haematin.\(^1^9\) Previous studies have discovered equivocal results for these findings as predictors of \( \text{H pylori} \) status,\(^1^9,2^2,4^2\) and the results of this analysis agree with this, suggesting that erosions or haem flecks do not carry predictive significance.

**Predictive classification models**

The recent work by Yoshii and colleagues to validate the Kyoto classification system has developed a prediction model for the diagnosis of \( \text{H pylori} \) on the basis of endoscopic findings. The proposed system is a two-stage model, which initially classifies patients into either ‘noninfection’ or ‘past and current infection’ and then divides the second group into ‘past’ and ‘current’ infection.\(^2^5\) This model was trained by machine learning techniques using endoscopic information reporting the presence or absence of each of the 16 findings of the Kyoto classification. The model was initially able to achieve diagnostic accuracy of 88.6%, and when information regarding a history of previous \( \text{H pylori} \) eradication was added, this increased to 93.4%.

A similar approach could be applied in the development of a simplified system for endoscopic real-time \( \text{H pylori} \) status classification. Our findings suggest that identifying the RAC would be an appropriate way of stratifying patients into naive versus past or active infection because of its distinctive appearance and high DOR. Patients could then be further grouped using features such as diffuse redness or mucosal oedema to signify \( \text{H pylori} \)-positive status, and map-like redness to signify past infection. Forest plots for the diagnostic performance of these endoscopic findings are shown in Figures 6–9. Although such an
Diffuse Redness:

Figure 7. Forest plot of studies analysing the diagnostic performance of diffuse redness for predicting active *Helicobacter pylori* infection. CI, confidence interval.

Mucosal Oedema:

Figure 8. Forest plot of studies analysing the diagnostic performance of diffuse redness for predicting active *Helicobacter pylori* infection. CI, confidence interval.
approach would lack the finer detail of the full Kyoto classification, it could be relatively straightforward to learn and apply to routine practice.

**Application to endoscopic practice**

The recognition of subtle endoscopic features may become of use in diagnosing or stratifying patients to different categories of *H pylori* status. There are, however, some limitations encountered in this analysis. In particular, the included studies show a high degree of heterogeneity, which may limit the applicability of the results. This is in part due to the methodological differences between the studies analysed, and the absence of randomised controlled trials.

As with all approaches to endoscopic lesion recognition, there may potentially be a large component of heterogeneity due to interoperator variability. Only one of the included studies controlled for interoperator variability; Yoshii and colleagues\textsuperscript{25} recruited seven endoscopists who were formally educated on the Kyoto classification features before starting endoscopic examination. We suggest that future prospective studies in this field could include multiple endoscopists and include analysis of the effects of interoperator variability, and of any change in diagnostic performance related to training or experience.

**Conclusion**

The current era of high definition endoscopy with increasing access to image enhancement has redefined what can be assessed endoscopically. This, together with increased impetus to make endoscopic predictions of *H pylori* status, has stimulated research on the important mucosal findings. Work in Japan and areas with high incidence of gastric cancer has raised the expectations of gastroscopy reporting, with 16 features of interest, and as a consequence, the nature of the studies in this area has been changing, with recent studies investigating a large number of well-defined findings.
This analysis aims to synthesise the body of evidence accumulated in this area into a coherent whole, as an aid to informing future prediction models, and for planning future research. We propose that prediction models could take account of these aggregated diagnostic accuracy data and should consider the complexity of the diagnostic process; an approach which incorporates a large number of findings and variables may achieve high levels of accuracy but may not be practical to apply by the endoscopist as part of routine practice.

Future directions in this field should include large, prospective validation studies of evidence-based diagnostic models. It will be important to consider the ease of use of these approaches, and to ensure that results are reproducible in a general population; interoperator variability should be considered, as should endoscopists with different levels of experience. There are also other factors besides \textit{H pylori} that can influence the appearances of the stomach, and to maximise the generalisability of results, prediction systems should attempt to take into account factors such as the changing appearances of the RAC with increasing age, and the effects of medications such as NSAIDs or proton pump inhibitors.

**Author Contributions**
BG helped in study concept and design, acquisition of data, analysis and interpretation of data, drafting of the manuscript and statistical analysis; JT helped in study concept, technical guidance and revisions of manuscript; HA helped in acquisition of data, analysis and interpretation of data and statistical analysis; and NP helped in study concept and design, acquisition of data, interpretation of data and revisions of the manuscript.

**Conflict of interest statement**
The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

**Funding**
The authors received no financial support for the research, authorship and/or publication of this article.

**ORCID iDs**
Ben Glover  
https://orcid.org/0000-0003-3043-0012

Hutan Ashrafian  
https://orcid.org/0000-0003-1668-0672

**References**
7. Yagi K, Nakamura A and Sekine A. Comparison between magnifying endoscopy and histological, culture and urease test findings from the gastric mucosa of the corpus. \textit{Endoscopy} 2002; 34: 376–381.


**Appendix 1**

**Search strategy for MEDLINE and Embase:** search performed on 2 October 2019

1. Helicobacter pylori/
2. Helicobacter pylori.mp. (mp = title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms)
3. HP.mp. (mp = title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms)
4. Gastritis.mp. (mp = title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms)
5. RAC.mp. (mp = title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms)
6. Regular Arrangement.mp. (mp = title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms)
7. 1 or 2 or 3 or 4 or 5 or 6
8. Endoscopy/
9. Narrow Band.mp. (mp = title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms)
10. NBI.mp. (mp = title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms)
11. Narrowband.mp. (mp = title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms)
12. i-scan.mp. (mp = title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms)
13. iscan.mp. (mp = title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms)
14. optical biopsy.mp. (mp = title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms)
15. chromoendoscopy.mp. (mp = title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms)
16. image enhanc*.mp. (mp = title, abstract, original title, name of substance word,
subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms)

17. IEE.mp. (mp = title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms)

18. high definition.mp. (mp = title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms)

19. HD.mp. (mp = title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms)

20. 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19

21. 7 and 20