Open access • Journal Article • DOI:10.1111/APT.14248

Systematic review with meta-analysis: diagnostic performance of the combination of pepsinogen, gastrin-17 and anti-Helicobacter pylori antibodies serum assays for the diagnosis of atrophic gastritis. — Source link

Rocco Maurizio Zagari, Stefano Rabitti, D. C. Greenwood, Leonardo Henry Eusebi ...+2 more authors

Institutions: University of Bologna, University of Leeds

Published on: 07 Aug 2017 - Alimentary Pharmacology & Therapeutics (Aliment Pharmacol Ther)

Topics: Atrophic gastritis, Helicobacter pylori and Gastritis

---

Article:
Zagari, RM, Rabitti, S, Greenwood, DC orcid.org/0000-0001-7035-3096 et al. (3 more authors) (2017) Systematic review with meta-analysis: diagnostic performance of the combination of pepsinogen, gastrin-17 and anti-Helicobacter pylori antibodies serum assays for the diagnosis of atrophic gastritis. Alimentary Pharmacology and Therapeutics, 46 (7). pp. 657-667. ISSN 0269-2813

https://doi.org/10.1111/apt.14248

Systematic review with meta-analysis: diagnostic performance of the combination of pepsinogen, gastrin-17 and anti-Helicobacter pylori antibodies serum assays for the diagnosis of atrophic gastritis.

Short title: The panel test for the diagnosis of atrophic gastritis.

Rocco Maurizio Zagari[1], Stefano Rabitti[1], Darren C. Greenwood[2], Leonardo H. Eusebi[1], Amanda Vestito[3], Franco Bazzoli[1].

[1] Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy

[2] Division of Biostatistics, University of Leeds, Leeds, UK

[3] Gastroenterology Unit, S. Orsola-Malpighi Hospital, Bologna, Italy

Corresponding author:

Prof. Rocco Maurizio Zagari,

Department of Medical and Surgical Sciences, University of Bologna,

Policlinico S. Orsola-Malpighi, Via Massarenti n. 9, 40138 Bologna, Italy.

Tel and Fax: +39 051 2144117

Email: roccomaurizio.zagari@unibo.it

Word count:

ABSTRACT

Background: The combination of pepsinogen, gastrin-17 and anti-H. pylori antibodies serological assays (panel test) is a non invasive tool for the diagnosis of atrophic gastritis. However, the diagnostic reliability of this test is still uncertain.

Aim: To assess the diagnostic performance of the serum panel test for the diagnosis of atrophic gastritis.

Methods: Medline via PubMed, Embase, Scopus, Cochrane Library databases and abstracts of international conferences proceedings were searched from January 1995 to December 2016 using the primary keywords "pepsinogens", "gastrin", "atrophic gastritis", "gastric precancerous lesions". Studies were included if they assessed the accuracy of the serum panel test for the diagnosis of atrophic gastritis using histology according to the updated Sydney System as reference standard.

Results: Twenty studies with a total of 4241 subjects assessed the performance of serum panel test for the diagnosis of atrophic gastritis regardless of the site in the stomach. The summary sensitivity was 74.7% (95% confidence interval (CI), 62.0-84.3) and the specificity was 95.6% (95%CI, 92.6-97.4). With a prevalence of atrophic gastritis of 27% (median prevalence across the studies), the negative predictive value was 91%. Few studies with small sample size assessed the performance of the test in detecting the site of atrophic gastritis.

Conclusions: The combination of pepsinogen, gastrin-17 and anti-H. pylori antibodies serological assays appears to be a reliable tool for the diagnosis of atrophic gastritis. This test may be used for screening subjects or populations at high risk of gastric cancer for atrophic gastritis; however, a cost-effectiveness analysis is needed.

INTRODUCTION

Atrophic gastritis is a loss of appropriate glands of the gastric mucosa, which are replaced by connective tissue and/or intestinal type-epithelium (intestinal metaplasia) (1). Atrophic gastritis, which is usually caused by Helicobacter (H.) pylori or may have an autoimmune origin, predisposes to gastric cancer and impairs gastric physiology leading to hypo- or achlorhydria, iron and vitamin B_{12} malabsorption (2). It is well known that the intestinal-type gastric adenocarcinoma develops in a stepwise manner with a sequence of events that evolves from atrophic gastritis and intestinal metaplasia to dysplasia and carcinoma. International guidelines recommend endoscopic follow-up and gastric biopsies for subjects with atrophic gastritis, even after H. pylori eradication, in order to early detect gastric cancer and reduce mortality (2,3).

However, identifying subjects with an underlying atrophic gastritis is still an issue. Gastroscopy and histology are the reference standard, but the use of endoscopy as a screening test is costly, uncomfortable and does not have good patient's compliance (2). International guidelines and a recent global consensus report have agreed that serological tests may be very useful in order to identify individuals with atrophic gastritis (2-4). A non-invasive tool able to easily identify individuals with atrophic gastritis, or those who are very likely to carry such precancerous lesion, is essential for improving the early diagnosis of gastric cancer. Such test would be ideal for screening subgroups of subjects, such as those with a positive family history, or populations at high risk of gastric cancer, to identify those patients which must undergo endoscopy. In addition, an accurate non-invasive test would be very helpful to improve our knowledge on the epidemiology of atrophic gastritis in the general population.

Over the last decade, the combination of serological assays including pepsinogen, gastrin-17 (G-17) and anti-H. pylori antibodies (panel test) has been proposed as a non-invasive test for the diagnosis of atrophic gastritis (2-4). The rationale of this test is based on the fact that pepsinogen-I (PG-I) is secreted only by oxintic glands of the corpus mucosa, while pepsinogen II (PG-II) is also produced in the gastric antrum and duodenum, and that gastrin-17 is only secreted by the G cells of the antral mucosa. Serum PG-I levels and/or the PG-I/PG-II ratio seem to be lower in patients with corpus atrophic gastritis, whereas a low G-17 serum level, in combination with positive anti-H. pylori antibodies (HpAb), would indicate the presence of antrum atrophic

gastritis (5). Thus the combination of the results of HpAb, PG-I or PGI/PGII ratio, and G-17 tests would allow us to detect the presence and site of atrophic gastritis (5).

However, although the panel test is commercially available and used in many countries worldwide, in particular in Europe, the diagnostic reliability of this test remains uncertain. Clarifying the diagnostic performance of this test is essential for its use in individuals and in the general population for gastric cancer screening and epidemiological studies on the prevalence and incidence of atrophic gastritis.

The aim of this study was to carry out a systematic review and meta-analysis to determine the diagnostic performance of the combination of pepsinogen, gastrin-17 and anti-H. pylori antibodies serum assays for the diagnosis of atrophic gastritis in adults. The primary outcome was to assess the diagnostic performance for the diagnosis of atrophic gastritis regardless of the location. The secondary outcome was to determine the accuracy in detecting the site of atrophic gastritis.

METHODS

We performed a systematic review and a meta-analysis following the recommendations of the Cochrane Collaboration's Diagnostic Test Accuracy Group (6).

Search strategy and study selection

We searched MEDLINE via PubMed, Ovid Embase, the Cochrane Library and Scopus databases up to 31st December 2016. The electronic search of literature was performed by using the following keywords: "pepsinogens", "pepsinogen I", "pepsinogen II", "gastrin", "panel test" or "gastropanel", and "atrophic gastritis", "gastric atrophy", "intestinal metaplasia", "gastric precancerous condition" or "gastric precancerous lesion". The search strategies are reported in the Supplementary Appendix 1. The first validation study of the panel test was published in 2002 (7); in order to identify earlier studies the search period was extended back to January 1995. In addition, we searched electronically and by hands abstracts of the conferences proceedings of Digestive Diseases Week, United European Gastroenterology Week, Asia Pacific Digestive Week and International Workshop on Helicobacter and Microbiota for the same period. We searched the references lists

of the included studies and relevant published reviews. We also searched the reference list of the manufacturer's website of the panel test GastroPanel® (Biohit Plc, Finland). We did not restrict for language or publication status. We obtained translation of any non-English articles.

Two authors (RMZ and LHE) did the initial selection on the basis of titles and abstracts. Subsequently, they independently performed a detailed full text assessment of potentially relevant studies, with any disagreement resolved through discussion or arbitration by a third reviewer (FB).

For the inclusion in the review, we selected studies if they met the following pre-specified criteria: diagnostic studies evaluating the accuracy of the combination of pepsinogen, gastrin-17 and anti-H. pylori antibodies serological assays for the diagnosis of atrophic gastritis in adults using the histological diagnosis of atrophic gastritis according to the updated Sydney System as reference standard (1). We excluded studies that did not meet the inclusion criteria or whether essential information was missing and could not be obtained by the authors.

Data Extraction and Quality Assessment

Two authors (RMZ and SR) extracted independently relevant data on the publication, study methods and results using a standardized data extraction form. We constructed 2 x 2 tables that contained the number of cases found to be true positives (subjects with positive panel test who had atrophic gastritis at histology), true negatives (subjects with negative panel test who did not have atrophic gastritis at histology), false positives (subjects with positive panel test who did not have atrophic gastritis at histology) and false negatives (subjects with negative panel test who did not have atrophic gastritis at histology). For the main outcome of this study we considered the panel test result as positive when it was positive for presence of atrophic gastritis regardless of the site in the stomach. Subjects with previous gastric surgical resection were excluded.

When possible, additional 2x2 tables were constructed for the site of atrophic gastritis. In addition, the following items were extracted from each study, when available: study design, country of the study, inclusion and exclusion criteria for participants, total number of participants, average age and number of males, indications for endoscopy, use of proton pump inhibitors (PPIs) over the last week, description of the index

test and the threshold values used for each test of the panel, description of the reference standard, number and site of gastric biopsy specimens, grade of severity of atrophic gastritis (atrophy at any grade of severity or moderate-severe atrophy) used for defining the target condition. When multiple articles for a single study were found, the latest publication was considered and supplemented, if necessary, with data from the previous publications.

Two authors (RMZ and SR) independently assessed the methodological quality of the included studies by using the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) tool (Supplementary Appendix 2) (8). We evaluated, in particular, the presence of potential bias in patients selection, blinding to the histological diagnosis, description of the reference standard and inclusion of all patients in the analysis. Any disagreements were resolved through discussion and, if necessary, arbitration by a third reviewer (FB).

Statistical Analysis

Using 2 x 2 tables we calculated sensitivity and specificity with 95% confidence intervals (95% CI) for each study, and created coupled forest plots for showing each set of data. We calculated summary estimates of sensitivity and specificity, positive and negative likelihood ratio using a random effect bivariate model and we fit a summary hierarchical receiving operating characteristic (HSROC) curve (9,10). We used summary estimates of sensitivity and specificity to estimate the summary negative and positive predictive values based on the median prevalence (pre-test probability) of atrophic gastritis across the studies, which was calculated from the median prevalence of the included studies.

We explored heterogeneity between studies through visual examination of the forest plot and HSROC curve (6). We planned to explore the following sources of heterogeneity adding them as covariates, if appropriate, to a bivariate regression model: index test, target condition, setting, study design, country, use of PPIs, type of publication, and methodological quality. We performed sub-group analyses for any covariates that showed a statistically significant association with the summary estimates. We performed separate meta-analyses to assess the performance of the panel test for the diagnosis of the site of atrophic gastritis: antrum-limited atrophic gastritis, corpus-limited atrophic gastritis, and both antrum and corpus atrophic gastritis. We used

Cook's distance to check for particular influential studies and produced a scatter plot of the standardized level 2 residuals to check for outliers (11). We did not investigate publication bias as standard funnel plot and tests for publication bias are not recommended in meta-analysis of diagnostic test accuracy studies (6). All analyses were performed with STATA version 14 (StataCorp, College Station, Texas, USA).

RESULTS

The electronic search identified 3924 records after duplicates were removed, of which 38 full text articles were assessed for eligibility (5,7,12-47). One additional full text article identified from the reference list of the manufacture's website of GastroPanel (Biohit Plc) was also assessed for eligibility (48). Of the 39 articles, 15 met the criteria for the inclusion in the review (7,12,15,18,19,20,23,26,30,33,35,38,42,47,48). In addition, a total of 5 abstracts that met inclusion criteria were identified from the conferences proceedings (49-53). Finally, a total of 20 studies consisting of 15 papers and 5 abstracts were included in the meta-analysis. Appendix Figure 1 shows the flowchart of references through the selection process and the reasons for study exclusion.

Study characteristics

The 20 included studies involved a total of 4241 participants, with 1143 having the target condition. The mean age of participants ranged from 39 (23) to 65 years (15) and the proportion of men from 20% (52) to 58.6% (30). The mean prevalence of atrophic gastritis in the included studies ranged from 8.1% (38) to 97.2% (18). Fifteen studies were conducted in Europe (7,12,15,19,20,23,26,33,35,42,49-53), three in Russia (18,47,48), one in Japan (30), and one in Africa (38). Fourteen studies were performed in a single centre (7,15,18,20,23,30,33,38,48-53), and six were multicentre studies (12,19,26,35,42,47). All studies had a cohort design, except three that were case-control studies (7,35,52). Seventeen studies assessed the diagnostic performance of the panel test in patients with upper gastrointestinal symptoms referred to endoscopy (7,12,15,18,19,20,23,30,33,35,38,42,48,50-53), while three studies were performed in the community

(26,47,49). All studies, but one where HpAb were detected using Helori-test IgG (Eurospital, Italy) (20), used the serum assays of the GastroPanel (Biohit, Finland).

Eleven studies used as target condition atrophic gastritis regardless of the grade of severity (from mild to severe atrophy) (18,20,23,33,38,42,47-50,52), seven studies used moderate-severe atrophic gastritis (7,12,19,26,30,35,51) and two studies reported the results for both target conditions separately (15,53).

Regarding the index test, fourteen studies measured the fasting serum level (basal) of G-17 (G-17b) (19,20,23,26,30,33,35,38,42,47,50-53), while six studies measured the serum level of stimulated G-17 (G-17s) taking a blood sample after a protein reach drink (7,12,15,18,48,49). Thirteen studies used PGI (7,12,15,18,19,20,33,35,42,48-50,53) while seven studies used both PGI and PGI/PGII ratio (23,26,30,38,47,51,52). The most commonly used cut-offs were HpAb < 30 enzime immune units (EIU), G17= 1-10 pmol/l, PGI=25-50 microgr/l, PGI/PGII ratio>3. In all studies HpAb, pepsinogens and G-17 serum levels were determined by ELISA. In sixteen studies participants stopped the use of PPIs at least a week before the enrolment (7,12,15,18,19,20,23,30,33,38,47-51,53), whereas in four studies a sub-group of participants used PPIs (26,35,42,52). Table 1 shows the characteristics of the included studies.

Only seven studies (1529 participants, including 471 with atrophic gastritis whose 261 with antrum-limited, 136 with corpus-limited and 74 with both antrum and corpus atrophic gastritis) reported data on the diagnostic performance of the panel test for the site of atrophic gastritis (7,12,30,35,47-49).

Supplementary Table 1 and Supplementary Table 2 show the results of the assessment of methodological quality of the included studies. All studies, but one (33), were at “high risk” or “unclear risk” in one or more domains concerning bias or applicability to the review question (Supplementary Table 2). Most studies were at high risk of bias in the selection of participants, mainly because they did not enroll a consecutive or random sample of subjects. In addition, about half of studies did not include all participants in the final analysis or did not report the time interval between gastroscopy and blood sampling, thus having a high or unclear risk of bias in the flow and timing domain.

Table 1. Characteristics of the studies included.

Study year	Study type and country	Biopsy specimens site and no	Target condition	Index test cut-off of panel tests	Total no included	Prevalence of target condition n. (%)	Age mean (range or SD)	Sex, male n. (%)	Use of PPIs
Sipponen 2002 [7]	Single centre, Finland	Antrum , 2 Corpus, 2	AG	HpAb = <30 EIU PGI= 25-50 µ/l G-17s=1-10 pmol/l	86	42 (48.8)	62 (14)	43 (43.0)	No
Zagari 2002[49]	Single centre, Italy	Antrum, 3 Corpus, 3	AG	HpAb = <30 EIU PGI=25-50 µ/l G-17s=1-10 pmol/l	104	35 (33.6)	55 (N/A)	56 (53.8)	No
Vaananen 2003 [12]	Multicentre Finland	Antrum, 2 Corpus, 2	M/S AG	HpAb = <30 EIU PGI=25-50 µ/l G-17s=1-10 pmol/l	398	60 (15.1)	58 (15)	164 (40.6)	No
Hartleb 2004 [15]	Single centre, Poland	Antrum, 2 Corpus, 2	AG	HpAb = <38 EIU PGI = 25-50 µ/l G-17s=1-10pmol/l	55	19 (34.5)	65 (55-81)	25 (45.4)	No
De Korwin 2004 [50]	Single centre, France	Antrum, 4 Corpus, 3	AG	HpAb = <38 EIU PGI=25-50 µ/l G-17b=1-10pmol/l	50	20 (40)	60 (19.8)	22 (43.3)	No
Pyurveyeva 2005 [48]	Single centre, Russian	Antrum, 2 Angulus,1 Corpus, 2	AG	HpAb = <38 EIU PGI= < 25 µ/l G-17s = < 5 pmol/l	100	89 (89)	N/A (21-77)	34 (34)	No
Pasechnikov 2005 [18]	Single centre, Russian	Antrum, 2 Corpus, 2	AG	HpAb = <32 EIU PGI=25-50 µ/l G-17s=5-10 pmol/l	178	173 (97.2)	N/A (38-80)	N/A	No
Germanà 2005 [19]	Multicentre Italy	Antrum, 2 Angulus,1 Corpus, 2	M/S AG	HpAb = <38 EIU PGI=25-100 µ/l G-17b=2.5-7.5pmol/l	287	60 (20.9)	50 (16)	121 (42.2)	No
Nardone 2005[20]	Single centre, Italy	Antrum, 2 Angulus, 1 Corpus, 2	AG	HpAb = N/A PGI=25-100 µ/l G-17b=2.5-7.5pmol/l	94	30 (31.9)	56 (N/A)	36 (38.3)	No
Cavallaro 2005 [51]	Single centre, Italy	Antrum, 2 Angulus, 1 Corpus, 2	M/S AG	HpAb=N/A PGI=N/A PGI/PGII=N/A G-17b=N/A	176	21 (11.9)	49 (17)	69 (39.2)	No
Valle Munoz 2007 [23]	Single centre, Spain	Antrum, 2 Corpus, 2	AG	HpAb = N/A PGI=N/A PGI/PGII=N/A G-17b=NA	56	80 (14.3)	39 (15)	24 (42.9)	No
Storskrubb 2008 [26]	Multicentre Sweden	Antrum, 2 Corpus, 2 Fundus, 2	M/S AG	HpAb= <38 EIU PGI=25-50 µ/l PGI/PGII= < 3 G-17b=5-10 pmol/l	976	86 (8.8)	54 (N/A)	473 (48.5)	Yes, 4.9%
Ijima 2009 [30]	Single centre, Japan	Antrum, 1 Corpus, 1	M/S AG	HpAb = <30 EIU PGI=30-165 µ/l PGI/PGII=3-20 G-17b=1-10 pmol/l	162	20 (12.3)	55 (22-79)	95 (58.6)	No
Lombardo 2010 [33]	Single centre, Italy	Antrum, 1 Angulus,1 Corpus, 2	AG	HpAb = <30 EIU PGI = 30-120 µ/l G-17b=2-10 pmol/l	400	64 (16.0)	46 (19)	186 (46.5)	No
Peitz 2011 [35]	Multicentre, Europe*	Antrum, 2 Corpus, 2	M/S AG	HpAb= <30 EIU PGI=25-50 µ/l G-17b=2.5-5 pmol/l	416	136 (32.7)	60 (11)	224 (53.8)	Yes, 45.5%
Di Mario 2011 [52]	Single centre, Italy	Antrum ,3 Corpus, 2	AG	HpAb=NA PGI=NA PGI/PGII= < 3 G-17b=N/A	20	19 (95)	65 (N/A)	4 (20)	Yes, 35%
Noah Noah 2012 [38]	Single centre, Africa	Antrum, 2 Corpus, 2	AG	HpAb = <30 EIU PGI=25-50 PGI/PGII= < 2.5 G-17b=1-5 pmol/l	86	7 (8.1)	46 (3)	22 (25.6)	No
Mc Nicholl 2014 [42]	Multicentre, Spain	Antrum, 2 Corpus, 2	AG	HpAb= <30 EIU PGI= 25-50 µ/l G-17b=1-10 pmol/l	85	10 (11.8)	44 (14)	20 (23.0)	Yes, 5%
Goni 2015 [53]	Single centre, Italy	Antrum, 2 Angulus,1 Corpus, 2	AG	HpAb=N/A PGI=N/A G-17b=N/A	249	151 (60.6)	49 (21-79)	129 (51.8)	No
Roman 2016 [47]	Multicentre, Russian	Antrum, 2 Corpus, 2	AG	HpAb = <30 EIU PGI = 30-160 µ/l PGI/PGII= < 3 G-17b=1-7 pmol/l	263	89 (33.8)	N/A	N/A	No

N/A, not available;*Germany, Austria and Swiss; AG,atrophic gastritis; M/S,moderate or severe; HpAb, antibodies to H. pylori; EIU, enzyme immune units; PGI,pepsinogen I;PGII, pepsinogen II;G-17b,basal Gastrin-17;G-17s, stimulated Gastrin-17; PPIs, proton pump inhibitors; thirteen patients with antrum resected and one with gastrectomy in the study by Sipponen and six patients with antrum resected in the study by Vaananen were excluded from 2 x 2 tables.

Diagnostic performance

Twenty studies reported the performance of the panel test for the diagnosis of atrophic gastritis regardless of the location. When available we included the results for atrophic gastritis at any grade of severity. Figure 2 shows the paired forest plots of sensitivity and specificity with 95% CIs for each study included. Pooling the results from the studies produced the following summary estimates: sensitivity 74.7% (95%CI, 62.0% to 84.3%), specificity 95.6% (95%CI, 92.6% to 97.4%), positive likelihood ratio 16.9 (95%CI, 9.5 to 30.1) and negative likelihood ratio 0.26 (95%CI, 0.17 to 0.41). The summary hierarchical receiver operating characteristic curve shows the summary sensitivity and specificity and the 95% confidence and prediction regions (Figure 3). Using the median prevalence of atrophic gastritis across the studies of 27%, the negative predictive value of the panel test was 91% and the positive predictive value was 86%.

Both Figure 2 and Figure 3 showed a large heterogeneity between studies, in particular for sensitivity. We investigated if the performance of the panel test varied between studies according to characteristics associated with the index test (G-17b vs G-17s, PGI vs combination of PGI and PG-I/PG-II ratio), target condition (atrophic gastritis at any grade of severity vs moderate-severe atrophic gastritis), use of PPIs (no vs yes), setting (clinical studies vs population-based studies), study design (cohort vs case-control studies), country (Europe vs others countries) and type of publication (full text vs abstract) adding these as a covariate (one at the time) to the bivariate regression model. The use of G-17s ($p< 0.001$) and PPIs ($p=0.01$) accounted for some of the heterogeneity between the studies. The sub-group analysis by G-17 produced the following summary statistics: sensitivity (G-17b) 62% (95%CI, 49% to 73.5%) and (G17s) 91% (95%CI, 81% to 96%), specificity (G-17b) 96.1% (95%CI, 92.3% to 98.1%) and (G-17s) 92.3% (95%CI, 87.5% to 95.4%). The sub-group analysis by the use of PPIs produced a summary sensitivity and specificity of 80.5% (95%CI 68.3% to 88.7%) and 96.1% (95%CI 93.8% to 97.5%) without use of PPIs and 46% (95%CI 18% to 74%) and 89% (95%CI 78% to 99%) with the use of PPIs. The use of PGI/PGII ratio ($p=0.12$), moderate-severe atrophic gastritis as target condition ($p=0.22$), setting ($p=0.80$), study design ($p=0.09$), country ($p=0.29$) and the type of publication ($p=0.36$), on the other hand, had not effect on the summary estimates**.**

After calculation of summary estimates we produced Cook's distance to identify influential studies. Cook's distance showed that the study by Peitz (35) was particularly influential, followed by the study by Pasechnikov (18) and the study by Goni (53) (Figure 4). All the three studies were identified as outliers having the largest standardized residuals for sensitivity, with the study by Peitz being an outlier also for specificity (Supplementary Figure 5). However, after exclusion of the study by Peitz sensitivity slightly increased (74.7% vs 76.8%), but specificity did not change (95.6% vs 96.0%); on the other hand, sensitivity slightly increased (76.6%) after the exclusion of the study by Goni and slightly decreased (71.1%) after the exclusion of the study by Pasechnikov.

Pooling data from the seven studies that assessed the performance of the panel test in the diagnosis of the site of atrophic gastritis, the summaries sensitivities and specificities were 65.4% (95%CI, 40.3% to 84.1%) and 95.1% (95%CI, 88.8% to 97.8%) for the diagnosis of antrum-limited atrophic gastritis (Supplementary Figure 6), 70.4 % (95%CI, 49.0% to 85.5%) and 98.4% (95%CI, 96.1% to 99.3%) for the diagnosis of corpus-limited atrophic gastritis (Supplementary Figure 7), and 42.6% (95%CI, 22.5% to 65.4%) and 99.1% (95%CI,98.4% to 99.5%) for the diagnosis of both antrum and corpus atrophic gastritis (Supplementary Figure 8), respectively.

DISCUSSION

This meta-analysis included 20 studies assessing the accuracy of the combination of pepsinogens, gastrin-17 and anti-H. pylori antibodies serum assays for the diagnosis of atrophic gastritis, compared to histology; pooling data from these studies yielded a summary sensitivity of 74.7% (62.0% to 84.3%) and a summary specificity of 95.6% (92.6% to 97.4%). Based on the median prevalence of atrophic gastritis across the studies of 27%, which is very close to that estimated worldwide in the general population (around 30%) (54), the negative predictive value of the panel test was 91% and the positive predictive value was 86%; this implies that 91 out of 100 subjects with a negative test will be true negative for the presence of atrophic gastritis, while 86 of 100 subjects with a positive test will be true positive. Using the pooled likelihood ratios, we

calculated that if the median pre-test probability of atrophic gastritis was 27%, the post-test probability was 9% for subjects with a negative test and 86% for subjects with a positive test result.

Pooling data from seven studies produced a summary sensitivity of the panel test of 65.4% for the diagnosis of antrum atrophic gastritis, 70.4% for the diagnosis of corpus atrophic gastritis and 42.6% for both antrum and corpus atrophic gastritis; the summary specificity was higher than 95% for any site of atrophic gastritis.

Strengths and weaknesses of the study

A strength of this review is the comprehensive search of literature without restrictions on the language of publications; we also identified and included unpublished studies, which were reported as abstracts in international conferences proceedings, minimizing the risk of missing relevant studies. As there is not a powerful method of testing for publication bias in a meta-analysis of diagnostic accuracy studies (6), we are not able to assess the likely impact of unpublished studies on our results. However, the studies included in this systematic review are likely to be the majority on this topic and, in addition, unpublished studies would have to have been very large to change the findings of this meta-analysis. If data was missing, in particular for abstracts, we attempted to contact the study authors to obtain additional data to improve the assessment of the methodological quality and data extraction. Another strength of this study is the use of a multilevel statistical approach with a bivariate model, that is recommended for meta-analysis of diagnostic accuracy studies (6,9).

A weakness of our findings was the substantial heterogeneity between the results of the studies, in particular for sensitivity. However, a substantial between-study heterogeneity is a commonplace in meta-analysis of diagnostic test accuracy studies. The meta-regression analysis showed that the measurement of serum G-17 and the use of PPIs were significant sources of heterogeneity. A relevant finding of our meta-analysis was that the measurement of stimulated gastrin-17 with a protein rich drink increased the sensitivity of the test; the summary sensitivity of the panel test increased to 91% with G-17s. Our finding is in line with previous reports that showed that the use of a protein rich meal before the blood sampling improved the performance of G-17 as a biomarker of antrum atrophic gastritis (13,14,16). It is well known that several factors, such as a physiologic high acid secretion, are involved in the output of G-17 from the antral G cells (5). The protein

stimulus, which usually increases the secretion of G-17, may help distinguishing if a low G-17 serum level is really due to the antrum atrophic gastritis with consequent loss of G cell (G-17 is still low after protein stimulus) or instead it is due to other factors, such as a high acid secretion (G-17 is higher after protein stimulus) (5,57). Our finding suggests that the measurement of serum level of stimulated G-17 could improve the diagnostic performance of the panel test, likely improving its sensitivity in the diagnosis of antrum atrophic gastritis.

We also found that the use of PPIs markedly reduced the sensitivity of the panel test. It is well known that PPIs increase serum levels of G-17 through the reduction of hydrochloric acid; in addition, the trophic effect of the gastrin on parietal cells increases serum levels of pepsinogens (5,29); These are the likely reasons that could explain the increased probability of false negative results due to the use of PPIs.

Unfortunately, only few studies with a small sample size assessed the reliability of the test for the diagnosis of the location of atrophic gastritis; with this limitation, we found a slightly lower sensitivity of the panel test in detecting the site of atrophic gastritis, except for cases with both antrum and corpus atrophic gastritis where the sensitivity of the test was very low (42.6%). Moreover, the sensitivity of the test in diagnosing antrum-limited atrophic gastritis was just slightly lower than corpus-limited atrophic gastritis (65.4% vs 70.4%).

Another weakness of our meta-analysis is that our findings are based on studies with low methodological quality. Most studies did not enroll a consecutive o random sample of subjects; thus, the presence of selection bias may have been a source of error in the estimation of the diagnostic performance. The interval between gastroscopy and blood sampling was often unclear; if the interval was too long, a treatment with PPIs and/or antibiotics, or unknown factors, during this period may have altered serum levels of pepsinogens and/or gastrin, thus causing a misclassification bias. Finally, some studies did not include all participants recruited in the analysis: in fact, the number of subjects enrolled was different from that included in the results, and this may have introduced a bias in the summary pooled estimates.

Comparison with other studies

To our knowledge, this is the first systematic review with meta-analysis to assess the performance of the serum panel test for the diagnosis of atrophic gastritis regardless of the site, using a) a comprehensive literature search, b) the Updated Sydney System classification of gastritis as reference standard, c) an appropriate tool for the evaluation of the methodological quality of studies and d) a multilevel statistical approach for meta-analysis. A recent meta-analysis by Syrjanen reported the performance of the panel test "GastroPanel" (Biohit, Finland) for the diagnosis of antrum atrophic gastritis and corpus atrophic gastritis, separately, but not the accuracy of the test regardless of the site of atrophic gastritis (55). This meta-analysis included studies that used different histological classifications of atrophic gastritis as reference standard. The Updated Sydney System is the most widely accepted system for classifying and grading gastritis both in clinical practice and research (2). When a different classification system was used, such as the Houston System, a different prevalence of atrophic gastritis was reported (21). Including only studies that used the Updated Sydney system, we have, most likely, reduced the heterogeneity between studies and minimized the introduction of bias related to the reference standard (8). In the previous meta-analysis, another limitation was the lack of assessment of the methodological quality of the included studies, which is essential for assessing the strength of the results. In contrast, we carried out an appropriate assessment of the quality of studies using the recommended QUADAS-2 tool (6,8). Finally, Syrjanen used traditional meta-analysis applications which essentially consist in pooling weighted averages of sensitivities and specificities across all studies. As well known, meta-analyses of data from diagnostic accuracy studies require more complex and rigorous statistical methods that account for the correlation between sensitivity and specificity, i.e. multilevel statistical approaches (6,9,10). In order to achieve meaningful summary estimates of sensitivity and specificity, we have indeed used such multilevel statistical methods.

The meta-analysis by Syrianen reported a low sensitivity (51.6%) for the diagnosis of antrum atrophic gastritis with a better sensitivity (70.2%) for corpus atrophic gastritis. We found a similar sensitivity for corpus atrophic gastritis (70.4%), but a higher sensitivity for the diagnosis of antrum atrophic gastritis (65.4%).

Two different meta-analyses assessed the diagnostic performance of pepsinogens (56) and gastrin-17 serum assays separately (57). The pooled sensitivity and specificity were, respectively, 69% and 88% for

pepsinogens and 48% and 79% for gastrin-17 tests. The panel test seems to have a higher sensitivity (74.7%) and specificity (95.6%) than serum pepsinogens and gastrin-17 tests alone, and this is likely due to the use of both biomarkers of atrophic gastritis.

Conclusions and implications

The results of this meta-analysis suggest that the combination of pepsinogen, gastrin-17 and anti-Helicobacter pylori antibodies serum assays is a reliable tool for the diagnosis of atrophic gastritis. Given a prevalence (pre-test probability) of atrophic gastritis of 27%, the panel test would miss only 9 subjects for every 100 with atrophic gastritis (negative predictive value = 91%). We also found that the measurement of stimulated gastrin-17 increased the sensitivity of the panel test, likely increasing its sensitivity in the detection of antrum atrophic gastritis, whereas the use of PPIs had a markedly negative impact on sensitivity. Thus, our findings would support the use of a rich protein drink before the blood sampling and confirm the manufacturer‘s recommendation to stop using PPIs at least a week before the test. With the limitation of only few studies included, we found a lower sensitivity of the panel test in detecting the site of atrophic gastritis.

Therefore, our findings would support the use of a combination of pepsinogens, gastrin-17 and anti-H. pylori antibodies serum assays for screening subjects or populations, in order to identify individuals who are very likely to have atrophic gastritis to refer to endoscopy. However, a cost-effectiveness analysis to determine the role of this test in screening programs, aimed to reduce gastric cancer mortality, is needed. This test may play a relevant role in screening programs aimed to reduce gastric cancer mortality. Moreover, this test may be useful for epidemiological studies on the prevalence and incidence of atrophic gastritis in the general population. As most studies were conducted in Europe we think that our findings are certainly applicable to the European population.

However, well-designed high quality studies with a large sample size are needed to confirm the performance of the panel test in the diagnosis of atrophic gastritis, especially in Asia and America.

REFERENCES

1. Dixon MF, Genta RM, Yardley JH, Correa P. Classification and grading of gastritis. The Update Sydney System. International Workshop on the Histopathology of Gastritis. Houston 1994. Am J Surg Pathol 1996;20:1161-81.
2. Dinis-Ribeiro M, Areia M, de Vries AC, et al. Management of precancerous conditions and lesions in the stomach (MAPS): guideline from the European Society of Gastrointestinal Endoscopy (ESGE), European Helicobacter Study Group (EHSG), European Society of Pathology (ESP), and the Sociedade Portuguesa de Endoscopia Digestiva (SPED). Endoscopy 2012;44:74-94.
3. Malfertheiner P, Megraud F, O'Morain CA, et al. Management of Helicobacter pylori infection – the Maastricht V/Florence Consensus Report. Gut 2017;66:6-.
4. Sugano K, Tack J, Kuipers E, et al. Kyoto global consensus report on Helicobacter pylori gastritis. Gut 2015;64:1353-67.
5. Agréus L, Kuipers EJ, Kupcinskas L, et al. Rationale in diagnosis and screening of atrophic gastritis with stomach-specific plasma biomarkers. Scand J Gastroenterol. 2012;47:136-47.
6. Cochrane Collaboration, Diagnostic Test Accuracy Working Group: handbook for DTA reviews. http://srdta.cochrane.org/handbook-dta-reviews.
7. Sipponen P, Ranta P, Helske T, et al. Serum levels of amidated gastrin-17 and pepsinogen I in atrophic gastritis: an Serum levels of amidated gastrin-17 and pepsinogen I in atrophic gastritis: an observational case-control study. Scand J Gastroenterol. 2002;37:785-91.
8. Whiting PF, RutjesAW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529-36.
9. Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005;58:982-90.
10. Wang F, Gatsonis CA. Hierarchical models for ROC curve summary measures: design and analysis of multireader, multimodality studies of medical tests. Stat Med. 2008; 27:243-56.

11. Skrondal A, Rabe-Hesketh. Generalized latent variable modeling: multilevel, longitudinal, and structural equation models. Boca Raton, FL: Chapman & Hall/CRC. 2004
12. Väänänen H, Vauhkonen M, Helske T, et al. Non-endoscopic diagnosis of atrophic gastritis with a blood test. Correlation between gastric histology and serum levels of gastrin-17 and pepsinogen I: a multicentre study. Eur J Gastroenterol Hepatol. 2003;15:885-91.
13. Broutet N, Plebani M, Sakarovitch, Sipponen P, Megraud F. Pepsinogen A, pepsinogen C, and gastrin as markers of atrophic chronic gastritis in European dyspeptics. British Journal of Cancer 2003;88:1239-47.
14. Di Mario F, Moussa AM, Caruana P, et al. "Serological biopsy" in first-degree relatives of patients with gastric cancer affected by Helicobacter pylori infection. Scand J Gastroenterol. 2003;28:1223-7.
15. Hartleb M, Wandzel P, Walunga M, et al. Non-endoscopic diagnosis of multifocal atrophic gastritis;efficacy of serum gastrin-17, pepsinogens and Helicobacter pylori antibodies. Acta Gastro-Enterologica Belgica 2004;67:320-6.
16. Ricci C, Vakil N, Rugge M, et al. Serological markers for gastric atrophy in asymptomatic patients infected with Helicobacter pylori. Am J Gastroenterol. 2004;99:1910-5.
17. Pasechnikov VD, Chukov SZ, Kotelevets SM, et al. Possibility of non-invasive diagnosis of gastric mucosal precancerous changes. World J Gastroenterol. 2004;10:3146-50.
18. Pasechnikov VD, Chucov SZ, Kotelevets SM, et al. Invasive and non-invasive diagnosis of Helicobacter pylori-associated atrophic gastritis: A comparative study. Scand J Gastroenterol. 2005;40:297-301.
19. Germaná B, Di Mario F, Cavallaro LG, et al. Clinical usefulness of serum pepsinogens I and II, gastrin-17 and anti-Helicobacter pylori antibodies in the management of dyspeptic patients in primary care. Dig Liver Dis. 2005;37:501-8.
20. Nardone G, Rocco A, Staibano S, et al. Diagnostic accuracy of the serum profile of gastric mucosa in relation to histological and morphometric diagnosis of atrophy. Aliment Pharmacol Ther. 2005;22:1139-46.
21. Graham DY, Nurgalieva ZZ, El-Zimaity HMT, et al. Non-invasive versus histological detection of gastric atrophy in a Hispanic population in North America. Clin Gastroenterol Hepatol. 2006; 4:306-14.

22. Cao Q, Ran ZH and Xiao SD. Screening of atrophic gastritis and gastric cancer by serum pepsinogen, gastrin-17 and Helicobacter pylori immunoglobulin G antibodies. J Dig Dis.2007;8:15-22.

23. Valle Munoz J, Artaza Varasa T, Lopez Pardo R, et al. Serological diagnosis of atrophic gastritis with a combination of pepsinogen I and II, Gastrin-17 and anti-Helicobacter pylori antibodies. Gastroenterologia y Hepatologia 2007; 30:567-71

24. Koivusalo AI, Pakarinen MP, Kolho KL. Is GastroPanel serum assay useful in the diagnosis of Helicobacter pylori infection and associated gastritis in children? Diagn Microbiol Infect Dis.2007;57:35-8.

25. De Angelis GL, Cavallaro LG, Maffini V, et al. Usefulness of a serological panel test in the assessment of gastritis in symptomatic children. Dig Dis. 2007;25:206-13

26. Storskrubb T, Aro P, Ronkainen J, et al. Serum biomarkers provide an accurate method for diagnosis of atrophic gastritis in a general population: The Kalixanda study. Scand J Gastroenterol. 2008; 43:1448-55

27. Haj-Sheykholeslami A, Rakhishani N, Amirzargar A, et al. Serum pepsinogen I, pepsinogen II, and gastrin 17 in relatives of gastric cancer patients: comparative study with type and severity of gastritis. Clin Gastroenterol Hepatol. 2008;6:174-9

28. Guariso G, Basso D, Bortoluzzi CF, et al. GastroPanel: evaluation of the usefulness in the diagnosis of gastro-duodenal mucosal alterations in children. Clinica Chimica Acta 2009; 402:54-60.

29. Agreus L, Storskrubb T, Aro P, et al. Clinical use of proton-pump inhibitors but not H2-blockers or antacid/alginates raises the serum level of amidated gastrine-17, pepsinogen I and pepsinogen II in a random adult population. Scand J Gastroenterol. 2009; 44:564-70.

30. Ijima K, Abe Y, Kikuchi R, et al. Serum biomarker tests are useful in delineating between patients with gastric atrophy and normal, healthy stomach. World J Gastroenterol. 2009;15:853-9.

31. Leja M, Kupcinskas L, Funka K, et al. The validity of a biomarker method for indirect detection of gastric mucosal atrophy versus standard histopatology. Dig Dis Sci.2009;54:2377-84.

32. Melton SD, Genta RM, Souza RF. Biomarkers and molecular diagnostic tests in gastrointestinal tract and pancreatic neoplasms. Nat Rev Gastrenterol Hepatol 2010;7:620-8.

33. Lombardo L, Leto R, Molinaro G, et al. Prevalence of atrophic gastritis in dyspeptic patients in Piedmont. A survey using the GastroPanel test. Clin Chem Lab Med. 2010;48:1327-32.

34. Kikuki R, Abe Y, Iijima K, et al. Low serum levels of pepsinogen and gastrin 17 are predictive of extensive gastric atrophy with high risk of early gastric cancer. Tohoku J Exp Med. 2011:223:35-44.

35. Peitz U, Wex T, Vieth M, et al. Correlation of serum pepsinogens and gastrin-17 with atrophic gastritis in gastroesophageal reflux patients: a matched-pairs study. J Gastroenterol Hepatol. 2011;26:82-9.

36. Nasrollhazadeh D, Aghcheli K, Sotoudeh M, et al. Accuracy and cut-off values of pepsinogen I, II,and gastrin 17 for diagnosis of gastric fundic atrophy: influence of gastritis. Plos One 2011;6:e26-957.

37. Bornschein J, Selgrad M, Wex T, Kuester D, Malfertheiner P. Serological assessment of gastric mucosal atrophy in gastric cancer. BMC Gastroenterology 2012;12:10.

38. Noah Noah D, Okomo Assoumou MC, Eloumou Bagnaka SAF, et al. Assessing GastroPanel serum markers as a non invasive method for the diagnosis of atrophic gastritis and Helicobacter pylori infection. Open J Gastroenterol. 2012, 2:113-8.

39. Yakut M, Ormeci N, Erdal H, et al. The association between precancerous gastric lesions and serum pepsinogens, serum gastrin, vascular endothelial growth factor, serum interleukin-1Beta, serum toll-like receptor-4 levels and Helicobacter pylori Cag A status. Clin Res Hepatol Gastroenterol. 2013;37:302-11.

40. Shafaghi A, Mansour-Ghanaei F, Joukar F, et al. Serum gastrin and the pepsinogen I/II ratio as markers for diagnosis of premalignant gastric lesions. Asian Pacific J Cancer Prev 2013;14:3931-6.

41. Hosseini M, Amoueian S, Attaranzadeh A, et al. Serum gastrin 17, pespinogen I and pepsinogen II in atrophic gastritis patients living in North-East of Iran. J Res Med Sci. 2013;18:225-9.

42. McNicholl AG, Fornè M, Barrio J, et al. Accuracy of GastroPanel for the diagnosis of atrophic gastritis. Eur J Gastroenterol Hepatol. 2014;26:941-8.

43. Masci E, Pellicano R, Mangiavillano B, et al. GastroPanel test for non invasive diagnosis of atrophic gastritis in patients with dyspepsia. Minerva Gastroenterol Dietol. 2014;60:79-83.

44. Mansour-Ghanaei F, Joukar F, Rajpout Y, Hasandokht T. Screening of precancerous gastric lesions by serum pepsinogen, gastrin-17, anti-Helicobacter pylori and anti-Cag A antibodies in dyspeptic patients over 50 years old in Guilan province, north of Iran. Asian Pacific J Cancer Prev. 2014;15:7635-8.
45. Nejadi-Kelarijiani F, Roshandel G, Semnani S, et al. Diagnostic values of serum levels of pepsinogens and gastrin-17 for screening gastritis and gastric cancer in a high risk area in Northern Iran. Asian Pacific J Cancer Prev. 2014;15:1733-6.
46. Tu H, Sun L, Dong X, et al. Temporal changes in serum biomarkers and risk for progression of gastric precancerous lesions: a longitudinal study. Int J Cancer 2015;136:425-34.
47. Roman LD, Lukyanchuk R, Sablin OA, et al. Prevalence of H. pylori infection and atrophic gastritis in a population-based screening with serum biomarker panel (GastroPanel) in St. Petersburg. Anticancer Research 2016; 4129-38.
48. Pyurveyeva KV, Lapina TL, Ivashkin VT, et al. Serum pepsinogen I, pepsinogen II and gastrin-17 in diagnostics of atrophic gastritis. RZGGK 2005;3:48-51.
49. Zagari RM, Nicolini G, Casanova S, et al. Diagnosis of atrophic gastritis in the general population based on a combination of three non invasive tests [Abstract]. GUT 2002;51(Suppl11):A39.
50. de Korwin JD, Prucha V, Lozniewski A, et al. Non endoscopic diagnosis of atrophic gastritis using the GastroPanel/Gastrosoft [Abstract]. Helicobacter 2004;9:A561.
51. Cavallaro LG, Merli R, Liatopoulou S, et al. Accuracy of "serologic gastric biopsy" in a cohort of dyspeptic patients [Abstract]. Helicobacter 2005;10:A503.
52. Di Mario F, Rugge M, Cavestro GM, et al. Usefulness of serum pepsinogens to identify chronic atrophic gastritis [Abstract]. Gastroenterology 2011;140:5(Suppl1):S318.
53. Goni E, Franceschi M, Scarpignato C, Rugge M, Di Mario F. Relationship between GastroPanel and OLGA in the detection of chronic atrophic gastritis in a primary care setting [Abstract]. UEGJ 2015;3(5S):A297.
54. Marques-Silva L, Areia M, Elvas L, Dinis-Riberio M. Prevalence of gastric precancerous conditions: a systematic review and meta-analysis. Eur J Gastroenterol Hepatol. 2014;26:378-87.

55. Syrjanen K. A panel of serum biomarkers (GastroPanel) in non-invasive diagnosis of atrophic gastritis. Systematic review and meta-analysis. Anticancer Research 2016;36:5133-44.
56. Huang YK, Yu JC, Kang WM, et al. Significance of serum pepsinogens as a biomarker for gastric cancer and atrophic gastritis screening: a systematic review and meta-analysis. PlosOne 2015;10 (11):e0142080.
57. Wang X, Ling L, Li S, et al. The diagnostic value of gastrin-17 detection in atrophic gastritis. A Meta-analysis. Medicine 2016;95:e3599.

Acknowledgement: We acknowledge Maurizio Zani from the University of Bologna for his contribution in the design of the search terms for the systematic review.

Funding information: None.

Declaration of personal interests: None.

Authorship: Guarantors of the article: Rocco Maurizio Zagari, MD. **Author contributions**: RMZ and FB drafted the protocol. RMZ, SR, LHE and AV performed electronic searches and study selection. RMZ and SR extracted data and assessed the methodological quality of the studies. RMZ and DG did the statistical analysis and generated the figures. RMZ wrote the manuscript, and all the other authors revised the manuscript and approved the final version.

FIGURE LEGENDS

Figure 1. Flowchart of systematic literature search.

Figure 2. Forest plots of coupled sensitivity and specificity for atrophic gastritis regardless of the site. TP=true positive, FP=false positive, FN=false negative, TN=true negative.

Figure 3. Summary receiver operating characteristics plot of sensitivity and specificity for atrophic gastritis regardless of the site. Each circle indicates an individual study and it is sized according to the total number of subjects; solid spot in middle is summary sensitivity and specificity; inner and outer ellipses indicate 95% confidence region and prediction regions, respectively.

Figure 4. Influence analysis**.** Cut-off for declaring Cook's distance to be large=1 (20, four times the number of parameters (5) of the model / 20, number of studies).

7=Pasechnikov 2005 , 15= Peitz 2011, 19= Goni 2015.

Supplementary Figure 5. Scatter plot of standardized residuals for outliers detection.

1=Sipponen 2002, 2= Zagari 2002, 3= Vaananen 2003, 4=Hartleb 2004, 5=De Korwin 2004, 6= Pyurveyeva 2005, 7=Pasechnikov 2005, 8=Germanà 2005, 9=Nardone, 10=Cavallaro 2005, 11=Valle Munoz 2007, 12=Storskrub 2008, 13=Ijima 2009, 14=Lombardo 2010, 15=Peitz 2011, 16=Di Mario 2011, 17=Noah Noah 2012, 18=McNicholl 2014, 19=Goni 2015, 20=Roman 2016.

Supplementary Figure 6. Forest plots of coupled sensitivity and specificity for antrum-limited atrophic gastritis. TP=true positive, FP=false positive, FN=false negative, TN=true negative.

Supplementary Figure 7. Forest plots of coupled sensitivity and specificity for corpus-limited atrophic gastritis. TP=true positive, FP=false positive, FN=false negative, TN=true negative.

Supplementary Figure 8. Forest plots of coupled sensitivity and specificity for both antrum and corpus atrophic gastritis. TP=true positive, FP=false positive, FN=false negative, TN=true negative.

Prisma Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	4
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	4
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	4
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	5
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	6
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	6

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	6

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	6
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	6
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	7
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	7 -8
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	8
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	8
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	10
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	8
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	10-11
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	11
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	12-13
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	14
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	21

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097