ANTICANCER RESEARCH
International Journal of Cancer Research and Treatment
ISSN: 0250-7005

A Panel of Serum Biomarkers (GastroPanel®) in Non-invasive Diagnosis of Atrophic Gastritis. Systematic Review and Meta-analysis

KARI SYRJÄNEN[1,2]

[1]*Department of Clinical Research, Biohit Oyj, Helsinki, Finland;*
[2]*Molecular Oncology Research Center, Barretos Cancer Hospital, Barretos, SP, Brazil*

Reprinted from
ANTICANCER RESEARCH *36*: 5133-5144 (2016)

ANTICANCER RESEARCH *36*: 5133-5144 (2016)
doi:10.21873/anticanres.11083

A Panel of Serum Biomarkers (GastroPanel®) in Non-invasive Diagnosis of Atrophic Gastritis. Systematic Review and Meta-analysis

KARI SYRJÄNEN[1,2]

[1]*Department of Clinical Research, Biohit Oyj, Helsinki, Finland;*
[2]*Molecular Oncology Research Center, Barretos Cancer Hospital, Barretos, SP, Brazil*

Abstract. *Background/Aim: To meet the increasing demand of non-invasive tests for screening of gastric cancer (GC) risk, biomarker panel (GastroPanel®) (GP) was designed by Biohit Oyj as the first serological test for stomach health. The aim of the present study was to perform a systematic review and meta-analysis of all studies on GP in diagnosis of atrophic gastritis (AG). Materials and Methods: Studies were eligible, if i) GP was used to diagnose biopsy-confirmed AG of the corpus (AGC) and/or antrum (AGA) and ii) exact numbers were available to enable calculating sensitivity (SE) and specificity (SP). Comprehensive Meta-Analysis software was used with maximum likelihood meta-regression (R^2 analog). Effect size estimates (SE; SP, 95% confidence interval (CI)) were tested for homogeneity with Cochran's Q and I^2 statistics. Potential publication bias was estimated by funnel plot statistics. Results: Altogether, 27 studies were eligible comprising of 8,654 patients from different geographic regions. Significant heterogeneity between studies reporting AGC (n=27) or AGA (n=13) warranted random effects (RE) model for summary statistics. GP performs better in diagnosing AGC than AGA with 70.2% vs. 51.6% pooled SE and 93.9% vs. 84.1% pooled SP, respectively. Limited number of studies erodes the Q test's power to detect true heterogeneity in meta-analysis stratified by geographic study origin. Few hypothetical missing studies had only marginal effect on pooled estimates of SE and SP. Conclusion: This first meta-analysis of GP literature corroborates the statement of international experts, advocating GP in diagnosis and screening of AG. Due to its high specificity for both AGA and AGC, GastroPanel® is truly a test for stomach health.*

Correspondence to: Prof. Kari Syrjänen, MD, Ph.D., FIAC, Department of Clinical Research, Biohit Oyj, Laippatie 1, FIN-00880 Helsinki, Finland. Tel: 358 405566810, e-mail: kari.syrjanen@biohit.fi

Key Words: GastroPanel®, meta-analysis, sensitivity, specificity, biomarkers, pepsinogen I, pepsinogen II, gastrin-17, *Helicobacter pylori*, stomach mucosa, antrum, corpus, atrophic gastritis, meta-regression, gastric cancer, risk.

Gastric cancer (GC) remains the fifth most common malignancy worldwide with almost one million new cases and far over 700,000 annual deaths (1). This uniformly poor outcome of GC is due to the fact that, once diagnosed, the disease has already progressed beyond reach of curative therapy in most cases (2, 3). To improve the ominous disease outcome due to delayed diagnosis, novel diagnostic measures are urgently needed to allow early detection of GC (4).

Population-based screening by (invasive) endoscopy for early GC and its precursors is not feasible except perhaps in Japan (5). Atrophic gastritis (AG) and its causative etiological agent *Helicobacter pylori* (HP) are well-established precursors of non-cardiac GC (6-9). Therefore, non-invasive diagnostic tests for detection of AG and HP are promising tools for systematic screening of GC risk groups (5, 10-16). For some time, serum pepsinogen (PG) tests have been used for this purpose (5, 15); however, their impact on global GC mortality has been modest (17-19).

To meet the increasing demand, the GastroPanel® test (hereafter referred as GP) was designed in the late 1990's by Biohit Oyj (Helsinki, Finland) representing the first non-invasive diagnostic test for stomach health (13, 20, 21). This ELISA-based biomarker panel includes 3 markers of mucosal atrophy (PGI and PGII for the corpus; G-17 for the antrum), combined with HP IgG antibody assay (13, 16, 22). During the past decade, GP has been tested in different settings, mostly in diagnosis of symptomatic (dyspeptic) subjects (10, 13, 23-30). As emphasized (20-22), GP is not a test for invasive GC but designed for screening of the subjects at risk for GC, *i.e.* those with HP infection and AG. Until now, however, GP has been less extensively validated in population-based screening of such risk groups (11, 12, 30-33).

The results on GP's sensitivity (SE) and specificity (SP) in diagnosis of AG have been variable (10-13, 23-34). There are several potential sources of bias affecting GP performance in different settings (22, 25). Apart from the different cut-off values used for PGI, PGII, G-17 and HP IgG in studies from different geographic regions, the prime culprit for the heterogeneity between the published studies is the misclassification bias or sampling error of gastric biopsies (4, 16, 22, 35-37). This is closely related to the use of appropriate histological end points in calculating the performance indicators of the GP test (10, 22, 30, 37).

We performed a systematic review and meta-analysis of the studies reporting the use of GP test in diagnosis of AG, covering all reports published until May 2016, with no restrictions to the language or geographic origin. The main objective was to estimate the SE and SP of the GP biomarkers in diagnosis of AGA (AG of the antrum) and AGC (AG of the corpus), separately. The maximum likelihood (ML) meta-regression was used to estimate the proportion of total between-study variance explained by the covariates in the model.

Materials and Methods

Data abstraction. We identified eligible studies by searching MEDLINE and reference lists from original articles, book chapters, reviews and congress abstracts until May 2016. No language limitations were imposed. The search terms included: GastroPanel, atrophic gastritis, biomarkers, pepsinogens, gastrin-17, *Helicobacter pylori*, stomach, gastric and human. All publications appeared in peer-reviewed journals, as well as published congress abstracts were eligible, irrespective of the study design (*i.e.*, clinical or a screening setting). Most importantly, the eligible reports had to include GP test (Biohit Oyj, Helsinki, Finland) as the diagnostic tool. Thus, all studies using either pepsinogens or G-17 as stand-alone markers were excluded from this analysis as subjected to comprehensive meta-analysis previously (19). To be eligible, the report had to contain the exact numbers of gastric biopsies analyzed, as well as the numbers of AGC and/or AGA, examined by the GP biomarkers, pepsinogens and G-17, respectively (10, 22, 30). If not directly specified, the report had to be complete enough to enable the assessors to calculate these numbers from the results.

Using the above search terms, a database of 5,093 reports could be built by Reference Manager (Professional version 12.0.3.) covering the years 1936 through 2016. For the present meta-analysis, a total of 27 studies were determined eligible, fulfilling all the defined criteria. The absolute vast majority of the ineligible studies was irrelevant in type and appeared in the search due to a matching search term but otherwise lacking the necessary key identifiers needed for the meta-analysis. Another large category consists of studies where pepsinogens and/or G-17 were used as stand-alone markers in diagnosis of AG (19), without using the GP test.

Data extracted for meta-analysis. From the summaries and/or body texts of each eligible study, we extracted the following data: sample size (number of patients studied), topography of AGs (AGA or AGC), number of AGA and AGC, geographic origin of the study, authors and publication year. To calculate the SE and SP for GP, we extracted the precise numbers of TP (true-positive), FN (false-negative), FP (false-positive) and TN (true-negative) (Table I). This list of items represents a modification of the PRISMA statements for systematic reviews and meta-analysis (38), adopted specifically for this purpose, where the effect size represents the SE and SP recorded for each original study. The meta-analysis file was arranged using AG as the study variable, whereby AGA and AGC represent subgroups within the study.

The authors defined GP performance indicators separately for PGI and G-17 used to diagnose AGC and AGA, respectively (10, 22, 30). In the next step, the authors identified from each eligible study the cut-off values for PGI and G-17 positivity (Table I). According to their geographic origin, the studies were stratified into the following categories: Central and South America, Europe, China or Taiwan and Other Asia. No studies were available from North America or from Africa/Middle East. At study level, the country of origin was classified into one of three categories, according to the incidence of GC; high-, intermediate- and low-incidence country. Publication year was arbitrarily divided into two categories; before year 2008 and after 2008, for older and more recent study, respectively.

Statistical analysis. A specific software, Comprehensive Meta Analysis™ (Version 3.3.070; Biostat Inc., Englewood, NJ, USA), was used to perform the meta-analysis and meta-regression. The software calculates the SE and SP (and their logit event rate, standard error and variance) automatically when the numbers for TP/(TP+FN) and TN/(FP+TN) from the original data are entered as the events and the sample size, respectively, into the software's effect size data entry. To assess the overall heterogeneity in effect size estimates between the different studies, Cochran's Q homogeneity *p*-value, as well as I^2 statistics (for percentage of variation) were used (39). To explore the eventual publication bias, funnel plots were drawn by plotting the logit event rates by their precision (1/standard error) (40), evaluated for asymmetry using the following statistics: i) Begg and Mazumdar rank correlation (41), ii) Egger's test of the intercept (regression) (42) and iii) Duval and Tweedie's "trim and fill" method (43), which imputes the results that are hypothetically missing due to the publication bias.

To assess the variation in event rates due to the differences between individual studies, we evaluated the study characteristics using stratified meta-analysis and maximum likelihood (ML) meta-regression separately for both study subgroups (AGA, AGC) and for both effect size estimates (SE and SP). Meta-regression formally compares the differences in SE and SP across the selected study-level covariates and estimates the among-study variance (44). In meta-regression, different models were tested using the following study-level covariates in the model (collectively or as individual combinations): country, geographic origin, GC incidence (high-, intermediate-, low-), biomarker cut-off values, cohort size and time of publication (early, recent). In the new version of the software, the ML meta-regression estimates the proportion of total between-study variance explained by the set of covariates in the model, expressed as the R^2 analog (44).

Sensitivity analysis was performed to assess the influence of each individual study on the strength and stability of the meta-analytical results, using the one-by-one study removal and evaluated by descriptively comparing the magnitude and precision of the random-effects summary event rates.

Table I. *Studies reporting GastroPanel sensitivity and specificity in diagnosing biopsy-confirmed atrophic gastritis of the corpus (AGC) and/or antrum (AGA).*

Sample size	AG topography	Sensitivity			Specificity			Country	Authors	Year	Ref No
		TP	FN	Sensitivity (95%CI)	FP	TN	Specificity (95%CI)				
104	AGC	32	3	91.4 (76.9-98.2)[1]	8	61	88.4 (78.4-94.9)	Italy	Zagari *et al*.	2002	33
	AGA	NA	NA		NA	NA					
13	AGC	2	2	50.0 (6.8-93.2)[6]	0	6	100 (66.4-100)	Italy	DiMario *et al*.	2003	45
	AGA	NA	NA		NA	NA					
404	AGC	36	9	80.0 (65.4-90.4)[2]	7	352	98.1 (96.0-99.2)	Finland	Väänänen *et al*.	2003	13
	AGA	2	2	50.0 (6.8-93.2)[3]	5	180	97.3 (93.8-99.1)				
55	AGC	6	0	100 (54.1-100)[4]	4	45	91.8 (80.4-97.7)	Poland	Hartleb *et al*.	2004	46
	AGA	7	4	63.6 (30.8-89.1)[5]	4	40	90.9 (78.3-97.5)				
178	AGC	46	16	74.2 (61.5-84.5)[6]	17	99	85.3 (77.6-91.2)	Russia	Pasechnikov *et al*.	2004	47
	AGA	105	30	77.8 (69.8-84.5)[5]	12	31	72.1 (56.3-84.7)				
109*	AGC	28	8	77.8 (60.8-89.9)[6]	3	70	95.9 (88.5-99.1)	Russia	Pasechnikov *et al*.	2005	48
	AGA	25	17	59.5 (43.3-74.4)[5]	9	58	86.6 (76.0-93.7)				
175	AGC	16	5	76.2 (52.8-91.8)[6]	4	150	97.4 (93.5-99.3)	Italy	Cavallaro *et al*.	2004	23
	AGA	NA	NA		NA	NA					
50	AGC	9	3	75.0 (42.8-94.5)[6]	0	38	100 (97.5-100)	France	De Korwin *et al*.	2004	24
	AGA	NA	NA		NA	NA					
287	AGC	38	22	63.3 (49.9-75.4)[6]	16	211	93.0 (88.8-95.9)	Italy	Germanà *et al*.	2005	18
	AGA	NA	NA		NA	NA					
94	AGC	7	3	70.0 (34.8.93.3)[6]	4	80	95.2 (88.3-98.7)	Italy	Nardone *et al*.	2005	29
	AGA	11	9	55.0 (31.5-76.9)[9]	4	70	94.6 (86.7-98.5)				
180^	AGC	2	12	14.3 (1.8-42.8)[6]	0	145	100 (97.5-100)	Mexico	Graham *et al*.	2006	49
	AGA	20	20	50.0 (33.8-66.2)[8]	50	69	58.5 (48.6-68.0)				
56	AGC	7	1	87.5 (47.3-99.7)[6]	0	48	100 (92.6-100)	Spain	Valle *et al*.	2007	50
	AGA	NA	NA		NA	NA					
71	AGC	5	2	71.4 (29.0-96.2)[6]	2	13	86.7 (59.5-98.3)	Russia	Reshetnikov *et al*.	2008	51
	AGA	3	11	21.4 (4.7-50.8)[10]	1	12	92.3 (64.0-99.8)				
976	AGC	44	18	71.0 (58.1-81.8)[5,6]	20	894	97.8 (96.6-98.7)	Sweden	Storskrubb *et al*.	2008	31
	AGA	NA	NA		NA	NA					
162	AGC	6	7	46.2 (19.2-74.9)[2]	4	145	97.3 (93.3-99.3)	Japan	Iijima *et al*.	2009	26
	AGA	2	11	15.4 (1.9-45.4)[7]	5	144	96.6 (92.3-98.9)				
313#	AGC	3	3	100 (29.2-100)[1]							

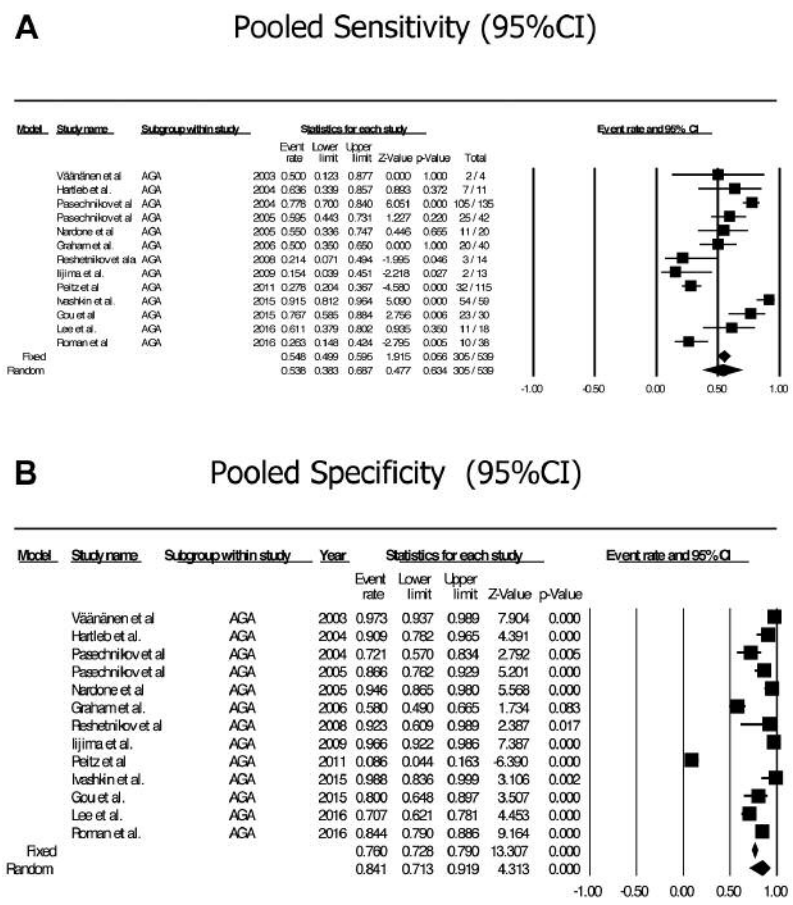

Figure 1. *Pooled sensitivity and specificity of GastroPanel in diagnosis of AG of the antrum.*

Results

Eligible studies. A total of 27 studies were considered eligible for the present analysis (13, 18, 23,-31, 33, 45-60) comprising of 8,654 patients analyzed by the GP test. Included are both small series (45, 54) and larger cohorts comprising up to 2,858 analyzed patients (31, 59, 60) (Table I). The studies include series where GastroPanel biomarkers were tested in diagnosing either AGA or AGC or both. In 14/27 reports, GP results on AGA are missing (18, 23-25, 28, 31, 33, 45, 50, 52-55, 57). As to the geographic origin of the studies, there were no studies published in North America and in Africa/Middle East.

Analytical results

Pooled estimates of sensitivity and specificity. The pooled estimates of sensitivity and specificity of the GP test in diagnosis of AG were analyzed separately for the AG subgroups: AGA and AGC. Starting from AGA (n=13), the crude sensitivity (305=TP/539=TP+FN) translates to the pooled sensitivity of 0.548 (95% confidence interval (CI)=0.499-0.595), using the fixed effects (FE) model, and 0.538 (95%CI=0.383-0.687), using the random effects (RE) model (Figure 1A). Homogeneity (Cochran's Q)=107.525; I^2=88.840; and *p* for homogeneity *p*=0.0001. For specificity, the crude figures (961=TN/1215=FP+TN) decode to a pooled specificity of 0.760 (95%CI=0.728-0.790) with the FE model and 0.841 (95%CI=0.713-0.919) with the RE model (Figure 1B). Homogeneity (Cochran's Q)=200.372; I^2=94.011; and *p* for homogeneity *p*=0.0001.

Figure 2 depicts the meta-analytical results of pooled sensitivity and specificity for GP in diagnosis of AGC. The crude sensitivity (487/685) translates to pooled sensitivity of 0.704 (95%CI=0.667-0.739) with the FE model and 0.702 (95%CI=0.643-0.775) using the RE model (Figure 2A). Cochran's Q=52.807; I^2=50.765; and *p* for homogeneity *p*=0.001. The crude specificity (3,946/4,171) corresponds to the pooled specificity of 0.924 (95%CI=0.914-0.934) using the FE model and 0.939 (95%CI=0.910-0.960) with the RE model (Figure 2B). Cochran's Q=196.475; I^2=86.767; and *p*=0.0001.

Table II summarizes the meta-analytical results for GP sensitivity and specificity in diagnosis of AGC, stratified by the geographic origin of the study. As to the test sensitivity,

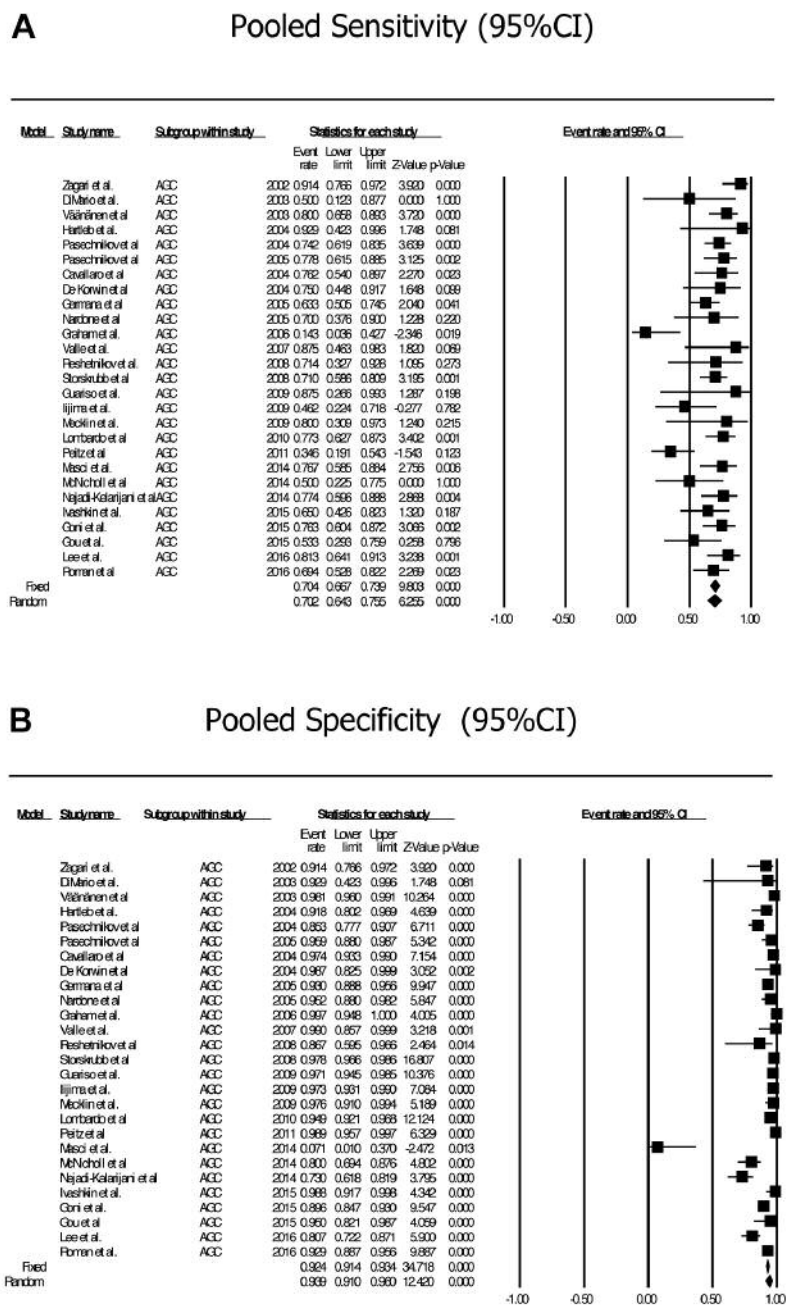

Figure 2. *Pooled sensitivity and specificity of GastroPanel in diagnosis of atrophic gastritis of the corpus (AGC).*

heterogeneity is significant in the overall comparison between strata (RE model, p=0.017). Using the RE model, studies from Europe give the highest pooled estimates of sensitivity (72.1%). As to the test specificity, there is a significant heterogeneity between the studies (n=26) from China/Taiwan, Europe and Other Asia, with p=0.048, p=0.0001 and p=0.0001, respectively. However, heterogeneity is not significant in the overall comparison between strata (RE model, p=0.199). Using the RE model, studies from Europe give the by far highest pooled estimates of specificity (0.943, *i.e.* 94.3%), followed by those from Other Asia (89.8%).

Table III shows the meta-analytical results for GP sensitivity and specificity in diagnosis of AGA, stratified by the geographic origin of the study. As to the test sensitivity, heterogeneity is not significant in the overall comparison between strata (RE model, p=0.483). Using the RE model, studies from China/Taiwan give the highest pooled estimates of sensitivity (69.6%), followed by those from Europe (54.5%). Regarding the test specificity,

Table II. *Pooled sensitivity and specificity of GastroPanel for atrophic gastritis of the corpus (AGC) in stratified meta-analysis by geographic study origin.*

Geographic origin of study	No. of studies	Events	Sample size	Point estimates of effect size (FE model)		Point estimates of effect size (RE model)		Heterogeneity (Cochran's Q)	**I-squared (I^2)	Heterogeneity (p-value)
				Point Estimate	95%CI	Point Estimate	95%CI			
Sensitivity										
China/Taiwan	2	34[1]	47[2]	0.709	0.555-0.826	0.701	0.496-0.848	3.756	73.374	p=0.053
Europe	22	421[1]	580[2]	0.715	0.675-0.752	0.721	0.663-0.773	32.519	35.423	p=0.052
Other Asia	2	30[1]	44[2]	0.671	0.512-0.799	0.656	0.444-0.820	3.890	74.293	p=0.049
S/C America	1	2[1]	14[2]	0.143	0.036-0.427	0.143	0.030-0.474	0.000	0.000	p=1.000
Summary	27	487[1]	685[2]	0.704	0.677-0.739	0.615	0.418-0.780	52.807	50.765	p=0.001
Total within (FE)								40.165		p=0.015
Total between (FE)								12.643		p=0.005
Total between (RE)								10.136		p=0.017
Specificity										
China/Taiwan	2	126[3]	149[4]	0.830	0.757-0.885	0.884	0.642-0.970	3.904	74.385	p=0.048
Europe	22	3.476[3]	3.654[4]	0.938	0.929-0.947	0.943	0.913-0.963	126.345	83.379	p=0.0001
Other Asia	2	199[3]	223[4]	0.823	0.747-0.880	0.898	0.686-0.972	20.729	95.176	p=0.0001
S/C America	1	145[3]	145[4]	0.997	0.948-1.000	0.997	0.914-1.000	0.000	0.000	p=1.000
Summary	27	3.946[3]	4.171[4]	0.924	0.914-0.934	0.936	0.846-0.975	196.475	86.767	p=0.0001
Total within (FE)								150.977		p=0.0001
Total between (FE)								45.498		p=0.0001
Total between (RE)								4.654		p=0.199

**Only calculated for fixed effects model: FE, fixed effects model; RE, random effects model; CI, confidence interval; [1]Events=TP; [2]Sample size=TP+FN; [3]Events=TN; [4]Sample size= FP+TN; S/C, South and Central.

heterogeneity is not significant in the overall comparison between strata (RE model, p=0.532). Excluding the regions with only one study, the studies from Europe give the highest pooled estimates of specificity (RE model) (86.0%), followed by those from China/Taiwan (75.6%).

Meta-regression. In ML meta-regression for GP sensitivity in diagnosis of AGA, the model contained the following covariates: GC incidence (high-, intermediate-, low-), G-17 cut-off values, cohort size, study timing, resulted in an R^2 analog of 0.610, *i.e.* 61.0% of the variance in true effects can be explained by these covariates. As to GP specificity for AGA, 95% of the variance (R^2 analog=0.950) could be explained by the model with the following covariates: geographic study origin, G-17 cut-off values and study timing. Changing the actual study year by the covariate early/recent, resulted in a model explaining the between-study variance in full (R^2 analog=1.00).

As to GP sensitivity for AGC, 96% of the variance (R^2 analog=0.960) in true effects could be explained by the model with the following covariates: geographic study origin, GC incidence (high-, intermediate-, low-) and PGI cut-off values. Adding the cohort size as the covariate, resulted in a model explaining the between-study variance in full (R^2 analog=1.00). As to GP specificity for AGC, the best fitting model included the following covariates: geographic study origin, GC incidence (high-, intermediate-, low-), PGI cut-off values, early/recent study) and explained 81% of the variance (R^2 analog=0.810) in true effects.

Publication bias. Publication bias was analyzed using precision funnel plots and the test statistics (Figure 3). Among studies reporting GP sensitivity in AGA end point, there was no evidence for publication bias; Begg p=0.541, Egger's p=0.923, the Duval and Tweedie's trim and fill method identified no missing studies (RE, to left or right of mean) (Figure 3A). For studies reporting GP specificity for AGA, the same was true with Begg p=0.179 and Egger's p=0.216; the Duval and Tweedie's trim and fill method imputed 4 hypothetically missing studies with a substantial effect on pooled estimate (RE) of specificity (from 84.1% to 72.8%) (Figure 3B).

Among studies reporting GP sensitivity for the AGC end point, there was no evidence for publication bias analyzed by the Begg (p=0.851) or Egger's (p=0.988) tests. The Duval and Tweedie's trim and fill method (RE) identified 4 missing

Table III. *Pooled sensitivity and specificity of GastroPanel for atrophic gastritis of the antrum (AGA) in stratified meta-analysis by geographic study origin.*

Geographic origin of study	No. of studies	Events	Sample size	Point estimates of effect size (FE model)		Point estimates of effect size (RE model)		Heterogeneity (Cochran's Q)	**I-squared (I^2)	Heterogeneity (p-value)
				Point Estimate	95%CI	Point Estimate	95%CI			
Sensitivity										
China/Taiwan	2	34[1]	48[2]	0.703	0.558-0.817	0.696	0.295-0.926	1.295	22.783	p=0.062
Europe	9	249[1]	438[2]	0.543	0.488-0.597	0.545	0.348-0.729	95.411	91.615	p=0.0001
Other Asia	1	2[1]	13[2]	0.154	0.039-0.451	0.154	0.012-0.728	0.000	0.000	p=1.000
S/C America	1	20[1]	40[2]	0.500	0.350-0.650	0.500	0.090-0.910	0.000	0.000	p=1.000
Summary	13	305[1]	539[2]	0.548	0.499-0.595	0.522	0.285-0.749	107.525	88.840	p=0.0001
Total within (FE)								96.706		p=0.0001
Total Between (FE)								10.818		p=0.013
Total Between (RE)								2.457		p=0.483
Specificity										
China/Taiwan	2	119[3]	163[4]	0.728	0.654-0.791	0.756	0.263-0.964	1.299	22.999	p=0.254
Europe	9	629[3]	784[4]	0.808	0.768-0.843	0.860	0.677-0.947	149.138	94.636	p=0.0001
Other Asia	1	144[3]	149[4]	0.966	0.922-0.986	0.966	0.559-0.998	0.000	0.000	p=1.000
S/C America	1	69[3]	119[4]	0.580	0.490-0.665	0.580	0.063-0.966	0.000	0.000	p=1.000
Summary	13	961[3]	1.215[4]	0.760	0.728-0.790	0.839	0.606-0.947	200.372	94.011	p=0.0001
Total within (FE)								150.436		p=0.0001
Total between (FE)								49.936		p=0.0001
Total between (RE)								2.200		p=0.532

**Only calculated for fixed effects model: FE, fixed effects model; RE, random effects model; [1]Events=TP; CI, confidence interval; [2]Sample size=TP+FN; [3]Events=TN; [4]Sample size= FP+TN; S/C, South and Central.

studies (left of mean) with a marginal effect on pooled estimate (RE) of sensitivity (from 70.2% to 68.1%) (Figure 3C). For studies reporting GP specificity for AGC, the same was true with Begg (p=0.851) and Egger's (p=0.229) tests, whereas the Duval and Tweedie's trim and fill (RE, left of mean) method imputed 5 hypothetically missing studies with only a slight effect on pooled estimates (RE) of specificity: from 93.9% (observed) to 92.5% (adjusted) (Figure 3D).

Sensitivity analysis. In general, the meta-analytical results seemed quite robust to all (n=27 AGC, n=13 AGA) one-by-one study removals with little change in the magnitude and precision of the FE- and RE-pooled estimates of GP sensitivity and specificity. However, among studies reporting sensitivity for AGA, there are two influential studies (27, 47) that, if omitted, have an impact on pooled sensitivity estimates (FE) from 54.8% up to 63.% (27) and down to 46.8% (47), respectively. For studies reporting GP sensitivity in AGC, no influential studies were identified by either FE or RE models. As to GP specificity in AGA, two studies had a moderate impact on pooled estimates but only in FE model (17, 49); once removed, the pooled specificity estimate improved from 76.0% to 79.6% and 80.0%, respectively. GP specificity was absolutely robust to all one-by-one study removals, with each study affecting the pooled estimates with less than 0.010, (*i.e.* <1% only), irrespective whether FE or RE model was used (data not in Tables).

Discussion

GastroPanel® test is a biomarker panel based on simultaneous analysis of PG-I, PG-II, amidated G-17 and HP IgG antibodies, designed to give information on both the structure and function of the stomach mucosa (13, 20-22). The added value of using this 4-biomarker combination instead of PGs and G-17 as stand-alone markers (14-16, 19) lies in the fact that this single test provides comprehensive information from the entire stomach, not restricted to either antrum or corpus only (10-13, 16, 20-22). GP results are interpreted by a specific software (GastroSoft®), classifying the test results into one of five categories, each with a specific biomarker profile (10-13, 16, 20-22, 30-32) and matching the diagnostic categories of the Updated Sydney System (USS) of classifying gastritis (35-37).

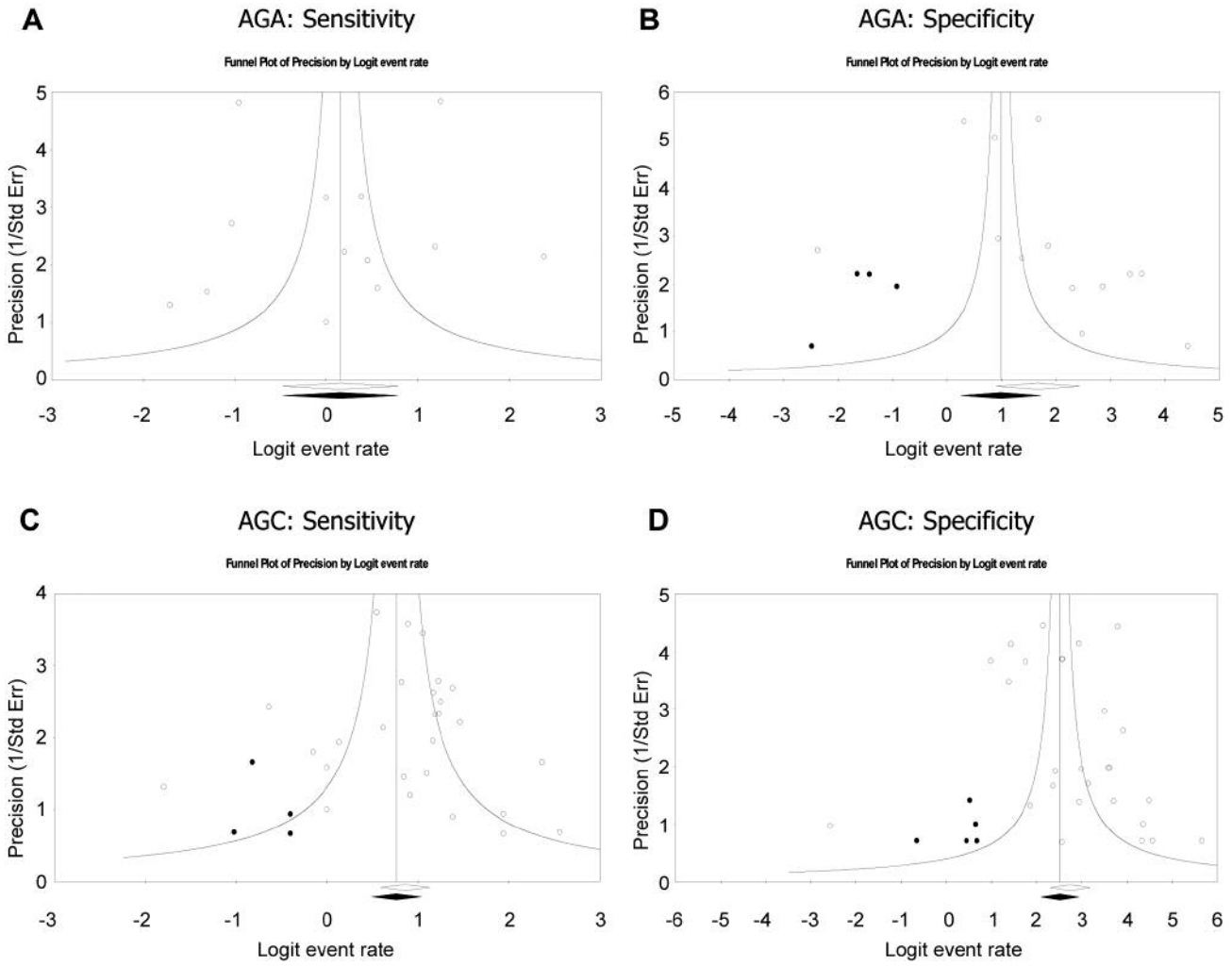

Figure 3. *Publication bias of the GastroPanel studies for atrophic gastritis of the antrum (AGA) and atrophic gastritis of the corpus (AGC) end points estimated by precision funnel plots.*

The experience on GastroPanel® test has been variable and, in part, conflicting, despite the recent consensus statement by an international panel of experts advocating its use in diagnosis and screening of AG (16). To cast further light on the potential causes of the between-study variance in the reported sensitivity and specificity of GP test, we performed a systematic review and meta-analysis covering all published studies where GP was used in diagnosis of AGA and AGC. Different GP biomarkers are used for AGA and AGC, G-17 and PGI, respectively (13, 16, 22, 30-32); AGA and AGC were treated as subgroups of AG in the present meta-analysis (Table I).

To assess the heterogeneity in meta-analysis is essential because the presence or absence of true heterogeneity (*i.e.* between-study variance) directly affects the statistical model that should be applied to analyze the meta-data (39, 64-67). Using the Q test, introduced by Cochran (67), a non-significant homogeneity *p*-value justifies the adoption of a fixed-effects (FE) model, assuming that the estimated effect sizes only differ by sampling error (64). In contrast, significant *p*-values in the Q test indicate true heterogeneity, advocating the use of a random effects (RE) model that includes both within- and between-studies variance. The I^2 index measures the extent of true heterogeneity, interpreted as the percentage of the total between-study variance among the effect sizes (68). One of its definite advantages is that the I^2 indices obtained from meta-analyses with different numbers of studies and different effect metrics are directly comparable (64, 68).

There is little doubt that a marked heterogeneity exists between the studies reporting GP sensitivity and specificity in both AGA and AGC (Figures 1 and 2; see Results for the exact Q-, I^2-index- and *p*-values). This advocates the use of the RE model to analyze all the summary statistics (39, 64-68). As shown by the forest plots for AGA and AGC (Figures 1 and 2), GP test performs better in diagnosing AGC than AGA. The difference is more marked in test sensitivity, with pooled (RE) estimates of 0.538 and 0.702 in diagnosis of AGA and AGC, respectively. The difference is less marked in test specificity for AGA and AGC, 0.841 and 0.939, respectively. Even more marked between-study heterogeneity was reported in a meta-analysis of the studies (n=12) using PGI and/or PGI/PGII as stand-alone markers for AGC, with SE varying between 5.8% and 98.6% and SP between 64.0% and 100% (19).

In practical terms, PGI biomarker of the GP detects AGC with 70.2% pooled sensitivity and 93.9% pooled specificity, whereas G-17 detects AGA with 53.8% pooled sensitivity and 84.1% pooled specificity. These meta-analytical results are not unexpected and the pooled estimates of GP sensitivity and specificity closely match those obtained in our own validation studies (13, 31, 32, 60). The reasons for this different GP performance in AGC and AGA are to be found in the different target-specificity and functional regulation of PGI and G-17 biomarkers (13, 22). As PGI is a product of gastric corpus, and the below-cut-off levels (30 μg/l) are only possible when the chief cells disappear as the result of mucosal atrophy, PGI is a highly sensitive and specific marker of AGC (10, 14-16, 22). G-17, in turn, is a specific biomarker of the antrum expressed by the G-cells that disappear in AG of the antrum (22, 48-50).

However, it is well-established that low levels of G-17b are not exclusively inherent to AGA but also result from high gastric acid output by the corpus (22, 69-71). On the other way round, G-17 is up-regulated (through a negative feedback loop) by low acid content of the corpus, caused by either i) AGC or ii) a prolonged use of proton pump inhibitor (PPI) medication (10, 14-16, 22, 69-71). By definition, any biomarker that is being regulated by more than one trigger cannot be a highly specific indicator of only one of them. In the case of fasting G-17b, the below-cut-off values can be due to either AGA or high acid output (22). Thus, for a distinction between these two potential causes of low G-17, G-17b measurement alone is not sufficient but it should be complemented by measuring G-17 levels after protein stimulation, *i.e.* G-17s, which is another component of the GP test (22). Failure to increase G-17s output after such a stimulation implicates the lack of G-cells and the presence of AGA (22, 69-71). Of the studies included in this meta-analysis, only four had included G-17s measurement in their GP analysis (13, 46-48). This precludes the possibility of making the distinction between AGA and high acid output as the cause of low G-17 levels (69-71). Failure to do so inevitably corrodes the sensitivity of G-17 as the marker of AGA, levelling off at 53.8% in the pooled (RE) analysis (Figure 1).

GC and its precursors (HP infection and AG) demonstrate a significant geographic variation in their global occurrence (1, 2, 10-13, 16, 19, 30-32). Accordingly, geographic origin of the GP studies clearly is a potential study-level covariate, here addressed by stratified meta-analysis. Interestingly, for AGA, the stratified meta-analysis using Q test and I^2 index did not disclose a marked heterogeneity between the studies from the four geographic regions (Table III). The most likely explanation is the limited number of studies reporting AGA, with two of the regions only having one single study, thus limiting the power of Q test to disclose true heterogeneity (39, 64-67). As to AGC, the stratified meta-analysis confirmed substantial heterogeneity only for the sensitivity (Table II). For specificity, there was no significant between-region variance (p=0.199). The small number of studies from three of the regions (n=2, n=2 and n=1) is probably the most important limiting factor for adequate power of the Q test to detect true heterogeneity.

Finally, we performed the ML meta-regression to estimate the proportion of total between-study variance explained by the set of covariates in the model. In the new version of the meta-analysis software, this is expressed as the R^2 analog (61-63), which is analogous to the R^2 index commonly reported for the proportion of variance explained by covariates in the primary studies. In the primary studies, all observations are given the same weight, while, in meta-regression, each study is given a different weight (61). In our meta-regression, the following study-level covariates were included in the model: country, geographic origin, GC incidence (high-, intermediate-, low-), biomarker cut-off values, cohort size, year of publication, early/recent study.

For the AGA end-point, we could build-up a regression model that explained at beast 61% of the variance in true sensitivity of the GastroPanel® test. For test specificity, in turn, we could find a set of covariates that explained the true variance by 100%, *i.e.* R2 analogue was 1.0. This is in aligment with the view that G-17b is not a highly sensitive marker of AGA as it is affected by both AGA and high acid output (22, 69-71). This information was not available as a covariate in the present meta-analysis. However, this information is not necessary to correctly classify the antrum as normal (AGA-negative) by normal G-17b levels, implicating that the existing covariates are sufficient to fully explain the true between-study variance in GP specificity.

In AGC, the reverse is true, *i.e.* the existing covariates resulted in a model explaining the between-study variance of sensitivity in full but the true variance of specificity by 81% only. This suggests that the below-cut-off levels of PGI are sensitive indicators of AGC, whereas additional covariates are needed to classify gastric corpus healthy (AGC-negative). One such candidate is the presence of HP infection, as well as the

PGI/PGII ratio, which both are essential components of the full GastroPanel® test; however, not included as separate covariates in the present meta-analysis (because incompletely reported). According to our experience, PGI/PGII ratio is another powerful marker of AGC, equivalent or sometimes superior to PGI alone (11-13, 16, 22, 30-32, 60). It might well be that studies using PGI alone might misclassify a few cases of AGC, being responsible for the between-study variance in test specificity that is not being explained by the set of available covariates. Importantly, the GastroSoft® algorithm also takes into account G-17 while interpreteing the GP results; the complete biomarker profile of AGC includes i) low PGI, ii) low PGI/PGII ratio, as well as iii) increased G-17b (16, 22, 30-32, 69-71). These data were not available from the original studies for use as covariates in this analysis. Similarly, diagnosis of AGA is made by GastroSoft® only when both conditions: i) low G-17 and ii) HP Ab titer above the cut-off (30 EIU) are fulfilled (22, 69, 70). As said, G-17s is needed in confirming the antrum atrophy.

Taken together, the present meta-analysis disclosed significant between-study heterogeneity in reported sensitivity and specificity for both AG end points. GastroPanel® performs better in diagnosing AGC than AGA, with 70.2% *vs*. 53.8% pooled sensitivity and 93.9% *vs*. 84.1% pooled specificity, respectively. In meta-analysis, stratified by the geographic origin of the study, the between-region heterogeneity was less pronounced than anticipated on the basis of the highly divergent geographic variation in GC and its precursor lesions (10-13, 16, 19, 30-32). Undoubtedly, the limited number of studies per strata erodes the power of the Q test to detect true heterogeneity in this stratified analysis.

Of the potential sources of errors affecting GastroPanel® performance in different studies, most important is the misclassification bias of gastric biopsies (22, 35, 37). This can result from two possible sources: a) failure to identify the correct biopsy site on gastroscopy (*e.g*. patchy AG) and b) unsatisfactory intra- and inter-observer reproducibility of grading the AG (22, 35-37). The different potential of G-17 and PGI to diagnose AGA and AGC, respectively, is inherent to their different physiological regulation (22, 69-71). GastroPanel® is an entity of 4 stomach-specific biomarkers but the value of this diagnostic test is much more than the sum of its components. GastroPanel® gives comprehensive information on the structure and function of the gastric mucosa and diagnosing AGA and AGC is just one part of the diagnostic range of this unique test. As suggested by the independent panel of international experts, GastroPanel® should be the first-line test in diagnosis and screening of AG (16). The present study, reporting the first formal meta-analysis of the GastroPanel® literature, clearly corroborates this statement; due to its high specificity in diagnosing both AGA and AGC, GastroPanel® is truly a test for stomach health (16, 22) with excellent longitudinal negative predictive value for GC (30).

References

1 Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C and Rebelo M: Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer *136*: E359-386, 2015.

2 Correa P, Haenszel W, Cuello C, Zavala D, Fontham E and Zarama G: Gastric precancerous process in a high risk population: cohort follow-up. Cancer Res *50*: 4737-4740, 1990.

3 Correa P: The epidemiology of gastric cancer. World J Surg *15*: 228–234, 1991.

4 Rugge M, Capelle LG, Cappellesso R, Nitti D and Kuipers EJ: Precancerous lesions in the stomach: from biology to clinical patient management. Best Pract Res Clin Gastroenterol *27*: 205-223, 2013.

5 Miki K: Gastric cancer screening by combined assay for serum anti-*Helicobacter pylori* IgG antibody and serum pepsinogen levels - "ABC method". Proc Jpn Acad Ser B Phys Biol Sci *87*: 405-414, 2011.

6 Correa P: Chronic gastritis: a clinico-pathological classification. Am J Gastroenterol *83*: 504-509, 1988.

7 Filipe MI, Munoz N, Matko I, Kato I, Pompe-Kirn V and Jutersek A: Intestinal metaplasia types and the risk of gastric cancer: a cohort study in Slovenia. Int J Cancer *57*: 324-329, 1994.

8 Ohata H, Kitauchi S, Yoshimura N, Mugitani K, Iwane M and Nakamura H: Progression of chronic atrophic gastritis associated with *Helicobacter pylori* infection increases risk of gastric cancer. Int J Cancer *109*: 138-143, 2004.

9 Correa P: Human gastric carcinogenesis:a multistep and multifactorial process –First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. Cancer Res *52*: 6735-6740, 1992.

10 Benberin V, Bektayeva R, Karabayeva R, Lebedev A, Akemeyeva K, Paloheimo L and Syrjänen K: Prevalence of *H. pylori* infection and atrophic gastritis among asymptomatic and dyspeptic adults in Kazakhstan. A Hospital-Based screening with a panel of serum biomarkers. Anticancer Res *33*: 4595-4602, 2013.

11 Vohlonen I, Syrjänen K and Karhu J: Serum biomarker-based screening of the risk groups of stomach cancer. Preliminary results of a feasibility study in Finland. Güncel Gastroenteroloji *17*: 170-175, 2013.

12 Vohlonen I, Syrjänen K and Karhu J: Serum biomarker-based screening of the risk groups of stomach cancer. Preliminary results of a feasibility study in Finland. Turk J Acad Gastroenterol *12*: 47-51, 2013.

13 Väänänen H, Vauhkonen M, Helske T, Kääriäinen I, Rasmussen M, Tunturi-Hihnala H, Koskenpato J, Sotka M, Turunen M, Sandström R, Ristikankare M, Jussila A and Sipponen P: 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 *15*: 885-891, 2003.

14 Samloff IM, Varis K, Ihamaki T, Siurala M and Rotter JI: Relationships among serum pepsinogen I, serum pepsinogen II, and gastric mucosal histology. A study in relatives of patients with pernicious anemia. Gastroenterol *83*: 204-209, 1982.

15 Miki K: Gastric cancer screening using the serum pepsinogen test method. Gastric Cancer *9*: 245-253, 2006.

16 Agréus L, Kuipers EJ, Kupcinskas L, Malfertheiner P, Di Mario F, Leja M, Mahachai V, Yaron N, van Oijen M, Perez Perez G,

Rugge M, Ronkainen J, Salaspuro M, Sipponen P, Sugano K and Sung J: Rationale in diagnosis and screening of atrophic gastritis with stomach-specific plasma biomarkers. Scand J Gastroenterol *47*: 136-147, 2012.

17 Lomba-Viana R, Dinis-Ribeiro M, Fonseca F, Vieira AS, Bento MJ and Lomba-Viana H: Serum pepsinogen test for early detection of gastric cancer in a European country. Eur J Gastroenterol Hepatol *24*: 37-41, 2012.

18 Germaná B, Di Mario F, Cavallaro LG, Moussa AM, Lecis P, Liatoupolou S, Comparato G, Carloni C, Bertiato G, Battiestel M, Papa N, Aragona G, Cavestro GM, Iori V, Merli R, Bertolini S, Caruana P and Franzé A: Clinical usefulness of serum pepsinogens I and II, gastrin-17 and anti-Helicobacterpylori antibodies in the management of dyspeptic patients in primary care. Dig Liver Dis *37*: 501-508, 2005.

19 Dinis-Ribeiro M, Yamaki G, Miki K, Costa-Pereira A, Matsukawa M and Kurihara M: Meta-analysis on the validity of pepsinogen test for gastric carcinoma, dysplasia or chronic atrophic gastritis screening. J Med Screen *11*: 141-147, 2004.

20 Oksanen A, Sipponen P, Miettinen A, Sarna S and Rautelin H: Evaluation of bood tests to normal gastric mucosa. Scand J Gastroenterol *35*: 791-795, 2000.

21 Varis K, Sipponen P, Laxen F, Samloff IM, Huttunen JK, Taylor PR, Heinonen OP, Albanes D, Sande N, Virtamo J, Härkönen M and The Helsinki Gastritis Study Group: Implications of serum pepsinogen I in early endoscopic diagnosis of gastric cancer and dysplasia. Scand J Gastroenterol *35*: 950-956, 2000.

22 Syrjänen KJ, Sipponen P, Härkönen M, Peetsalu A and Korpela S: Accuracy of GastroPanel testing in detection of atrophic gastritis. Eur J Gastroenterol Hepatol *27*: 102-104, 2015.

23 Cavallaro LG, Moussa AM, Liatopoulou S, Comparato G, Bertolini S, Merli R, Dal Bo N, Maino M, Bosi S, Cavestro GM, Prati B, Nervi G, Rugge M, Franzé A and DiMario F: Accuracy of "serological gastric biopsy" in a cohort of dyspeptic patients. 6th European Bridging Meeting in Gastroenterology, Kaunas, Lithuania. A5: p. 32, 2004.

24 De Korwin JD, Prucha V, Lozniewski A, Vignaud JM, Conroy MC and Legoedec J: Non-endoscopic diagnosis of atrophic gastritis using the Gastro Panel/GastroSoft. Helicobacter *9*: 561-562. (A10.15), 2004.

25 McNicholl AG, Forné M, Barrio J, de la Coba C, González B and Rivera R: *Helicobacter pylori* Study Group of the Asociación Española de Gastroenterología (AEG). Accuracy of GastroPanel for the diagnosis of atrophic gastritis. Eur J Gastroenterol Hepatol *26*: 941-948, 2014.

26 Iijima K, Abe Y, Kikuchi R, Koike T, Ohara T, Sipponen P and Shimosegawa T: Serum biomarker tests are useful in delineating between patients with gastric atrophy and normal, healthy stomach. World J Gastroenterol *15*: 853-859, 2009.

27 Peitz U, Wex T, Vieth M, Stolte M, Willich S, Labenz J, Jaspersen D, Lind T and Malfertheiner P: Correlation of serum pepsinogens and gastrin-17 with atrophic gastritis in gastroesophageal reflux patients: A matched-pairs study. J Gastroenterol Hepatol *26*: 82-89, 2011.

28 Lombardo L, Leto R, Molinaro G, Migliardi M, Ravarino N, Rocca R and Torchio B: Prevalence of atrophic gastritis in dyspeptic patients in Piedmont. A survey using the GastroPanel test. Clin Chem Lab Med *48*: 1327-1332, 2010.

29 Nardone G, Rocco A, Staibano S, Mezza E, Autiero G, Compare D, De Rosa G and Budillon G: Diagnostic accuracy of the serum profile of gastric mucosa in relation to histological and morphometric diagnosis of atrophy. Aliment Pharmacol Ther *22*: 1139-1146, 2005.

30 Kurilovich SA, Belkovets AV, Reshetnikov OV, Openko TG, Malyutina SK, Ragino YI, Scherbakova LV, Leja M, Paloheimo L, Syrjänen K and Voevoda MI: Stomach-specific biomarkers (GastroPanel) can predict the development of gastric cancer in Caucasian population: A longitudinal nested case-control study in Siberia. Anticancer Res *36*: 247-254, 2016.

31 Storskrubb T, Aro P, Ronkainen J, Sipponen P, Nyhlin H and Talley NJ: Serum biomarkers provide an accurate method for diagnosis of atrophic gastritis in a general population: the Kalixanda study. Scand J Gastroenterol *43*: 1448-1455, 2008.

32 Telaranta-Keerie A, Kara R, Paloheimo L, Härkönen M and Sipponen P: Prevalence of undiagnosed advanced atrophic corpus gastritis in Finland: an observational study among 4,256 volunteers without specific complaints. Scand J Gastroenterol *45*: 1036-1041, 2010.

33 Zagari RM, Nicolini G, Casanova S, Fiocca R, Pozzato P, Lunedei V, DiMarco M, Luca D, Ricciardiello L, Martuzzi C, Fuccio L, Maltoni S, Grigioni W, Panuccio D and Sipponen P: Diagnosis of atrophic gastritis in the general population based on a combination of three non-invasive tests. Gut *51(suppl 11)*: A39, 2002.

34 Belkovets AV, Kurilovich SA, Reshetnikov O, Ragino YI and Scherbakova LV: Non-invasive diagnosis of the phenotype of gastritis: analysis of the first thousand cases. Eksp Klin Gastroenterol *3*: 26-30, 2015.

35 Dinis-Ribeiro M, Areia M, de Vries AC, Marcos-Pinto R, Monteiro-Soares M, O'Connor A, Pereira C, Pimentel-Nunes P, Correia R, A. Ensari A, Dumonceau M, Machado JC, Macedo G, Malfertheiner P, Matysiak-Budnik T, Megraud F, Miki K, O'Morain C, Peek RM. Ponchon T, Ristimäki A, Rembacken B, Carneiro F, Kuipers EJ and behalf of MAPS Participants: 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 *44*: 74-94, 2012.

36 Dixon MF, Genta RM, Yardley JH and Correa P: Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. Am J Surg Pathol *20*: 1161-1181, 1996.

37 Sipponen P and Stolte M: Clinical impact of routine biopsies of the gastric antrum and body. Endoscopy *29*: 671-678, 1997.

38 Moher D, Liberati A, Tetzlaff J, Altman DG. PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ *339*: b2535, 2009.

39 Hardy RJ and Thompson SG: Detecting and describing heterogeneity in meta-analysis. Stat Med *17*: 841-856, 1998.

40 Sterne JA, Gavaghan D and Egger M: Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol *53*: 1119-1129, 2000.

41 Begg CB and Mazumdar M: Operating characteristics of a rank correlation test for publication bias. Biometrics *50*: 1088-1101, 1994.

42 Egger M, Davey SG, Schneider M and Minder C: Bias in meta-analysis detected by a simple, graphical test. BMJ *315*: 629-634, 1997.

43 Duval S and Tweedie R: Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics *56*: 455-463, 2000.

44 Thompson SG and Sharp SJ: Explaining heterogeneity in meta-analysis: a comparison of methods. Stat Med *18*: 2693-2708, 1999.
45 Di Mario F, Moussa AM, Caruana P, Merli R, Cavallaro LG, Cavestro GM, Dal Bo N, Iori V, Pilotto A, Leandro G, Franzé A and Rugge M: Serological biopsy in first-degree relatives of patients with gastric cancer affected by *Helicobacter pylori* infection. Scand J Gastroenterol *38*: 1223-1227, 2003.
46 Hartleb M, Wandzel P, Waluga M, Matyszczyk B, Bołdys H and Romanczyk T: Non-endoscopic diagnosis of multifocal atrophic gastritis; efficacy of serum gastrin-17, pepsinogens and *H. pylori* antibodies. Acta Gastroenterol Belg *67*: 320-326, 2004.
47 Pasechnikov VD, Chukov SZ, Kotelevets SM, Mostovov AN, Mernova VP and Polyakova MB: Possibility of non-invasive diagnosis of gastric mucosal precancerous changes. World J Gastroenterol *10*: 3146-3150, 2004.
48 Pasechnikov VD, Chukov SZ, Kotelevets SM, Mostovov AN, Mernova VP and Polyakova MB: Invasive and non-invasive diagnosis of *Helicobacter pylori*-associated atrophic gastritis: a comparative study. Scand J Gastroenterol *40*: 297-301, 2005.
49 Graham DY, Nurgalieva ZZ, El-Zimaity HM, Opekun AR, Campos A and Guerrero L: Noninvasive versus histologic detection of gastric atrophy in a Hispanic population in North America. Clin Gastroenterol Hepatol *4*: 306-314, 2006.
50 Valle MJ, Varasa TA, Pardo RL, Merlo RR, Pérez Grueso MJ, Escobedo RM, Torres MA, Boyd RC and Carrobles Jiménez JM: Diagnóstico serológico de gastritis atrófica con una combinación de pepsinógeno I y II, gastrina-17 y anticuerpos anti-*Helicobacter pylori*. Gastroenterol Hepatol *30*: 567-571, 2007.
51 Reshetnikov OV, Kurilovich SA, Krotov SA, Krotova VA, Bessonov PP, Vasil'ev RR, Tatarinova OV and Muchina EG: Prevalence of atrophic gastritis in different populations in Siberia on medical evidence of the serological survey. Klin Med (Mosk) *86*: 35-38, 2008.
52 Guariso G, Basso D, Bortoluzzi C-F, Meneghel A, Schiavon S, Fogar P, Farina M, Navaglia F, Greco E, Mescoli C, Zambon C-F and Plebani M: GastroPanel: Evaluation of the usefulness in the diagnosis of gastro-duodenal mucosal alterations in children. Clinica Chimica Acta *402*: 54-60, 2009.
53 Mecklin J-P, Ristimäki A, Sipponen P, Nuorva K, Kärjä V, Sarna S, Renkonen-Sinisalo L, Aarnio M, Heikkinen M, Pylvänäinen K and Järvinen HJ: Serological markers in evaluation of individual risk of gastric cancer among mutation carriers in Lynch syndrome. The International Society for Gastrointestinal Hereditary Tumours Incorporated. Dusseldorf, Germany. June 24-27, 2009.
54 Masci E, Pellicano R, Mangiavillano B, Luigiano C, Stelitano L, Morace C, Viale E, Freschi M, Locatelli M, Ieri R, Cavallaro A, Testoni S and Testoni PA: GastroPanel test for non-invasive diagnosis of atrophic gastritis in patients with dyspepsia. Minerva Gastroenterol Dietol *60*: 79-83, 2014.
55 Nejadi-Kelarijani F, Roshandel G, Semnani S, Ahmadi A, Faghani B, Besharat S, Akhavan-Tabib A and Amiriani T: 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 Pac J Cancer Prev *15*: 7433-7436, 2014.
56 Ivashkin VT, Lapina TL, Purveeva KV, Sklanskaya OA, Konkov MU, Serova AG and Vasilieva NY: The serum levels of pepsinogen I, pepsinogen II, Gastrin-17 as the markers of atrophic gastritis in the dyspeptic patients. Submitted; 2016.
57 Goni E, Franceschi M, Scarpignato C, Rugge M and Di Mario F: Relationship between Gastro Panel and O.L.G.A. in the detection of chronic atrophic gastritis in a primary care setting. P-0514. 23rd United European Gastroenterology Week, Barcelona, Spain, October 24-28, 2015.
58 Gou Y, Wang W, Li Y, Chen H, Jin A, He W and Huang X: Diagnostic values of serum pepsinogens, Gastrin 17 and *Helicobacter pylori* IgG antibodies for atrophic gastritis or gastric cancer patients screening in a high risk area in China. P-1790. 23rd United European Gastroenterology Week, Barcelona, Spain, October 24-28, 2015.
59 Lee, Y-C, Chiang T-H, Chou C-K, Chiu H-M and Wu M-S: Serological biomarkers for gastric cancer screening in a population that has undergone mass eradication of *Helicobacter pylori*. Digestive Disease Week, San Diego, CA, May 21-24, 2016.
60 Roman LD, Lukyanchuk R, Sablin OA, Araslanova EI, Holmström C, Hendolin P, Paloheimo L and Syrjänen K: Prevalence of *H. pylori* infection and atrophic gastritis in a population-based screening with serum biomarker panel (GastroPanel®) in St. Petersburg. Anticancer Res *36*: 4129-4138, 2016.
61 Borenstein M, Hedges LV, Higgins JPT and Rothstein H: Introduction to Meta-Analysis. Chichester: Wiley, 2009.
62 Hartung J, Knapp G and Sinha BK: Statistical Meta-Analysis with Applications. Hoboken: Wiley, 2008.
63 Sutton AJ, Abrams KR, Jones DR and Song F: Methods for Meta-Analysis in Medical Research. Chichester: Wiley, 2000.
64 Huedo-Medina T, Sanchez-Meca J, Marin-Martinez F, and Botella J: Assessing heterogeneity in metaanalysis: Q statistic or I2 index? CHIP Documents. 2006; Paper 19. Available at: http://digitalcommons.uconn. edu/chip_docs/19. Accessed on March 22, 2016.
65 Field AP: Meta-analysis of correlation coefficients: a Monte Carlo comparison of fixed- and random-effects methods. Psychol Meth *6*: 161-180, 2001.
66 Field AP: The problems in using fixed-effects models of meta-analysis on real-world data. Understand Stat *2*: 77-96, 2003.
67 Cochran WG: The combination of estimates from different experiments. Biometrics *10*: 101-129, 1954.
68 Higgins JPT, Thompson SG, Deeks JJ and Altman DG: Measuring inconsistency in meta-analyses. BMJ *327*: 557-560, 2003.
69 Sipponen P, Ranta P, Helske T, Kääriäinen I, Mäki T and Linnala A: Serum levels of amidated gastrin-17 and pepsinogen I in atrophic gastritis: an observational case-control study. Scand J Gastroenterol *37*: 785-791, 2002.
70 Sipponen P, Valle J, Varis K, Kekki M, Ihamäki T and Surala M: Fasting levels of serum gastrin in different functional and morphologic states of the antro-fundal mucosa. An analysis of 860 subjects. Scand J Gastroenterol *25*: 513-519, 1990.
71 Leja M, Kupcinskas L, Funka K, Sudraba A, Jonaitis L and Ivanauskas A: Value of gastrin-17 in detecting ntral atrophy. Adv Med Sci *56*: 145-150, 2011.

Received August 18, 2016
Revised August 26, 2016
Accepted August 29, 2016