www.medscape.com

Serology Is More Sensitive Than Urea Breath Test or Stool Antigen for the Initial Diagnosis of *Helicobacter pylori* Gastritis When Compared With Histopathology

Dustin E. Bosch, MD, PhD; Niklas Krumm, MD, PhD; Mark H. Wener, MD; Matthew M. Yeh, MD, PhD; Camtu D. Truong, MD; Deepti M. Reddi, MD; Yongjun Liu, MD; Paul E. Swanson, MD; Rodney A. Schmidt; Andrew Bryan, MD, PhD

Am J Clin Pathol. 2020;154(2):255-265.

Abstract and Introduction

Abstract

Objectives: To assess the concordance and performance characteristics of *Helicobacter pylori* laboratory tests compared with histopathology and to propose algorithms for the diagnosis of *H pylori* that minimize diagnostic error.

Methods: *H pylori* diagnostics were reviewed from a 12-year period within a health system (2,560 cases). Analyses were performed to adjust diagnostic performance based on treatment and consensus histopathologic diagnoses among pathologists. Markers of access to care, including test cancellation frequency and turnaround time, were assessed. Costs and performance of candidate noninvasive testing algorithms were modeled as a function of disease prevalence.

Results: Serum *H pylori* IgG demonstrated a higher sensitivity (0.94) than urea breath and stool antigen tests (0.64 and 0.61, respectively). Evidence of an advantage in access to care for serology included a lower cancellation rate. Interobserver variability was higher (κ = 0.34) among pathologists for cases with a discordant laboratory test than concordant cases (κ = 0.56). A model testing algorithm utilizing serology for first-time diagnoses minimizes diagnostic error.

Conclusions: Although *H pylori* serology has modestly lower specificity than other noninvasive tests, the superior sensitivity and negative predictive value in our population support its use as a noninvasive test to rule out *H pylori* infection. Reflexive testing with positive serology followed by either stool antigen or urea breath test may optimize diagnostic accuracy in low-prevalence populations.

Introduction

Helicobacter pylori infection is common, with serologic prevalence exceeding 50% worldwide.[1] However, there is considerable variation by location, for example, with prevalence of approximately 35% in the United States but as high as 88% in Nigeria.[1] Within the United States, the prevalence of *H pylori* is related to geographic region and ethnicity, among other factors.[1,2] Infection with *H pylori* is associated with gastric ulcers, adverse outcomes in bariatric surgery patients, and neoplasia including adenocarcinoma and gastric marginal zone lymphoma.[3] Eradication is effective in preventing malignancy and can reverse nonneoplastic lesions.[4–7]

Diagnostic tests for *H* pylori can be categorized as invasive/tissue-based or noninvasive. Invasive tests include mucosal biopsy with histopathologic examination and *H pylori* culture. Histopathologic evaluation with routine H&E stain has a reported sensitivity of 70% to 95%[8] that can be increased by special stains such as the silver-based Genta or *H pylori* immunohistochemistry (IHC).[8] The morphology of *H pylori* and thus the performance of histopathology is affected by treatment, including proton pump inhibitors (PPIs).[8]*H pylori* culture is highly specific and may be considered the gold standard for definitive active infection, but the organism is fastidious and the sensitivity of culture is low.[9,10]

Commonly used noninvasive tests include the stool antigen test, the urea breath test (UBT), and *H pylori* serum IgG.[10] Stool antigen testing is performed by ELISA or lateral flow chromatography.[11] The UBT involves administration of either ^{14}C- or ^{13}C-labeled urea.[12] After a waiting period, the breath specimen is captured in a bag and assessed by scintigraphy, spectrophotometry, or mass spectrometry. The stool antigen test and UBT perform similarly in direct comparison studies,[12,13] and both suffer from decreased sensitivity in the context of acid reduction therapy and recent antibiotic exposure.[14]*H pylori* IgG serology is a highly sensitive test unaffected by PPI or antibiotic therapy, but active infection cannot be distinguished from past exposure.[8,15] Multiple *H pylori* serology kits are available, and the performance of some are related to patient factors including age, sex, and ethnicity.[16,17] Thus, appropriate kit selection and local validation of performance are important. The limitations of *H pylori* serology have prompted recommendations to avoid its use (eg, Choosing Wisely Campaign).[18] The American College of Gastroenterology's 2017 guidelines similarly recommend tests that detect active infection for most presentations, except if the pretest probability is high (eg, peptic ulcer disease) and serology may be acceptable.[19]

There are clear clinical benefits of *H pylori* eradication. Guidelines have tended to expand recommendations for diagnostic testing to broader populations over time and have recommended routinely confirming successful *H pylori* eradication more than 4 weeks after therapy.[19] Multimodal testing creates complexity of interpretation but also presents an opportunity to retrospectively compare test performance and concordance in a real-world clinical population. To our knowledge, we report the largest data set comparing invasive and noninvasive *H pylori* diagnostic tests and model diagnostic errors and costs of noninvasive testing algorithms.

Materials and Methods

Case Selection

This study was approved by the University of Washington institutional review board (IRB STUDY00003621). All *H pylori* laboratory tests ordered at the University of Washington and Harborview Medical Centers from 2005 to 2017 and from Northwest Hospital and Medical Center from April 2016 through 2017 were included in the study, as obtained through the laboratory information system. Noninvasive *H pylori* tests included serum IgG, stool antigen test, and UBT. Invasive testing included collection of gastric biopsies for *H pylori* culture. Gastric biopsies within 1 year of laboratory testing were identified using the institutional pathology database. Genta silver-based stain and *H pylori* IHC were utilized in the majority of cases.

Laboratory Test Performance

UBT was performed by spectrophotometry of ^{13}C-labeled urea (BreathTek UBT) in an outpatient phlebotomy center. At the time of collection, laboratory staff screened patients for contraindications including antibiotics, PPIs, or bismuth agents within 2 weeks. *H pylori* culture was performed on biopsies processed ideally within 3 hours, and culture incubated on chocolate and Brucella agar (microaerophilic) and a urea slant (ambient) at 35 °C for 7 days. Serum IgG for *H pylori* was performed by 2 manual ELISAs, with transition from the Trinity *H pylori* assay to the Inova QUANTA Lite assay after 2013. Antigens are proprietary. The assays are qualitative, as approved by the US Food and Drug Administration, but yield optical density ratios. For analyses pertaining to the change in IgG over time, the optical density ratio was investigated as a function of time and referred to as a "titer" for brevity, without connoting a correlation to traditional 1:2 serological dilutions. *H pylori* stool antigen testing was sent to the Mayo Clinic Laboratories, with performance currently on the HpSA microwell-based ELISA (Meridian Biosicence). Counseling patients on the contraindications for stool antigen testing at our institutions is primarily the responsibility of the ordering provider.

Histopathology Review

Sixty gastric biopsies with H&E stain, together with either Genta stain or *H pylori* IHC slides, were deidentified and independently reviewed by 2 gastrointestinal pathologists and 1 experienced gastrointestinal pathology fellow who were blind to the initial report. Cases were randomly selected within each of 3 groups based on concordance with noninvasive *H pylori* tests: 23 biopsies concordant with stool antigen, UBT, or serology; 23 discordant; and 14 interpreted as equivocal at the time of clinical diagnosis. *H pylori* status was scored as positive, negative, or equivocal by each independent observer.

Data Analysis

Demographic characterization, time trending, receiver operating characteristic (ROC) analysis, cancellation rates, turnaround time, and correlation of laboratory data with histopathology were performed with RStudio and Prism 7 (GraphPad). Pathology reports were reviewed manually and interpreted based on the diagnostic line and ancillary studies (Genta stain, *H pylori* IHC). Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were derived from 2 × 2 tables (equivocal serology and/or biopsy excluded), treating the biopsy result as the reference standard. Discordant biopsy and laboratory pairs explained by treatment or prior infection were estimated by random chart review (20-50 false-positive and false-negative cases per test) and excluded from the test performance analysis. Inpatient medication administration and outpatient prescription data including therapeutics used in *H pylori* eradication and acid suppression were obtained from an institutional data warehouse, derived from the electronic health record (EHR) system.

Statistical Analysis

Statistical analyses were performed with RStudio and Prism 7. Comparison of cancellation rates was performed with χ^2 tests. Sampling error associated with random chart review of discordant biopsy and laboratory result cases was calculated as $se = \sqrt{\frac{p(1-p)}{n}}$, where p is the proportion explained by treatment, and n is the number of charts assessed. Sampling error was propagated through calculations of sensitivity, specificity, PPV, and NPV. Interobserver agreement was assessed with the Fleiss κ. To assess test performance accounting for the imperfect reference test (histopathology), calculations of sensitivity, specificity, PPV, NPV, and associated standard errors were also performed using the consensus diagnosis among slide reviewers as an "imperfect resolver" test by previously described methods.[20]

Testing Algorithm Modeling

Diagnostic accuracy and cost were modeled as functions of disease prevalence. Assumptions included constancy of test sensitivity and specificity, absence of a prior known infection or positive *H pylori* test, and laboratory cost for each test fixed at the 2018 Medicare reimbursement rate or internal actual laboratory cost estimates (inclusive of reagents and labor). The model for "2017 utilization pattern" assumes 1 diagnostic test per diagnosis and test utilization similar to the year 2017 at our institution. The diagnostic error rate is defined as the sum of false positives plus false negatives, divided by the total number of patients.

Cost per diagnosis was assessed by 3 different metrics: (1) the product of test number and Medicare reimbursement, divided by the sum of true positives and true negatives; (2) the product of test number and laboratory cost estimate, divided by the sum of true positives and true negatives; and (3) test number multiplied by the difference of laboratory cost estimate minus Medicare reimbursement, divided by the sum of true positives and true negatives).

Results

H pylori laboratory tests were frequently utilized within our multiple sites of practice and increased over time Figure 1A. Positivity rates across tests ranged from 14% to 21%, indicating the prevalence in our population. As overall test utilization increased over time, more patients underwent both gastric biopsy and laboratory testing Figure 1B and . Common scenarios included esophagogastroduodenoscopy after noninvasive *H pylori* testing in patients with "alarm symptoms" such as weight loss, dysphagia, or iron deficiency anemia or a noninvasive test for confirmation of eradication after positive histopathology and treatment.[21]

Table 1. Characteristics of Subjects Having Biopsy and Laboratory Test Within 90 Days

No. of cases	1,653
No. (%) biopsy positive	683 (41)
% Female	55
Mean age (quartiles), y	52 (39–63)
Mean time interval (quartiles), d	31 (5–57)

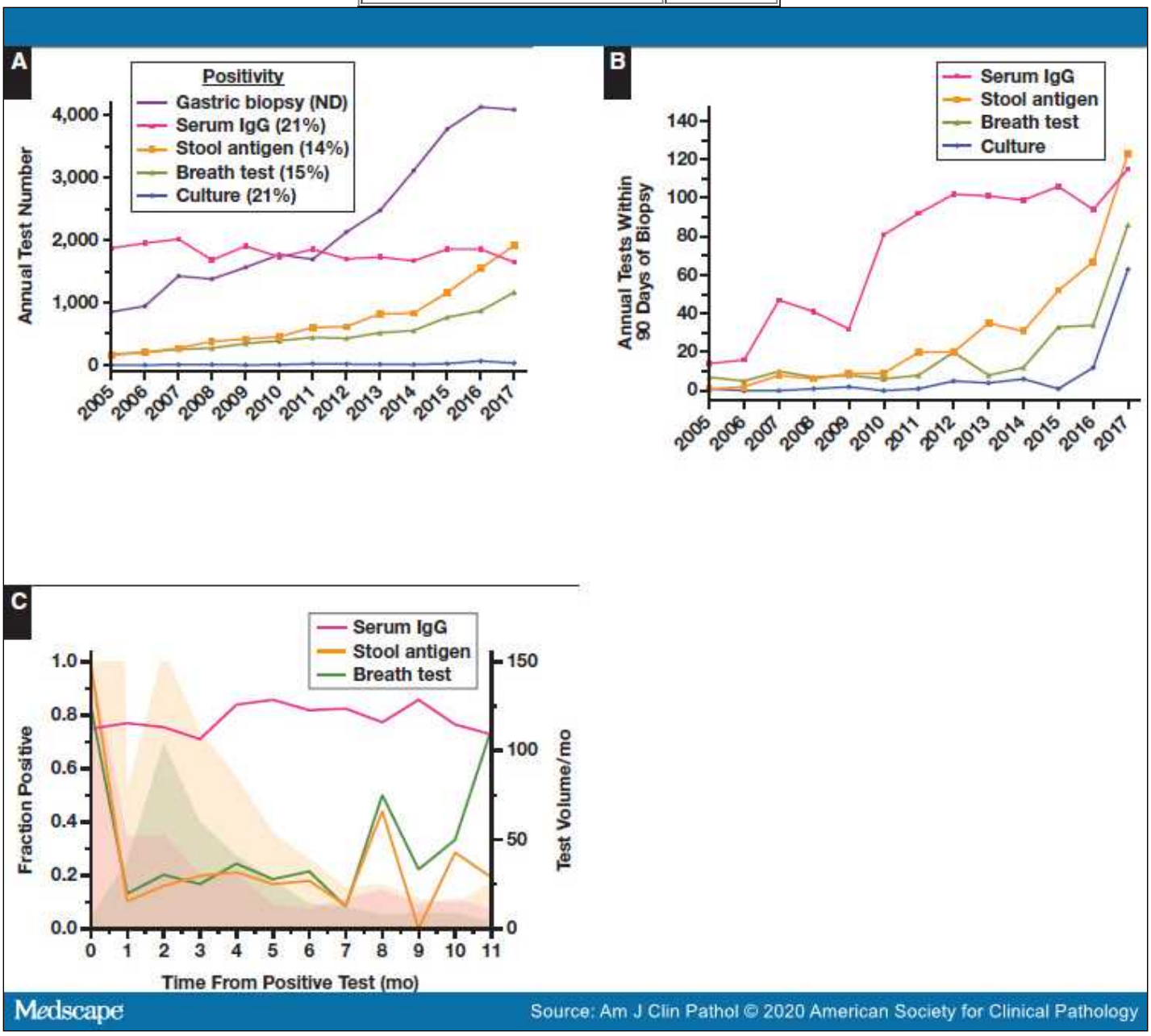


Figure 1.

Utilization of *Helicobacter pylori* laboratory tests in combination with gastric biopsy. **A**, Annual noninvasive test and gastric biopsy volumes increased, as did multimodal testing (**B**). **C**, Noninvasive laboratory testing was assessed after an initial positive biopsy, stool antigen, or culture. Util ization (fill) peaked at approximately 2 months, reflecting testing for confirmation of eradication. Positivity rates (lines) of stool antigen and urea breath test drop sharply due to treatment, whereas *H pylori* serology is unaffected.

Of the 2,560 patients with gastric biopsy and laboratory testing within 1 year, concordance of biopsy and noninvasive test results was related to the time interval between tests, which likely reflects testing to confirm eradication after therapy. For patients with an initial positive biopsy, stool antigen, or *H pylori* culture, approximately 80% of subsequent stool antigen and UBTs were negative Figure 1C. This pattern, and a peak in repeat testing at 2 to 3 months, reflects the common practice of treatment followed by confirmation of eradication.[22] As expected, a consistently high proportion of patients with diagnosed infection remained seropositive over time.

To assess noninvasive laboratory test and biopsy concordance with minimal confounders, results within 14 days were compared Figure 2A. Because most antibiotic regimens are prescribed for 14 days, completion of therapy and eradication of infection

between biopsy and the noninvasive laboratory test would be minimized in this cohort. Case numbers limited assessment of UBT performance compared with biopsy, but a notable sensitivity advantage (0.92 compared with 0.33–0.63) and corresponding higher NPV (0.97 compared with 0.64–0.90) were identified for *H pylori* serology (Figure 2A and Supplemental Figure 1; all supplemental material can be found at American Journal of Clinical Pathology online). Because stool antigen and culture exhibited essentially 100% specificity, an additional sensitivity analysis was performed for tests within 14 days of positive biopsy, stool antigen, or culture (). With additional test numbers, serology still exhibited a substantial sensitivity advantage (0.90) over stool antigen and UBT (0.70–0.71). As previously described,[15] the specificity (0.82) and corresponding PPV (0.64) of *H pylori* serum IgG were lower than those of stool antigen and *H pylori* culture and was likely explained by the inability of serology to distinguish between active infection and past exposure.[8] A possible confounding effect of histopathologic interpretation biased by already available laboratory results was evaluated. Among the cohort with biopsy and noninvasive test within 14 days, 49% had laboratory results preceding biopsy, and there was no correlation to biopsy/noninvasive test concordance ($\chi^2 = 0.8$, $P = .4$).

Table S1. ***Sensitivity analysis of H. pylori tests compared to positive histopathology, stool antigen test, or culture.*** **Results of laboratory tests within 14 days of a positive test were tabulated. Sensitivity patterns were similar to the comparison to biopsy only (Figure 1), although sensitivities of stool antigen test and urea breath test were slightly higher with increased case numbers.**

	Positive biopsy, stool antigen, or culture	Sensitivity
Serum *H. pyori* IgG		0.90
Positive	149	
Equivocal	10	
Negative	16	
***H. pylori* stool antigen**		0.70
Positive	14	
Negative	6	
***H. pylori* culture**		0.46
Positive	21	
Negative	25	
Urea breath test		0.71
Positive	5	
Negative	2	

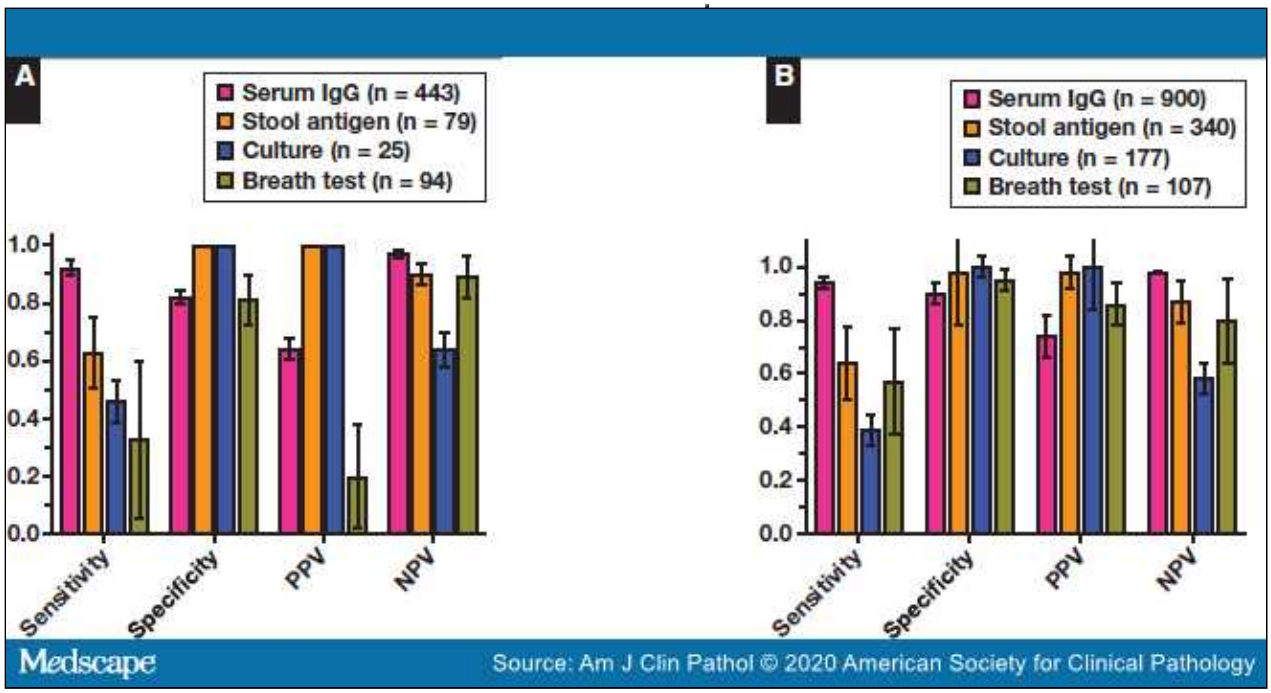
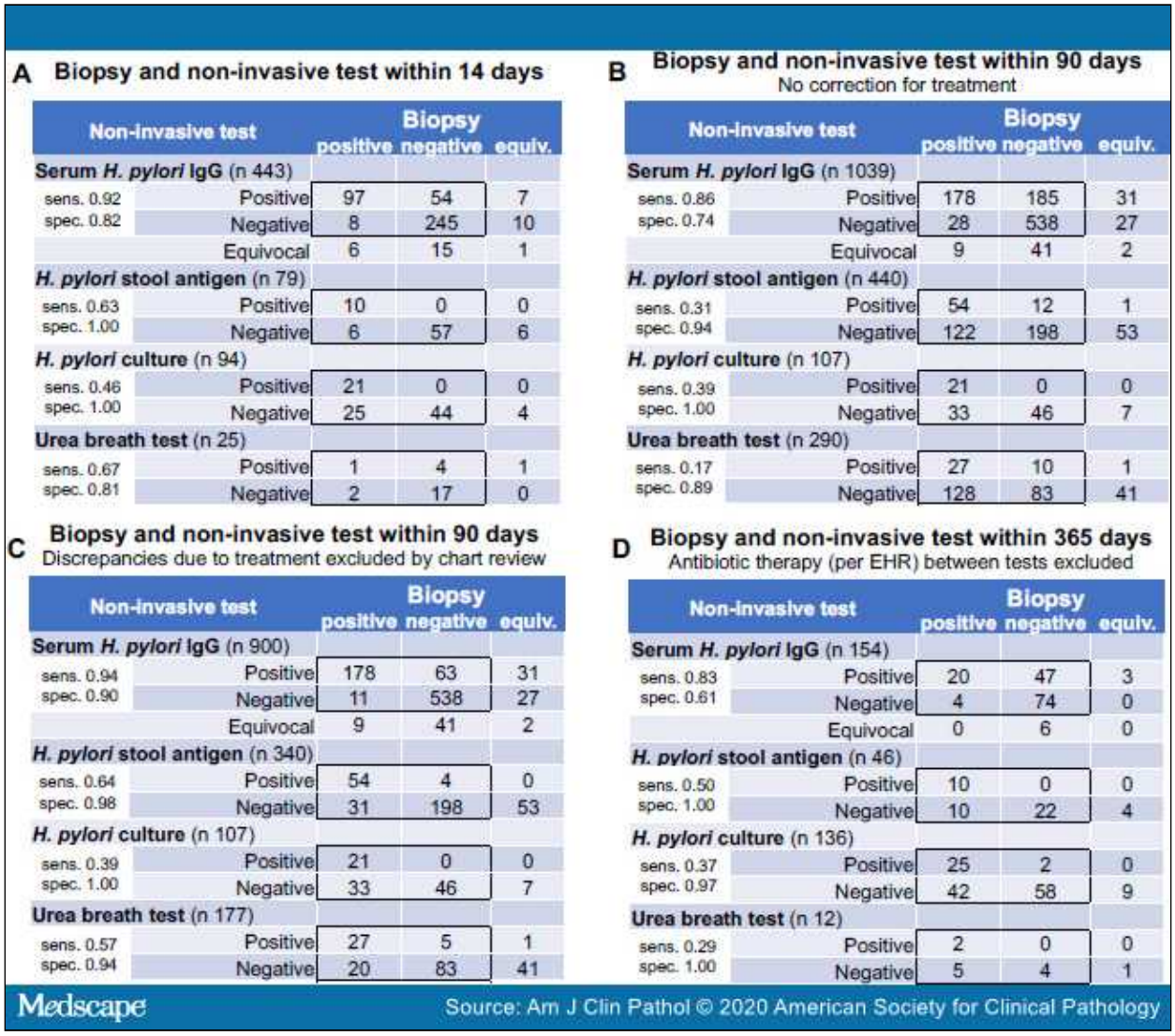


Figure 2.

Performance of laboratory tests compared with biopsy revealed superior sensitivity of serology. **A**, Tests within 14 days of biopsy showed highest sensitivity and lowest specificity for serology, using histopathology as the standard. **B**, Tests within 90 days of biopsy were correlated, with exclusion of treatment-related discordance by randomized chart review. Error bars represent sampling error. NPV negative predictive value; PPV, positive predictive value.



A Biopsy and non-invasive test within 14 days

Non-invasive test		Biopsy positive	Biopsy negative	Biopsy equiv.
Serum *H. pylori* IgG (n 443)				
sens. 0.92	Positive	97	54	7
spec. 0.82	Negative	8	245	10
	Equivocal	6	15	1
***H. pylori* stool antigen** (n 79)				
sens. 0.63	Positive	10	0	0
spec. 1.00	Negative	6	57	6
***H. pylori* culture** (n 94)				
sens. 0.46	Positive	21	0	0
spec. 1.00	Negative	25	44	4
Urea breath test (n 25)				
sens. 0.67	Positive	1	4	1
spec. 0.81	Negative	2	17	0

B Biopsy and non-invasive test within 90 days
No correction for treatment

Non-invasive test		Biopsy positive	Biopsy negative	Biopsy equiv.
Serum *H. pylori* IgG (n 1039)				
sens. 0.86	Positive	178	185	31
spec. 0.74	Negative	28	538	27
	Equivocal	9	41	2
***H. pylori* stool antigen** (n 440)				
sens. 0.31	Positive	54	12	1
spec. 0.94	Negative	122	198	53
***H. pylori* culture** (n 107)				
sens. 0.39	Positive	21	0	0
spec. 1.00	Negative	33	46	7
Urea breath test (n 290)				
sens. 0.17	Positive	27	10	1
spec. 0.89	Negative	128	83	41

C Biopsy and non-invasive test within 90 days
Discrepancies due to treatment excluded by chart review

Non-invasive test		Biopsy positive	Biopsy negative	Biopsy equiv.
Serum *H. pylori* IgG (n 900)				
sens. 0.94	Positive	178	63	31
spec. 0.90	Negative	11	538	27
	Equivocal	9	41	2
***H. pylori* stool antigen** (n 340)				
sens. 0.64	Positive	54	4	0
spec. 0.98	Negative	31	198	53
***H. pylori* culture** (n 107)				
sens. 0.39	Positive	21	0	0
spec. 1.00	Negative	33	46	7
Urea breath test (n 177)				
sens. 0.57	Positive	27	5	1
spec. 0.94	Negative	20	83	41

D Biopsy and non-invasive test within 365 days
Antibiotic therapy (per EHR) between tests excluded

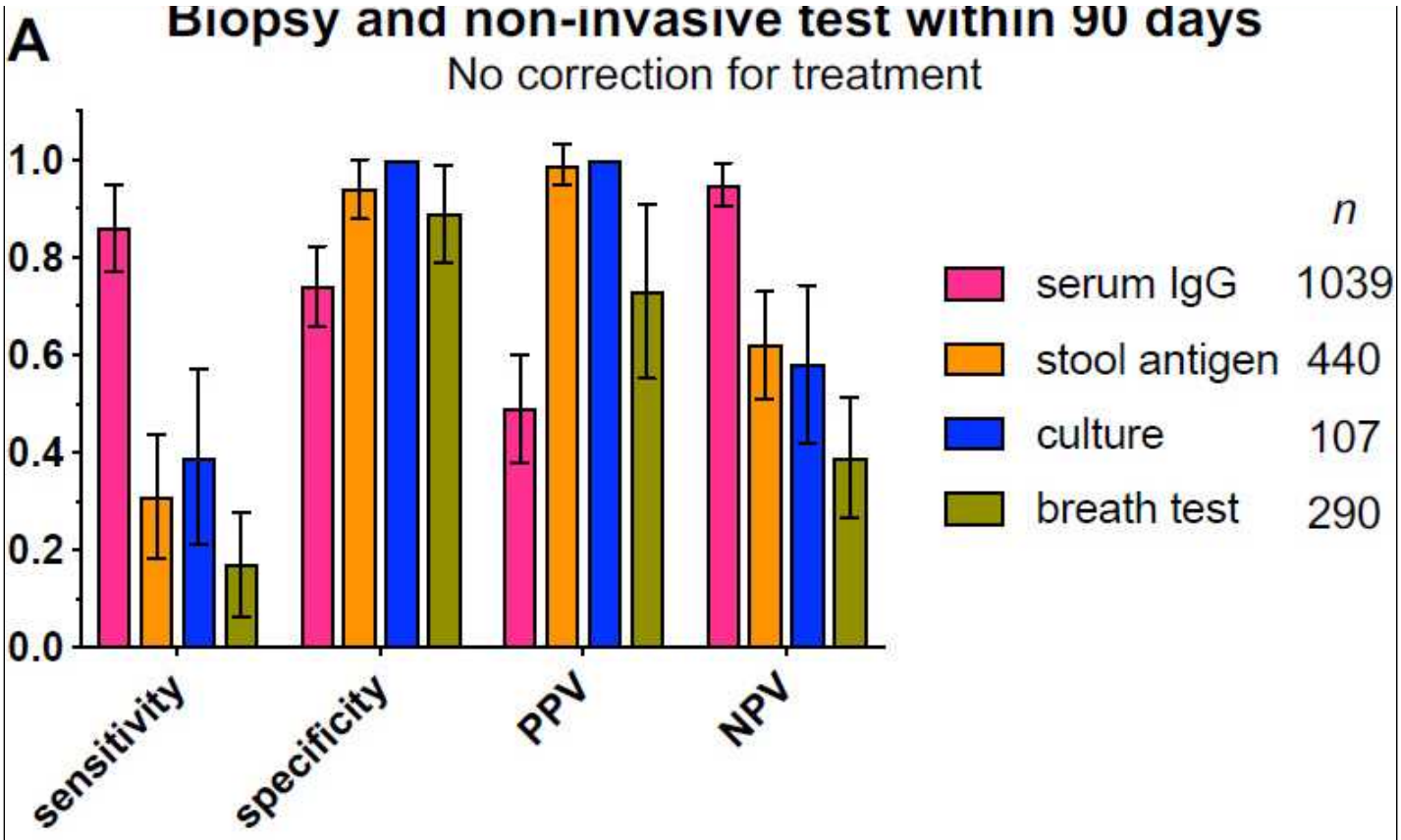
Non-invasive test		Biopsy positive	Biopsy negative	Biopsy equiv.
Serum *H. pylori* IgG (n 154)				
sens. 0.83	Positive	20	47	3
spec. 0.61	Negative	4	74	0
	Equivocal	0	6	0
***H. pylori* stool antigen** (n 46)				
sens. 0.50	Positive	10	0	0
spec. 1.00	Negative	10	22	4
***H. pylori* culture** (n 136)				
sens. 0.37	Positive	25	2	0
spec. 0.97	Negative	42	58	9
Urea breath test (n 12)				
sens. 0.29	Positive	2	0	0
spec. 1.00	Negative	5	4	1

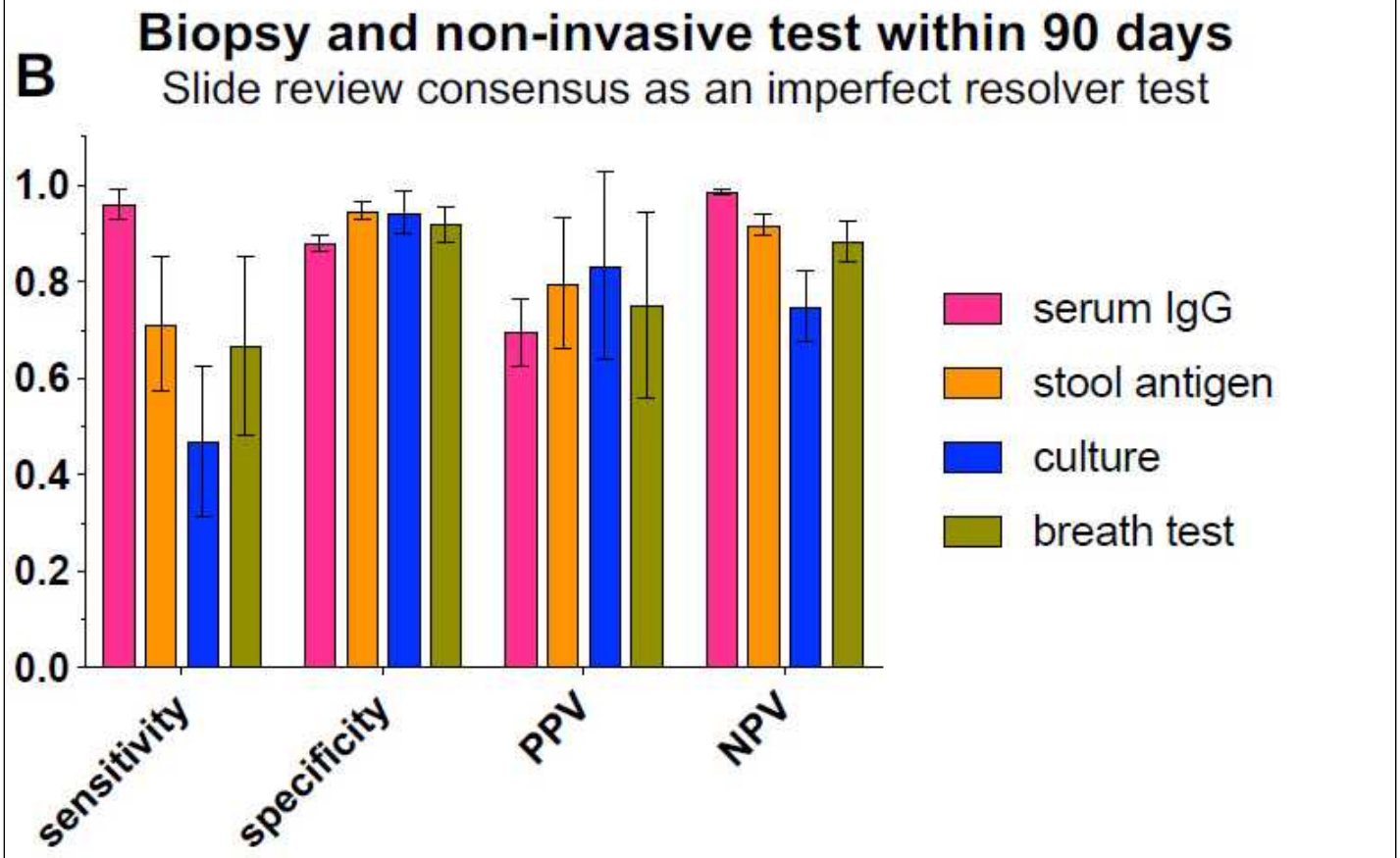
Medscape Source: Am J Clin Pathol © 2020 American Society for Clinical Pathology

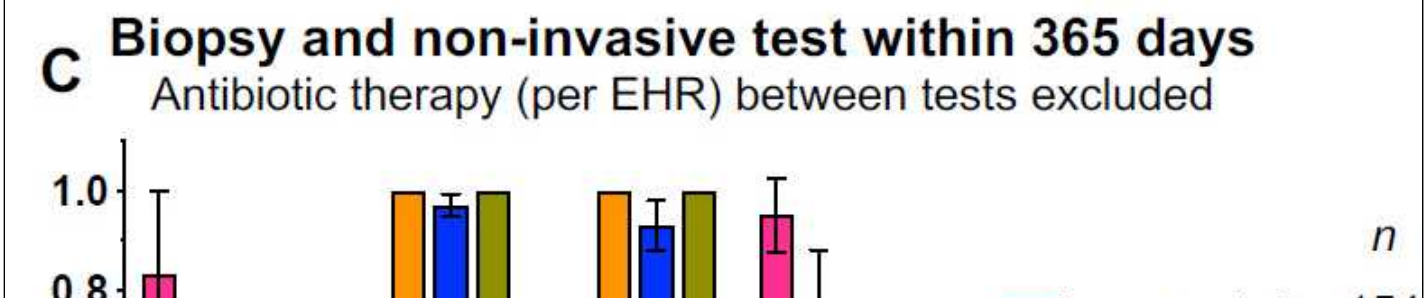
Figure S1.

Numeric data for performance of non-invasive H. pylori tests with biopsy as the reference standard. Contingency tables for non-invasive tests and histopathology correspond to graphical representations of test performance in Figures 1, S2. Analyses include biopsy and non-invasive test within 14 days (A), biopsy and non-invasive test within 90 days without (B) and with (C) correction for treatment effect by randomized chart review, and biopsy and non-invasive test within a year corrected for treatment effect by evidence of antibiotic therapy derived from the electronic health record (EHR) (D). Black boxes correspond to classical 2 x 2 tables. Sensitivity (sens.) and specificity (spec.) are calculated by exclusions of equivocal biopsy and serology results.

Concordance of biopsy and laboratory testing within 90 days was also assessed Figure 2B (Supplemental Figures 1 and 2). Two methods of correction for treatment effects were used. First, discordant results explained by treatment or known prior infection, as assessed by random chart review, were excluded (Figure 2B). A similar performance trend was identified, with serum IgG having higher sensitivity (0.94 compared with 0.39–0.64) and modestly lower specificity (0.90) than the other laboratory tests (0.95–1.00). Although review of all charts was not feasible, sampling error related to the randomized chart review was calculated, and its propagation in the calculation of performance characteristics is reflected in the error bars of Figure 2B. To corroborate the performance estimates based on randomized chart review, discordant cases with antibiotics administered or prescribed between biopsy and noninvasive *H pylori* test were excluded from the performance assessment (Supplemental Figure 2), independent from the chart-review assessment. Important limitations to this approach include the absence of treatment data outside of our institution and inability to distinguish *H pylori* eradication therapy from other antibiotic indications. A similar trend appeared in this analysis, with superior sensitivity and mildly inferior specificity of serology compared with stool antigen, UBT, and *H pylori* culture (Supplemental Figure 2).


A
Biopsy and non-invasive test within 90 days
No correction for treatment
1.0
0.8
0.6
0.4
0.2
0.0
sensitivity
specificity
PPV
NPV
n
serum IgG 1039
stool antigen 440
culture 107
breath test 290


B
Biopsy and non-invasive test within 90 days
Slide review consensus as an imperfect resolver test
1.0
0.8
0.6
0.4
0.2
0.0
sensitivity
specificity
PPV
NPV
serum IgG
stool antigen
culture
breath test


C
Biopsy and non-invasive test within 365 days
Antibiotic therapy (per EHR) between tests excluded
1.0
0.8
n

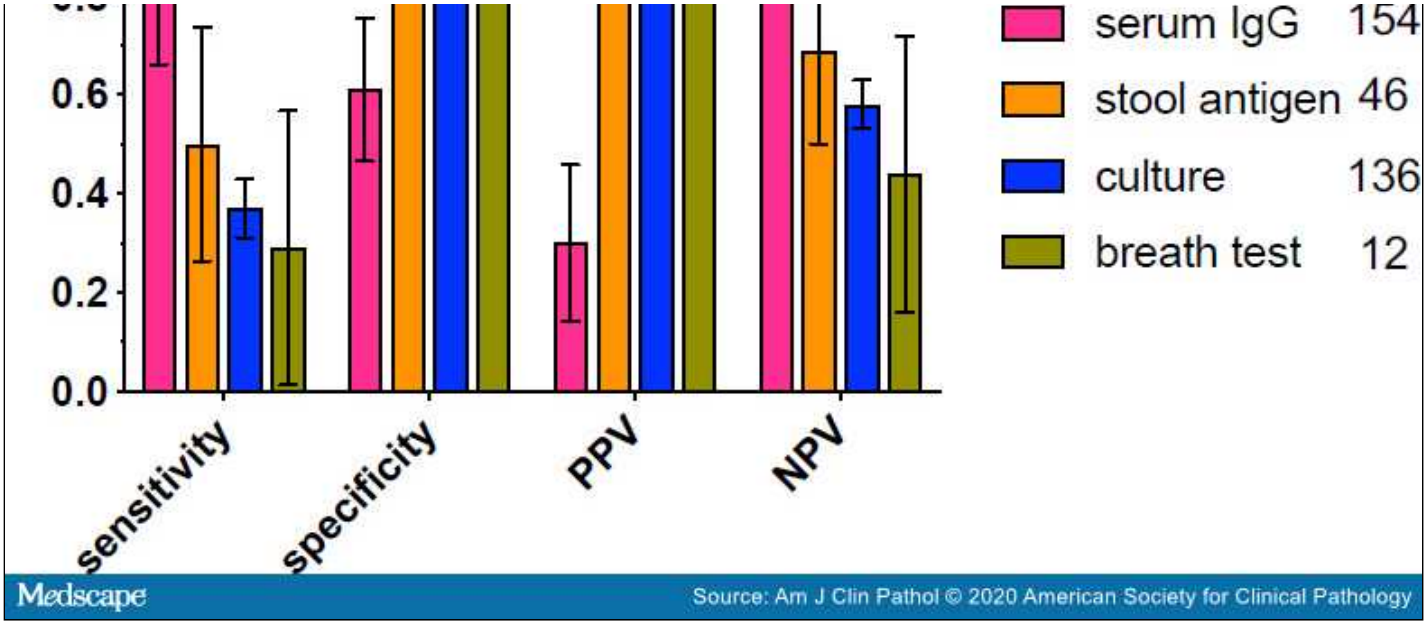


Figure S2.

Performance of non-invasive H. pylori tests, with slide review as a resolver and corrected for antibiotic therapy. Performance data for biopsy and non-invasive tests within 90 days (A) was adjusted using slide review consensus as an imperfect resolver test (B) (20). (C) Non-invasive tests were correlated to histopathology for patients with documented inpatient medication administration between laboratory test and biopsy. Patients with antibiotic therapy between tests were excluded. Performance metrics were derived from raw data in Figure S1, and error bars represent sampling error. Abbreviations: PPV positive predictive value, NPV negative predictive value.

Agreement of biopsy and *H pylori* serum IgG was related to titer, as demonstrated with ROC analysis of 1 ELISA kit in use from 2005 to 2013 (Figure 3; area under the curve, 0.88). Some patients had multiple serum IgG tests, and titers were generally stable over time, although a modest statistically significant trend of titer decrease was detected (−0.0024 fold/mo) (Supplemental Figure 3). The titer trend was not significantly different in patients with evidence of successful treatment (an intervening negative stool antigen or UBT) than in those with evidence of persistent or recurrent infection (intervening positive stool antigen or UBT). Because of the relative stability of *H pylori* serum IgG over time, the utility of trending titers is low.

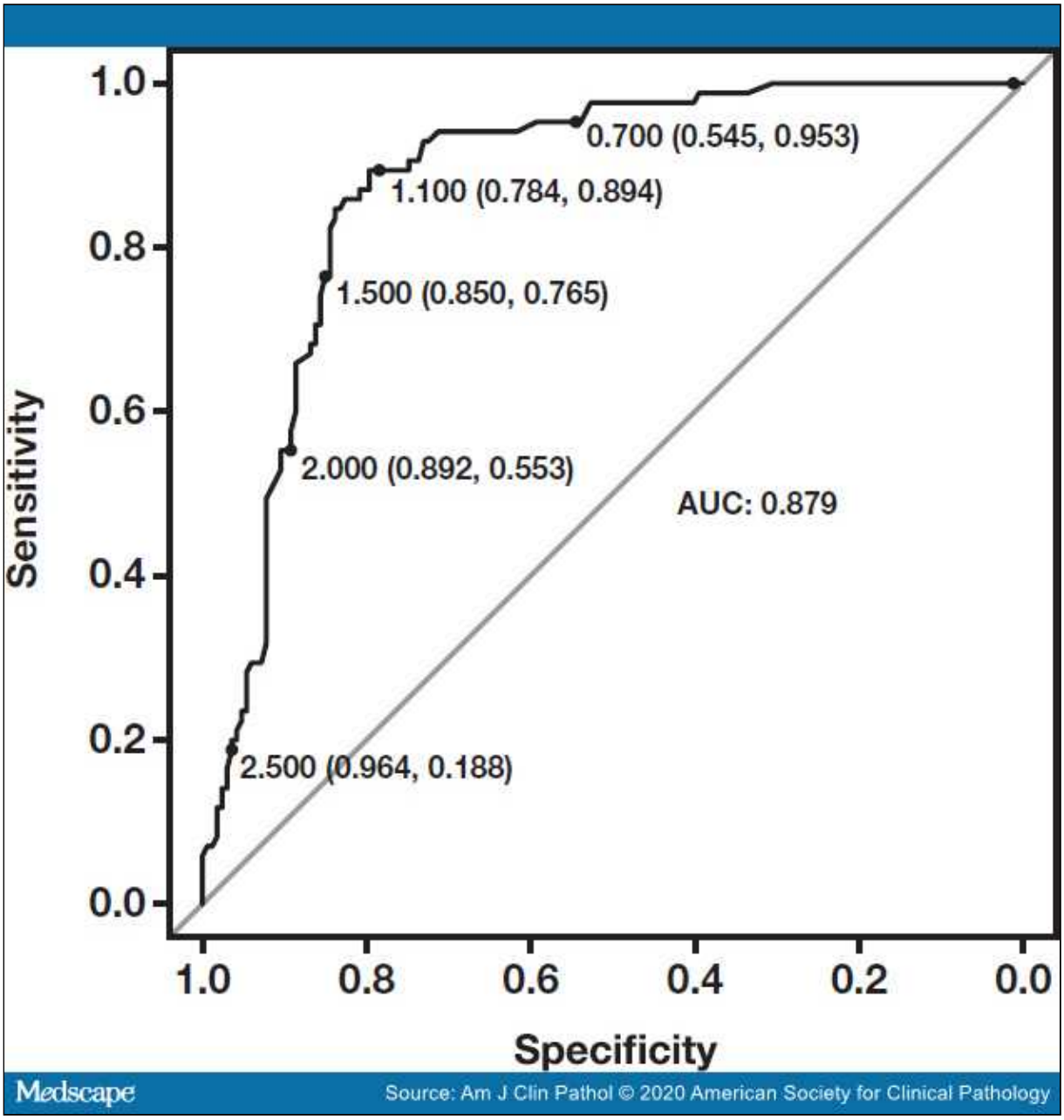


Figure 3.

Serum IgG titer correlated to biopsy result. Receiver operating characteristic analysis with points labeled as cutoff (specificity, sensitivity) demonstrated an area under the curve (AUC) of 0.88.

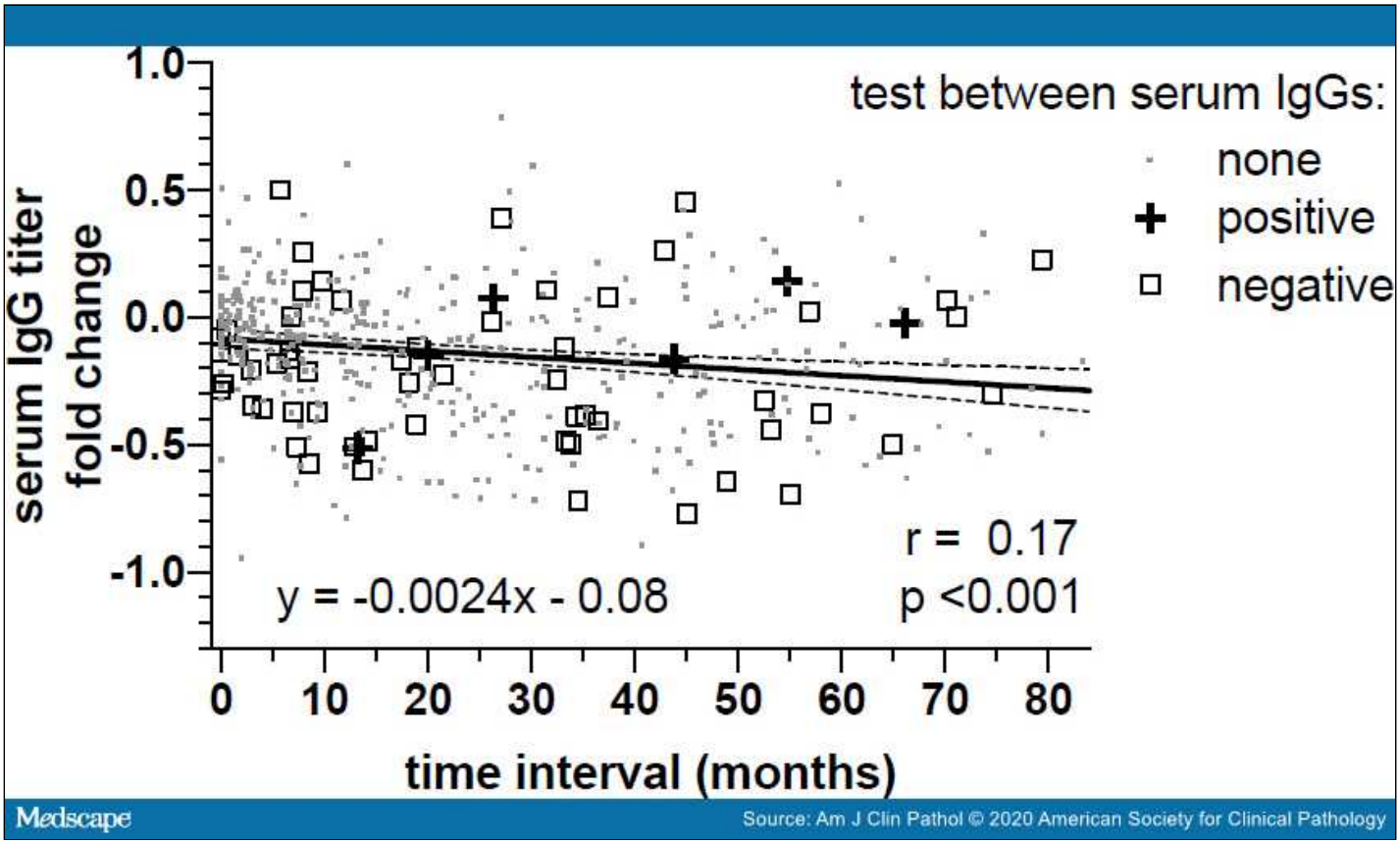


Figure S3.

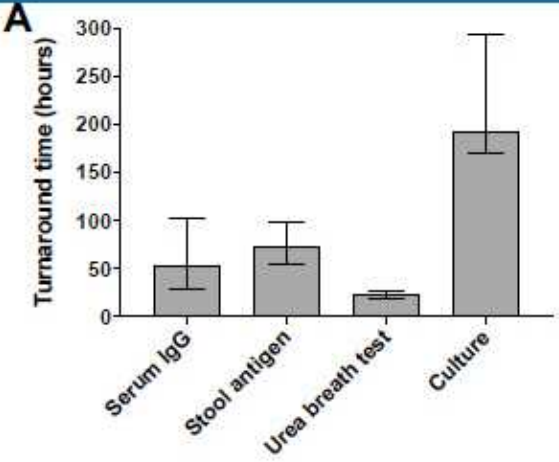
Serum IgG titer decreased slowly. Rate of titer change was not related to evidence of eradication (negative intervening test) or evidence of persistent/recurrent infection (positive intervening test). A very slow, but statistically significant decrease in titer over time was detected by linear regression.

H pylori serology is anecdotally perceived as a more convenient testing modality that can be performed by routine phlebotomy compared with a stool specimen or timed breath specimen collection. Availability of serology may theoretically improve access to care for some patients. Cancellation rates and turnaround time were thus assessed as measurable indicators of access to care and convenience. There were significantly more cancellations of stool antigen and culture than *H pylori* serum IgG (χ^2 test, $P <$ .05) . A breakdown of reasons for cancellation revealed significantly more improper specimen collections and fewer corrections by the ordering provider for the stool antigen test compared with the other laboratory tests (Supplemental Figure 4). Consistently, 25% to 30% of patients with any test cancellation had a subsequent *H pylori* test, with no significant differences between initial test ordered. The turnaround time for most tests was 1 to 3 days, except culture (approximately 7 days) (Supplemental Figure 4).

Table 2. Cancellation Rates of Noninvasive Helicobacter pylori Tests[a]

Test	Cancellation Rate (%)	*P* Value (χ^2)
Serum IgG	1.5	—
Stool antigen	2.5	<.001
Urea breath test	1.8	.05
Culture	3.3	.02

[a]Overall test cancellation rates were higher for stool antigen and culture relative to serology and breath test. Serum IgG was used as the comparator for χ^2 tests.



B

	Serum IgG n (%)	Stool antigen n (%)	Urea breath test n (%)	Culture n (%)
Total cancelations	406	242	118	7
Specimen collection	62 (15)	112 (46)	21 (18)	3 (43)
Laboratory correction	53 (13)	34 (14)	23 (19)	2 (29)
Ordering provider correction	87 (21)	22 (9)	40 (34)	2 (29)
Patient availability	12 (3)	1 (0)	2 (2)	0
Other / Unknown	192 (47)	73 (30)	32 (27)	0
Chi square p value	comparator	<0.001	<0.001	0.08

Medscape Source: Am J Clin Pathol © 2020 American Society for Clinical Pathology

Figure S4.

Laboratory test turnaround time and cancellations suggest relative convenience of urea breath testing and H. pylori serum IgG. (A) Turnaround times from specimen receipt to final result was shortest for the in-house urea breath test and longest for *H. pylori* culture. (B) Reasons for cancellation were classified according to annotation in the laboratory information system.

Review of histopathologic diagnostic reports in patients having noninvasive *H pylori* testing within 1 year identified approximately 5% of cases with an equivocal report (Supplemental Figure 1), for example, "rare immunoreactive forms suspicious for *H pylori*." There is likely enrichment for equivocal diagnoses in the study group because a nondefinitive biopsy interpretation may prompt alternative testing, and *H pylori* organism number and phenotype are frequently altered in the setting of treatment Image 1 and .[8] We hypothesized that additional biopsy/laboratory test-discordant cases may reflect treatment effect and challenging histopathology. Sixty gastric biopsies with H&E and either Genta stain or *H pylori* IHC (to maximize sensitivity), including 23 biopsy/laboratory concordant, 23 biopsy/laboratory discordant, and 14 biopsy-equivocal and laboratory-negative cases that were deidentified and independently reviewed by 3 pathologists. Interpretation of biopsy/laboratory-concordant cases had high agreement with the original diagnosis and substantial interobserver agreement (Image 1 and Fleiss κ = 0.56). Reviewer agreement with original diagnosis and interobserver concordance were lower for biopsy/laboratory-discordant cases and lowest for cases initially diagnosed as equivocal for *H pylori* gastritis. These findings suggest that equivocating on difficult histopathologic interpretations is appropriate because of poor agreement among observers and with noninvasive laboratory tests. However, the large majority of patients fall in the concordant biopsy/laboratory category, which is characterized by good agreement among observers.

Table 3. Pathologist Interobserver Agreement[a]

	Agreement With Initial Diagnosis, No./Total No. (%)	Fleiss κ
Overall (n = 60)	98/180 (54)	0.43
Concordant lab plus biopsy (n = 23)	56/69 (81)	0.56
Discordant lab plus biopsy (n = 23)	38/69 (55)	0.34
Equivocal biopsy (n = 14)	4/42 (10)	0.04

[a]Pathologist interobserver agreement is highest when histopathology and laboratory test are concordant. Agreement of slide review consensus with the original diagnosis and interobserver concordance were highest for cases with a concordant noninvasive test and lowest with an initial equivocal diagnosis.

Table 3. Pathologist Interobserver Agreement[a]

	Agreement With Initial Diagnosis, No./Total No. (%)	Fleiss κ

Overall (n = 60)	98/180 (54)	0.43
Concordant lab plus biopsy (n = 23)	56/69 (81)	0.56
Discordant lab plus biopsy (n = 23)	38/69 (55)	0.34
Equivocal biopsy (n = 14)	4/42 (10)	0.04

[a]Pathologist interobserver agreement is highest when histopathology and laboratory test are concordant. Agreement of slide review consensus with the original diagnosis and interobserver concordance were highest for cases with a concordant noninvasive test and lowest with an initial equivocal diagnosis.

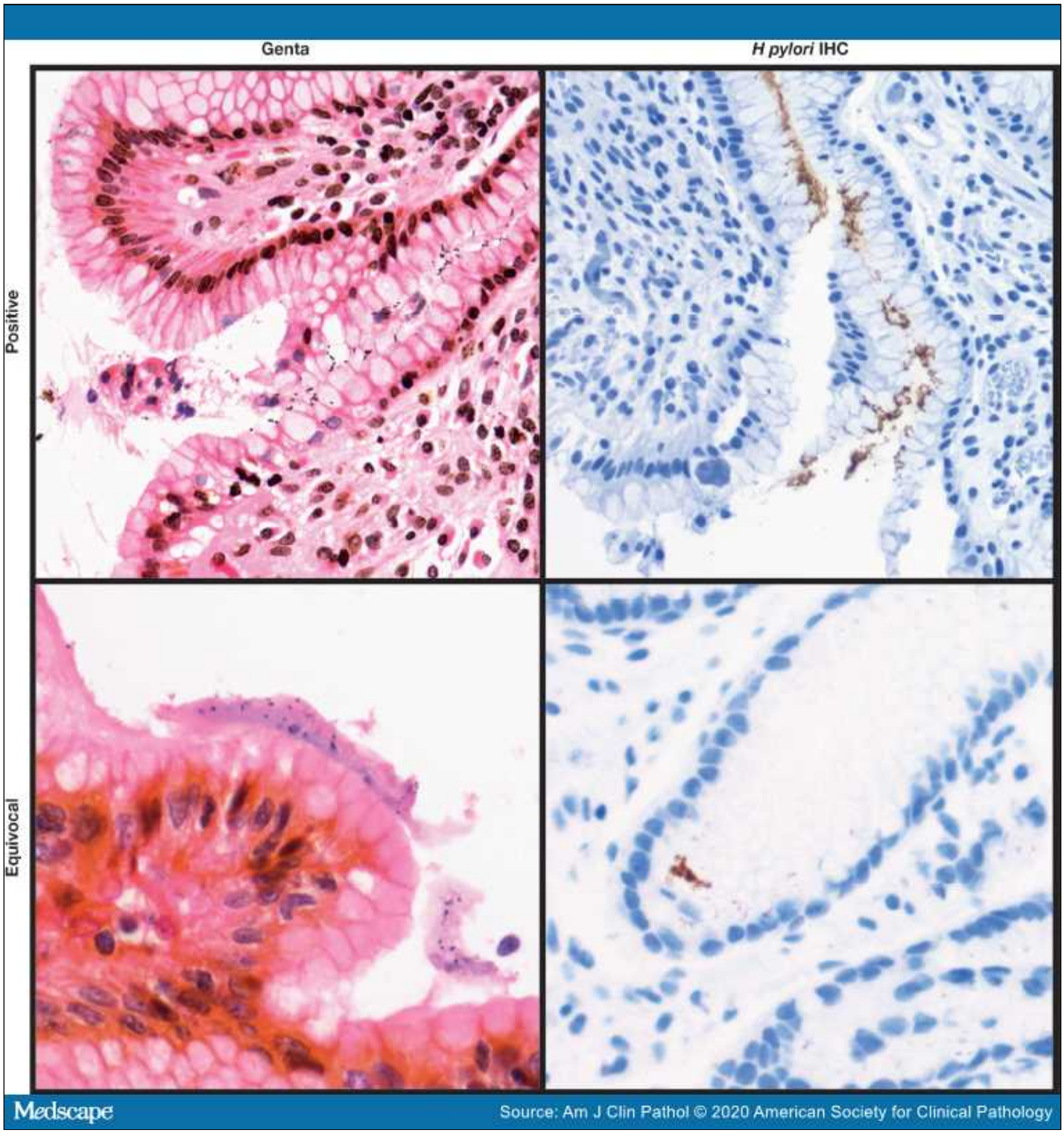


Image 1.

Histopathology with concurrent noninvasive laboratory testing. Coccoid forms at the epithelial surface and immunohistochemical staining of forms lacking the classic rod shape of *Helicobacter pylori*. Agreement of slide review consensus with the original diagnosis and interobserver concordance were highest for cases with a concordant noninvasive test and lowest with an initial equivocal diagnosis. IHC, immunohistochemistry.

The occasional disagreement of consensus slide review interpretations and the original histopathologic diagnoses reflects that histopathology, albeit the best available reference test, is not a perfect gold standard. Comparison to a reference standard is common practice, and the performance characteristics of the reference test may affect the apparent performance of the index tests in complex ways.[23] One way to account for error in the reference test (eg, histopathology) is to utilize a third test as a "resolver."[20] Noninvasive *H pylori* test performance characteristics were also calculated using slide review consensus interpretation as an imperfect resolver (Supplemental Figure 2). This composite reference standard method uniformly increased the apparent sensitivity and minimally decreased apparent specificity across tests. However, the trend of superior sensitivity and modestly inferior specificity of *H pylori* serum IgG was unchanged.

Based on the noninvasive test sensitivities and specificities measured in Figure 2B, the performance of 3 candidate diagnostic algorithms was modeled over a range of disease prevalence Figure 4 and Figure 5. The primary outcome measures selected were diagnostic error rate and testing cost, crudely estimated as the laboratory's estimated cost, the 2018 Medicare reimbursement rate, or the laboratory's loss per diagnosis if all payers reimbursed the Medicare rate (laboratory cost minus Medicare reimbursement). Serology-first algorithms had lower predicted diagnostic error than a stool antigen–first algorithm (Figure 4). A testing algorithm with serology first and reflexive stool antigen for positive serology was more accurate than serology alone and essentially eliminated false positives related to the nonspecificity of *H pylori* IgG (Figure 4 and Supplemental Figure 5). Cost associated with tiered testing was approximately linearly related to disease prevalence; added cost per diagnosis is likely acceptable for populations with low disease prevalence. The stool antigen–only algorithm offers an advantage of simplicity, as it is inclusive of patients with prior infection. However, modeled diagnostic accuracy is inferior to algorithms including serology. UBT performed similarly to stool antigen in the current study and in previous comparisons[24] and can be substituted for stool antigen. UBT has higher associated laboratory cost than stool antigen, but Medicare reimbursement (approximately $93) is better matched to laboratory cost (Figure 4). Importantly, any patients with previous infection or prior positive tests should be directed away from serology, regardless of the testing approach. The proposed laboratory-testing algorithms do not account for the clinical decision to pursue endoscopy and tissue-based testing, a process that should be based on clinical context, "alarm symptoms," etc.[10,25] The overall, most cost-effective approach to noninvasive *H pylori* testing is likely to be that with the lowest diagnostic error because it allows optimal targeting of therapy and informs appropriate selection of patients for endoscopy.

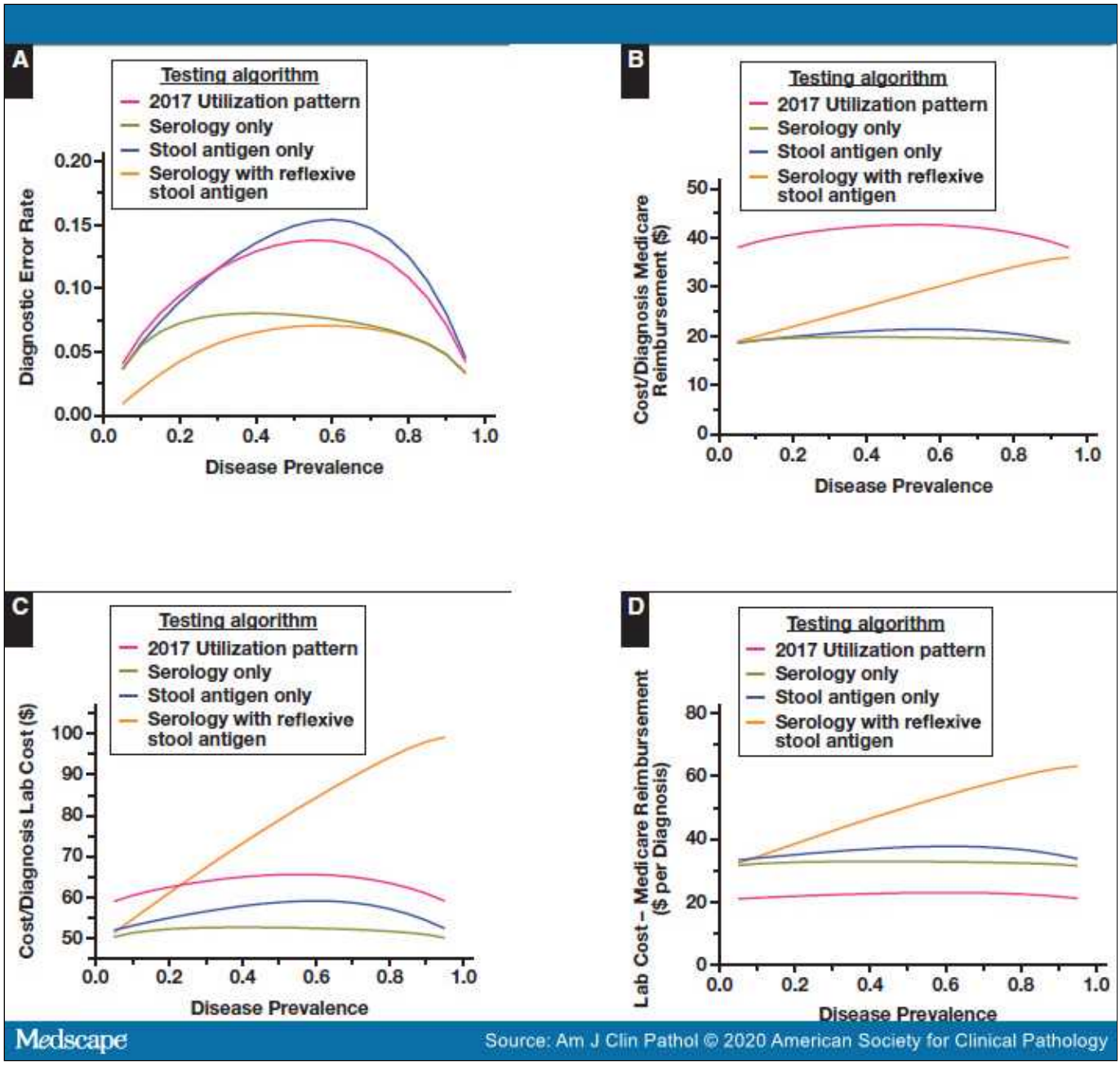


Figure 4.

Modeled performance of candidate *Helicobacter pylori* testing algorithms. **A**, Diagnostic error rate depends on disease prevalence and is lowest for algorithms utilizing serology. Cost per diagnosis is modeled by 2018 Medicare reimbursement (**B**) and laboratory cost (**C**) and difference between laboratory cost and Medicare reimbursement rates (**D**). Reflexive confirmation of positive serology with stool antigen improves accuracy with added cost.

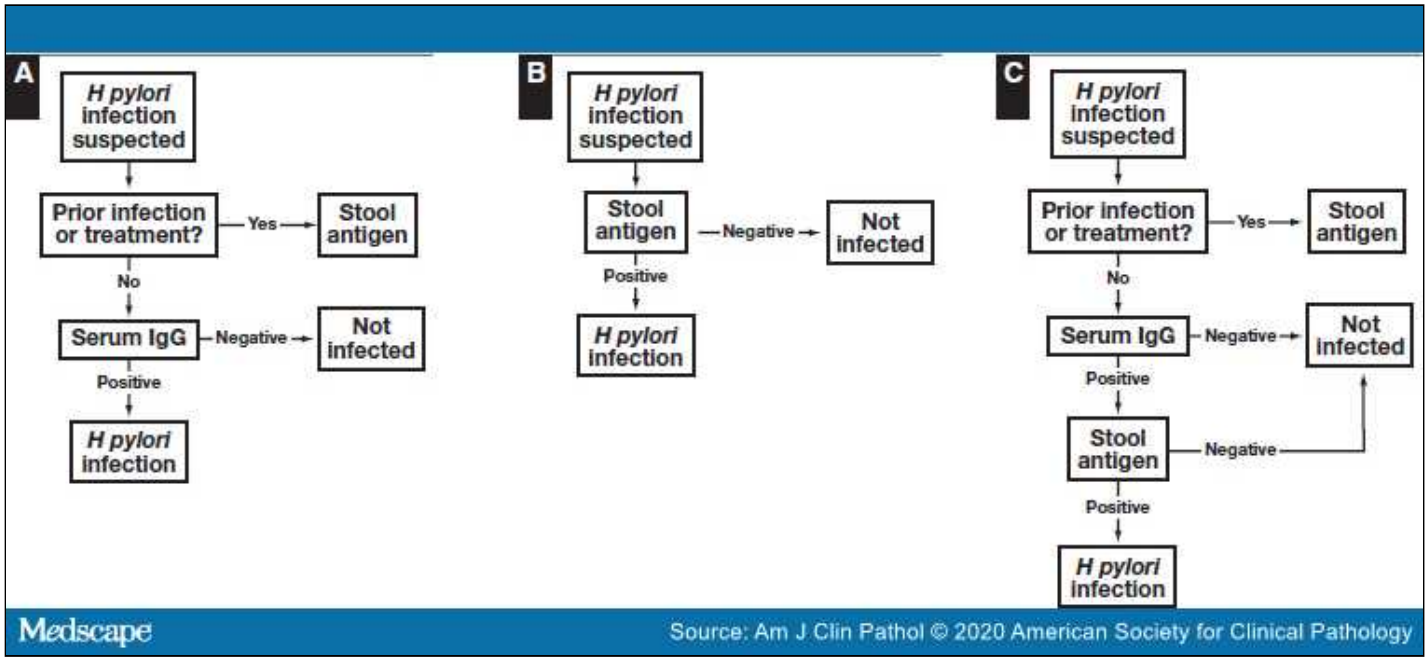


Figure 5.

Proposed testing algorithms for noninvasive diagnosis of *Helicobacter pylori* infection. Candidate testing algorithms include serology only (**A**), stool antigen only (**B**), or a reflexive combination of sensitive serology followed by specific stool antigen (**C**). Algorithms utilizing serology require exclusion of prior infection and/or treatment. Urea breath test may be substituted for stool antigen.

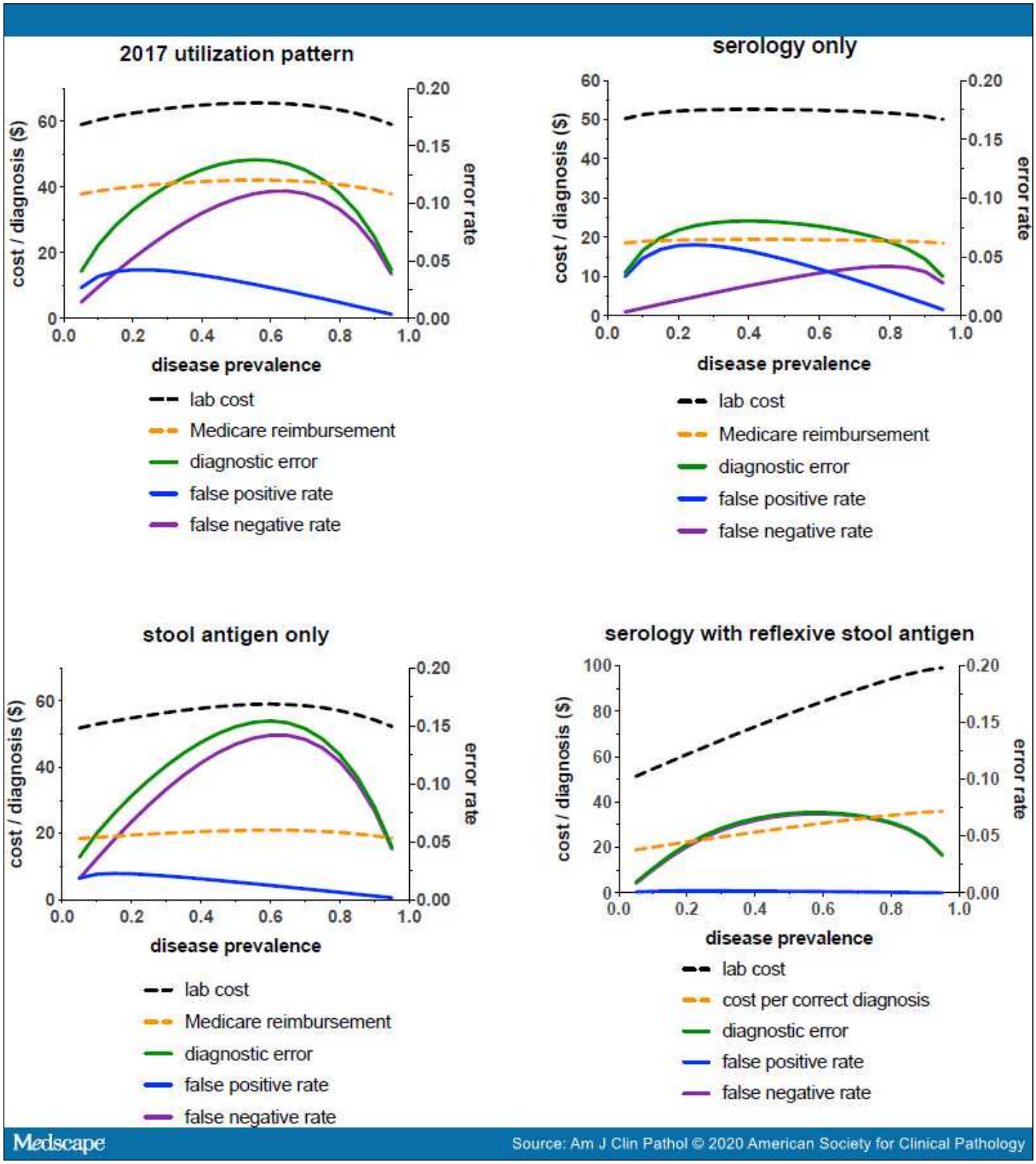


Figure S5.

Modeled performance of H. pylori testing algorithms. The diagnostic error rates of stool antigen and serology are dominated by false negatives and false positive, respectively. Serology with reflexive stool antigen testing essentially eliminates false positives but introduces a significant false negative rate in the population with combination of positive serology and negative stool antigen.

Conclusions

Perhaps the most striking finding in this review of *H pylori* testing is the low observed sensitivity of stool antigen and UBT (best estimates, 0.6–0.7), treating histopathology as the reference standard. The contrast to prior sensitivity measurements of stool antigen (0.77–0.90) and UBT (0.72–0.99)[11–13] may reflect differences in patient populations, increasing utilization of interfering therapies (eg, PPIs) over time, difficulty of excluding patients on interfering therapy in routine clinical practice, and the reference method used. As overall utilization of PPIs increases[26] and an expanding fraction of acid suppression therapy is obtained over

the counter,[27] more noninvasive *H pylori* testing will inevitably be done on patients taking interfering therapy. The anticipated effect will be reduced sensitivity of UBTs and stool antigen tests (Figure 2).

Serum *H pylori* IgG had sensitivity in our patient population similar to previously published performance studies.[15] Previous studies have highlighted performance differences among serology kits and patient populations, underscoring the importance of local validation of performance with implementation of *H pylori* serology. The relatively high sensitivity and corresponding NPV indicate the clinical utility of serology as a "rule-out" test, particularly in patients with low to moderate pretest probability. In our population, when excluding patients with a known past infection, the specificity of serum IgG was lower than other noninvasive tests and *H pylori* culture but still acceptable (approximately 0.75–0.90) and had a better ROC (0.88) than described previously.[15] Serology also offers advantages of no or minimal interference from treatment and the convenience of blood-based testing as opposed to stool specimen collection and timed breath specimen collection.

The unexpectedly low sensitivity of stool antigen and UBT in comparison to serology prompted an investigation of candidate testing algorithms. The superior sensitivity of *H pylori* serum IgG is attractive for an initial test or for screening in an appropriate clinical context, such as before bariatric surgery. The specificity limitation of serology due to persistent seropositivity can be minimized by exclusion of previously infected patients on clinical grounds, particularly in a low-prevalence population (estimated as 14%-21%; Figure 1). Clinical exclusion may be facilitated by clinical decision support alerting the clinician to previous positive test results. Placing serology at the start of diagnostic testing algorithms for patients without prior infection is predicted to improve accuracy of noninvasive diagnosis (Figures 3 and 4). Although the decision to eradicate with antibiotics based on positive serology alone is generally discouraged because of specificity limitations, a PPV of approximately 0.75 (Figure 2) may be sufficient in an appropriate clinical context. The quantitation (titer) of serum *H pylori* IgG also may also influence the PPV (Figure 3). The nonspecificity of serology can be mitigated by reflexive stool antigen or UBT, both of which are highly specific. Tiered testing drives the diagnostic error rate lower, particularly among low-prevalence populations, when most predicted errors are missed infection (false negatives) in patients with positive serology and negative stool antigen or UBT (Supplemental Figure 5). The high sensitivity and NPV of serology imparts utility as a rule-out test, including outside the context of a testing algorithm.

Another important factor in the interpretation of noninvasive test-performance studies is the known imperfect sensitivity of histopathology.[8] *H pylori* infection may be "patchy," leading to sampling error, and variable organism burden and treatment effects impact performance of invasive tests.[25] In the current study, cases with discordance between the initial pathologic diagnosis and noninvasive *H pylori* tests were enriched for challenging histopathology, reflected by lower interobserver concordance among gastrointestinal pathologists (Image 1). Agreement on biopsies initially diagnosed with equivocal language was very low. The histology of many equivocal cases included coccoid and intracellular forms (Image 1), which are associated with PPI therapy.[8] Discordance with noninvasive testing or equivocation by the pathologist should alert the clinician to interpret biopsy results in the clinical context, particularly for patients with recent antibiotic or acid-suppressing therapy.

Histopathology, stool antigen, and UBT all are affected by therapeutic interreferences to different extents, and this may be reflected in performance studies. Previous work has largely, although not exclusively, compared noninvasive tests with each other.[11–13,15] We believe histopathology, when corrected for consensus review by multiple pathologists, offers the gold standard for diagnosing acute *H pylori* infection because it can incorporate organism morphology in the context of host inflammatory response.

Tiered or reflexive testing has historically been an effective way to reduce diagnostic error for infectious diseases, exemplified by treponemal and nontreponemal testing for the diagnosis of syphilis. Our institution uses a "reverse" algorithm (treponemal test first) using in-house assays, incorporates results from independent tests performed at the state department of health, and delivers a single final interpretive comment. Other common examples of reflexive panels for the diagnosis of infectious disease include testing for hepatitis C virus, HIV, and Lyme disease. Optimized diagnostic stewardship for the proposed *H pylori*–infection reflexive diagnostic algorithm might include order–interface rules based on previous laboratory results and/or diagnosis codes. Tools for implementation and stewardship exist in several commercially available EHR systems. When informatics/EHR resources are limited, reflexive algorithms can be provider driven, exemplified by follow-up invasive testing for Pap smears. Whether EHR or provider driven, a reflexive algorithm has advantages in diagnostic performance and patient convenience, requiring a blood draw for most patients and limiting requirements for stool or breath samples. Access to care may also potentially be improved by eliminating the need for a return visit to drop off a stool sample or a lengthy breath test for most patients.

This study highlights the importance of ongoing assessment of *H pylori* laboratory test performance at the local and regional level. Selection and interpretation of *H pylori* testing should be informed by disease prevalence, potential for therapeutic interference, and cost, in addition to published analytical performance studies. Despite multiple guidelines recommending against its use, serology contributes substantially to accurate diagnosis of *H pylori* infection in our patient population.

Sidebar

Key Points

- *Helicobacter pylori* serology has been recommended against in current guidelines, but few recent studies have assessed its performance compared with other tests in clinical practice.
- In our study, *H pylori* serology performed with higher sensitivity and negative predictive value than urea breath test, stool antigen, or culture when compared with histopathology.
- In low-prevalence populations, reflexive stool antigen or urea breath test after a positive initial *H pylori* serology may optimize diagnostic accuracy.

References

1. Hooi JKY, Lai WY, Ng WK, et al. Global prevalence of *Helicobacter pylori* infection: systematic review and meta-analysis. *Gastroenterology.* 2017;153:420–429.

2. Venerito M, Vasapolli R, Rokkas T, et al. *Helicobacter pylori*, gastric cancer and other gastrointestinal malignancies. *Helicobacter*. 2017;22(suppl 1):1–5.

3. Li Z, Ying X, Shan F, et al. The association of garlic with *Helicobacter pylori* infection and gastric cancer risk: a systematic review and meta-analysis. *Helicobacter.* 2018;23:e12532.

4. Lu B, Li M. *Helicobacter pylori* eradication for preventing gastric cancer. *World J Gastroenterol.* 2014;20:5660–5665.

5. Kong YJ, Yi HG, Dai JC, et al. Histological changes of gastric mucosa after *Helicobacter pylori* eradication: a systematic review and meta-analysis. *World J Gastroenterol.* 2014;20:5903–5911.

6. Doorakkers E, Lagergren J, Engstrand L, et al. Eradication of *Helicobacter pylori* and gastric cancer: a systematic review and meta-analysis of cohort studies. *J Natl Cancer Inst*. 2016;108:1–10.

7. Schulman AR, Abougergi MS, Thompson CC. *H. pylori* as a predictor of marginal ulceration: a nationwide analysis. *Obesity (Silver Spring).* 2017;25:522–526.

8. Batts KP, Ketover S, Kakar S, et al; Rodger C Haggitt Gastrointestinal Pathology Society. Appropriate use of special stains for identifying *Helicobacter pylori*: recommendations from the Rodger C. Haggitt Gastrointestinal Pathology Society. *Am J Surg Pathol.* 2013;37:e12–e22.

9. Blanchard TG, Nedrud JG. Laboratory maintenance of *Helicobacter* species. *Curr Protoc Microbiol.* 2006;Chapter 8:Unit8B.1.

10. Bessède E, Arantes V, Mégraud F, et al. Diagnosis of *Helicobacter pylori* infection. *Helicobacter*. 2017;22(suppl 1):1–6.

11. Kesli R, Gokturk HS, Erbayrak M, et al. Comparison of the diagnostic values of the 3 different stool antigen tests for the noninvasive diagnosis of *Helicobacter pylori* infection. *J Investig Med.* 2010;58:982–986.

12. Ferwana M, Abdulmajeed I, Alhajiahmed A, et al. Accuracy of urea breath test in *Helicobacter pylori* infection: meta-analysis. *World J Gastroenterol.* 2015;21:1305–1314.

13. Best LM, Takwoingi Y, Siddique S, et al. Non-invasive diagnostic tests for *Helicobacter pylori* infection. *Cochrane Database Syst Rev*. 2018;3:CD012080.

14. Calvet X. Diagnosis of *Helicobacter pylori* infection in the proton pump inhibitor era. *Gastroenterol Clin North Am.* 2015;44:507–518.

15. She RC, Wilson AR, Litwin CM. Evaluation of *Helicobacter pylori* immunoglobulin G (IgG), IgA, and IgM serologic testing compared to stool antigen testing. *Clin Vaccine Immunol.* 2009;16:1253–1255.

16. Burucoa C, Delchier JC, Courillon-Mallet A, et al. Comparative evaluation of 29 commercial *Helicobacter pylori* serological kits. *Helicobacter.* 2013;18:169–179.

17. Biranjia-Hurdoyal SD, Seetulsingh-Goorah SP. Performances of four *Helicobacter pylori* serological detection kits using stool antigen test as gold standard. *PLoS One.* 2016;11:e0163834.

18. Admon AJ, Gupta A, Williams M, et al. Appraising the evidence supporting choosing Wisely® recommendations. *J Hosp Med.* 2018;13:688–691.

19. Chey WD, Leontiadis GI, Howden CW, et al. ACG clinical guideline: treatment of *Helicobacter pylori* infection. *Am J Gastroenterol.* 2017;112:212–239.

20. Hess AS, Shardell M, Johnson JK, et al. Methods and recommendations for evaluating and reporting a new diagnostic test. *Eur J Clin Microbiol Infect Dis.* 2012;31:2111–2116.

21. Malfertheiner P, Megraud F, O'Morain CA, et al; European *Helicobacter* and Microbiota Study Group and Consensus panel. Management of *Helicobacter pylori* infection-the Maastricht V/Florence consensus report. *Gut.* 2017;66:6–30.

22. Attumi TA, Graham DY. Follow-up testing after treatment of *Helicobacter pylori* infections: cautions, caveats, and recommendations. *Clin Gastroenterol Hepatol.* 2011;9:373–375.

23. Hawkins DM, Garrett JA, Stephenson B. Some issues in resolution of diagnostic tests using an imperfect gold standard. *Stat Med.* 2001;20:1987–2001.

24. Choi J, Kim CH, Kim D, et al. Prospective evaluation of a new stool antigen test for the detection of *Helicobacter pylori*, in comparison with histology, rapid urease test, (13)C-urea breath test, and serology. *J Gastroenterol Hepatol.* 2011;26:1053–1059.

25. Dixon MF, Genta RM, Yardley JH, et al. Classification and grading of gastritis: the updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. *Am J Surg Pathol.* 1996;20:1161–1181.

26. Lazarus B, Chen Y, Wilson FP, et al. Proton pump inhibitor use and the risk of chronic kidney disease. *JAMA Intern Med.* 2016;176:238–246.

27. Sheikh I, Waghray A, Waghray N, et al. Consumer use of over-the-counter proton pump inhibitors in patients with gastroesophageal reflux disease. *Am J Gastroenterol.* 2014;109:789–794.

Acknowledgments
We thank personnel of the University of Washington Department of Laboratory Medicine Immunology laboratory, especially Anu Chaudhary, for helpful discussions about *Helicobacter pylori* serology, and the University of Washington Departments of Laboratory Medicine and Pathology informatics team for assistance in obtaining data.

Am J Clin Pathol. 2020;154(2):255-265.