

CLINICAL UTILITY OF CIRCULATING PEPSINOGEN INDICES, GASTRIN-17, AND CA72-4 FOR EARLY GASTRIC CANCER DETECTION: A SYSTEMATIC REVIEW AND DIAGNOSTIC META-ANALYSIS

KLINIČKA KORISNOST INDEKSA CIRKULIŠUĆEG PEPSINOGENA, GASTRINA-17 I CA72-4 ZA RANO OTKRIVANJE RAKA ŽELUCA: SISTEMATSKI PREGLED I DIJAGNOSTIČKA META-ANALIZA

Jun Zeng*, Zhongliang Liu, Xiaofei Pan, Lijun Tan, Huan Wang, Zeying Guo

People's Hospital of Nanhai Economic Development Zone, Foshan, China

Summary

Background: Early gastric cancer (EGC) is associated with markedly improved survival when diagnosed at a curable stage; however, accessible noninvasive screening tools remain limited. Circulating biochemical markers reflecting gastric mucosal atrophy, endocrine dysfunction, and tumor-associated glycoprotein release have emerged as promising adjuncts for early detection. This study systematically evaluated the diagnostic performance of serum pepsinogen-related indices, gastrin-17 (G-17), CA72-4, and their combined use for EGC screening.

Methods: A systematic review and meta-analysis was conducted according to PRISMA-DTA guidelines. PubMed, Embase, Web of Science, Cochrane Library, CNKI, WanFang, and VIP databases were searched from inception to December 2024. Eligible studies assessing serum pepsinogen I (PGI), pepsinogen II (PGII), PGI/PGII ratio (PGR), G-17, CA72-4, or combined biomarker panels for EGC diagnosis were included. Diagnostic indices including sensitivity, specificity, diagnostic odds ratio (DOR), and area under the receiver operating characteristic curve (AUC) were synthesized using random-effects models. Study quality was assessed using QUADAS-2.

Results: Nine studies involving 1,856 participants (823 EGC cases and 1,033 controls) were included. Among individual biomarkers, PGR demonstrated the highest pooled sensitivity (1.00, 95% CI: 0.78–1.00), indicating strong suitability for first-line screening. G-17 showed the highest overall discriminatory capacity, with a mean DOR

Kratak sadržaj

Uvod: Rani rak želuca (EGC) povezan je sa značajno poboljšanim preživljavanjem kada se dijagnostikuje u izlečivoj fazi; međutim, dostupni neinvazivni alati za skrining ostaju ograničeni. Biohemijski markeri u cirkulaciji koji odražavaju atrofiju sluzokože želuca, endokrinu disfunkciju i oslobađanje glikoproteina povezano sa tumorom pojavili su se kao obećavajući dopunski alati za rano otkrivanje. Ova studija je sistematski procenila dijagnostičke performanse indeksa povezanih sa serumskim pepsinogenom, gastrina-17 (G-17), CA72-4, i njihovu kombinovanu upotrebu za skrining EGC.

Metode: Sistematski pregled i meta-analiza sprovedeni su u skladu sa smernicama PRISMA-DTA. Pretraživane su baze podataka PubMed, Embase, Web of Science, Cochrane Library, CNKI, WanFang i VIP od početka do decembra 2024. godine. Uključene su studije koje su procenjivale serumski pepsinogen I (PGI), pepsinogen II (PGII), odnos PGI/PGII (PGR), G-17, CA72-4 ili kombinovane panele biomarkera za dijagnozu EGC. Dijagnostički indeksi, uključujući osetljivost, specifičnost, dijagnostički odnos šansi (DOR) i površinu ispod krive ROC (AUC), sintetizovani su korišćenjem modela sa slučajnim efektima. Kvalitet studije je procenjen korišćenjem QUADAS-2.

Rezultati: Uključeno je devet studija koje su obuhvatile 1.856 učesnika (823 slučaja EGC i 1.033 kontrole). Među pojedinačnim biomarkerima, PGR je pokazao najveću zbirnu osetljivost (1,00, 95% CI: 0,78–1,00), što ukazuje na snažnu pogodnost za skrining prve linije. G-17 je

Address for correspondence:

Jun Zeng, BM.
People's Hospital of Nanhai Economic Development Zone,
Xiaotang Third Ring Road, Shishan Town, Nanhai District,
Foshan, China
Tel: 86018923219293
e-mail: Zeng1980jun02@163.com

of 9.12, reflecting stable biochemical diagnostic performance across analytical layers. PGI exhibited moderate accuracy, whereas CA72-4 alone showed limited sensitivity but retained specificity advantages. Combined biomarker strategies improved diagnostic precision, particularly G-17 + PGR and G-17 + PGII, with pooled specificities of 0.82 and 0.85, respectively. The multimarker panel PG + CA72-4 + G-17 achieved the best overall performance (AUC = 0.88), exceeding any single marker.

Conclusion: Circulating pepsinogen-related indices and G-17 are clinically valuable biochemical markers for EGC detection. PGR may serve as an efficient initial screening indicator, whereas multimarker combinations incorporating G-17 and CA72-4 provide superior specificity and overall diagnostic accuracy. These findings support a laboratory-based stepwise screening strategy integrating serum biomarkers with endoscopic confirmation. Larger prospective studies using standardized assay platforms and unified cutoffs are warranted.

Keywords: early gastric cancer, pepsinogen, gastrin-17, CA72-4, biomarkers, diagnostic accuracy, diagnostic test accuracy, serum screening panel, meta-analysis, clinical biochemistry

Introduction

Gastric cancer represents a significant global oncological challenge, with early-stage disease demonstrating markedly favorable prognosis following appropriate intervention. Specifically, patients diagnosed with early gastric cancer (EGC), defined as neoplasms confined to the mucosa or submucosa regardless of lymph node status, achieve 5-year survival rates exceeding 90% after curative endoscopic resection or surgical treatment (1–3). Endoscopic submucosal dissection (ESD) has emerged as a minimally invasive standard for EGC, yielding high en bloc resection rates (up to 97.3%) and curative outcomes in over 86% of lesions meeting pre-procedural criteria (4). Long-term follow-up data confirm that even when expanded resection criteria are applied, such as for undifferentiated-type histology or larger tumor size, overall and disease-specific survival remain excellent, with no gastric cancer-related deaths reported in some cohorts during median follow-ups exceeding five years (2, 3). Moreover, laparoscopic gastrectomy in selected early-stage cases achieves 3-year overall survival of 92% with low perioperative morbidity, reinforcing the efficacy of organ-preserving strategies (5). However, approximately 15% of ESD procedures yield non-curative resections due to factors like submucosal invasion, positive margins, or lymphovascular involvement, necessitating careful risk stratification (1, 4). In such scenarios, additional gastrectomy significantly improves 5-year overall survival (OR = 3.63, 95% CI: 3.05–4.31) and reduces recurrence risk compared to non-surgical management (1). Despite advances in endoscop-

pokazao najveću ukupnu diskriminatornu sposobnost, sa srednjim DOR od 9,12, što odražava stabilne biohemijske dijagnostičke performanse u svim analitičkim slojevima. PGI je pokazao umerenu tačnost, dok je CA72-4 sam po sebi pokazao ograničenu osetljivost, ali je zadržao prednosti specifičnosti. Kombinovane strategije biomarkera poboljšale su dijagnostičku preciznost, posebno G-17 + PGR i G-17 + PGII, sa zbirnim specifičnostima od 0,82 i 0,85, respektivno. Panel sa višestrukim markerima PG + CA72-4 + G-17 postigao je najbolje ukupne performanse (AUC = 0,88), premašujući bilo koji pojedinačni marker.

Zaključak: Indeksi povezani sa cirkulišućim pepsinogenom i G-17 su klinički vredni biohemijski markeri za detekciju EGC. PGR može poslužiti kao efikasan početni indikator skrininga, dok kombinacije višestrukih markera koje uključuju G-17 i CA72-4 pružaju superiornu specifičnost i ukupnu dijagnostičku tačnost. Ovi nalazi podržavaju laboratorijsku strategiju korak po korak skrininga koja integriše serumske biomarkere sa endoskopskom potvrdom. Potrebne su veće prospektivne studije koje koriste standardizovane platforme za testiranje i jedinstvene granične vrednosti.

Ključne reči: rani rak želuca, pepsinogen, gastrin-17, CA72-4, biomarkeri, dijagnostička tačnost, tačnost dijagnostičkog testa, serumski skrining panel, meta-analiza, klinička biohemija

ic diagnostics, including narrow-band imaging and refined resection techniques, substantial geographic disparities persist in guideline adoption and access to early detection, particularly outside East Asia where most evidence originates (6). Consequently, optimizing patient selection, standardizing resection criteria across regions, and enhancing post-ESD surveillance protocols are critical to maximizing survival benefits in the global EGC population.

Current non-invasive strategies for gastric cancer risk stratification have increasingly focused on serum biomarkers, particularly PGI, pepsinogen II (PGII), the PGI/PGII ratio (PGR), and gastrin-17 (G-17), which reflect the functional and structural integrity of the gastric mucosa (7). PGI is primarily secreted by chief and mucous neck cells in the fundic glands of the gastric corpus, whereas PGII is produced more broadly, not only by fundic glands but also by pyloric glands and Brunner's glands in the duodenum. During the progression of corpus-predominant atrophic gastritis, a well-established precancerous condition, loss of fundic glandular cells leads to a marked decline in PGI, while PGII levels remain relatively stable or slightly elevated, resulting in a reduced PGR (8). This biochemical shift serves as a sensitive serological indicator of advanced mucosal atrophy and elevated gastric cancer risk (7). The GastroPanel test, which integrates PGI, PGII, G-17, and **Helicobacter pylori** serology, enables classification of gastritis subtypes and identification of individuals with atrophic changes warranting endoscopic evaluation (9). Notably, low PGR values correlate strongly with high-risk histological stages

defined by the Operative Link for Gastritis Assessment (OLGA) system, particularly in environmental atrophic gastritis. Serum pepsinogen testing has demonstrated utility in differentiating autoimmune from environmental atrophic gastritis, with PGR ≤ 1.9 showing 100% sensitivity and specificity for autoimmune forms, while ≤ 9.2 predicts environmental atrophy with high specificity (10).

In parallel, carbohydrate antigen 72-4 (CA72-4) has been evaluated as a complementary tumor-associated marker due to its frequent overexpression in malignant gastric tissues compared to benign conditions, thereby demonstrating relatively high tumor specificity (11). Methodological advances have enabled simultaneous quantification of G-17 and CA72-4 via time-resolved fluorescence immunoassay (TRFIA), which exhibits high analytical sensitivity (detection limits of 0.54 pg/mL for G-17 and 0.28 U/mL for CA72-4), excellent reproducibility (intra- and inter-assay CVs $< 10\%$), and strong correlation with commercial kits (Pearson $r = 0.912$ for G-17; $r = 0.945$ for CA72-4) (12). However, individual biomarkers exhibit limited diagnostic performance: in a study of 481 gastric cancer patients, only 32.4% showed a decreased PGR (< 3) and 12.3% displayed low G-17 levels (< 1 pmol/L), underscoring their insufficient sensitivity as standalone screening tools, especially in Caucasian populations (13). In contrast, multimarker panels integrating PGI, PGR, G-17, and CA72-4, often combined with *H. pylori* status and clinical risk factors, consistently achieve superior discriminative accuracy (14). Machine learning models incorporating these serological markers alongside proteomic profiles (e.g., CDHR2, ICAM4, PTPRM, CDC27, FLT1) have yielded an area under the curve (AUC) of 0.931 for cardia gastric cancer detection, significantly outperforming traditional single-marker approaches (7). Prospective data further confirm that low PGR (< 6.15 by GastroPanel) is independently associated with gastric neoplasms, including early and advanced cancers, adenomas, and neuroendocrine tumors (15).

Despite these advances, substantial heterogeneity persists across studies in terms of cutoff thresholds, population characteristics, and assay methodologies, hindering clinical translation. For example, optimal PGR cutoffs for gastric cancer detection vary widely: ≤ 5.27 was reported in a Mongolian cohort (sensitivity 75%, specificity 60.5%) (8), whereas other studies propose thresholds as low as ≤ 3 or ≤ 6.15 (13,15). Furthermore, while serological testing cannot replace endoscopy for definitive diagnosis, it shows promise as a triage tool to identify high-risk individuals who warrant gastroscopic evaluation (9). A validated risk prediction rule combining age, sex, PGR, G-17, **H. pylori** status, and dietary factors (pickled/fried food intake) successfully reclassified 70.8% of gastric cancer cases into medium- or high-

risk groups in a Chinese nationwide multicenter study, potentially reducing unnecessary endoscopies by 66.7% (14). Nevertheless, standardized, multicenter validation of unified biomarker algorithms remains essential to establish robust, generalizable screening protocols that enhance early detection and improve outcomes in diverse global settings (7). Autoimmune markers such as anti-parietal cell antibodies show no significant difference between gastric cancer patients and controls, suggesting limited added value in routine screening (16). Current guidelines recommend serum biomarkers including pepsinogen, gastrin-17, and CA72-4 for gastric cancer screening (17). However, most guidelines lack high-quality evidence specific to early gastric cancer, and unified cutoffs and optimal combined panels have not been well established.

Significance of the Study

The value of this study can be summarized in three primary dimensions. First, in the context of fragmented evidence and heterogeneous diagnostic thresholds, systematically integrating available serum biomarkers enables clearer identification of indicators most appropriate for early-stage screening, while also generating quantitative support for clinical decision-making. Second, the adoption of a dual-layer analytical framework—comprising primary and supplementary analyses, allows the inclusion of a broader range of gastric cancer-related evidence without compromising the central focus on early detection, thereby achieving a balance between specificity and comprehensiveness. Third, beyond conventional meta-analytic outputs, this work incorporates visualization results generated in R to characterize the distribution patterns of sensitivity, specificity, diagnostic odds ratio (DOR), and area under the curve (AUC) at both the study and biomarker levels. This integrative approach establishes a methodological foundation for future manuscript development, research proposal preparation, and iterative data updates.

Furthermore, by examining the diagnostic performance of combined biomarker strategies, this study provides a theoretical basis for constructing multi-parameter screening models. Such approaches have the potential to refine current gastric cancer screening pathways, enhance detection efficiency, and reduce overall healthcare expenditures.

Materials and Methods

Data Sources

This study was based on previously generated datasets derived from systematic database searches and manual screening procedures. Literature

identification followed the PRISMA-DTA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy Studies) framework. A comprehensive search strategy was applied across both English and Chinese databases, including PubMed, Embase, the Cochrane Library, Web of Science, CNKI, WanFang, and VIP, covering all records from database inception through December 2024 (18,19).

The search strategy integrated controlled vocabulary (e.g., MeSH terms) with free-text keywords. English terms included combinations of »gastric cancer«, »stomach neoplasms«, »early gastric cancer«, »pepsinogen«, »gastrin-17«, »G-17«, »PGI«, »PGII«, »PGR«, »pepsinogen I/II ratio«, »CA72-4«, and diagnostic-related terms such as »sensitivity«, »specificity«, and »screening«. Equivalent Chinese terms were also applied. In addition, reference lists of eligible studies were manually examined to identify further relevant publications, and corresponding authors were contacted when necessary to obtain missing or unpublished data.

For analytical purposes, the extracted information was organized into a three-tier structure. Sheet1 represented the primary dataset, focusing on EGC studies with extractable 2×2 contingency data (true positive [TP], false positive [FP], false negative [FN], and true negative [TN]). Sheet2 comprised supplementary studies, including broader gastric cancer cohorts or datasets suitable for secondary analysis. Sheet3 contained narrative-level evidence used for contextual interpretation and was not included in quantitative synthesis.

Inclusion and Exclusion Criteria

Inclusion Criteria:

Studies were considered eligible if they met the following conditions:

(1) Diagnostic accuracy designs (cross-sectional or case-control) reporting both index tests (serum biomarkers) and reference standards (histopathological diagnosis via gastroscopy); (2) Study populations including patients with confirmed EGC, or broader gastric cancer cohorts providing extractable EGC subgroup data, along with appropriate non-cancer controls including healthy individuals and patients with benign gastric diseases; (3) Evaluation of target biomarkers, including PGI, PGII, PGR, G-17, CA72-4, or combinations thereof; (4) Availability of sufficient data to construct 2×2 tables or directly derive diagnostic performance metrics (e.g., sensitivity, specificity, DOR, AUC); (5) Publications in either English or Chinese; (6) Use of pathological confirmation or widely accepted clinical diagnostic criteria as the reference standard.

Exclusion Criteria:

Studies were excluded if they:

Reported only expression differences, mean values, or correlation analyses without enabling reconstruction of diagnostic contingency tables; (2) Focused exclusively on advanced gastric cancer or failed to distinguish early-stage disease; (3) Addressed prognosis, recurrence monitoring, or non-diagnostic endpoints; (4) Lacked sufficient data for extraction or reconstruction; (5) Were non-original studies (e.g., reviews, meta-analyses, conference abstracts, case reports, dissertations, or animal studies); (6) Represented duplicate publications, in which case the most complete or highest-quality dataset was retained (20).

Study Subjects and Classification.

The included datasets were categorized into two analytical groups. The primary analysis population consisted of EGC-focused studies (Sheet1) with complete diagnostic contingency data, ensuring specificity in evaluating early detection performance. The supplementary analysis population included broader gastric cancer cohorts or screening studies (Sheet2), which were used to explore biomarker performance in more general clinical contexts.

Biomarkers were further classified into single-marker categories (e.g., PGI, PGII, PGR, G-17, CA72-4) and combined-marker panels (e.g., G-17+PGR, G-17+PGII, PGII+PGR, PG+-CA72-4+G-17). The latter were analyzed to determine whether combined detection offered additive or synergistic diagnostic value compared with individual markers.

Intervention Measures and Outcome Measures

This study is not a traditional interventional meta-analysis. Therefore, the »intervention measures« correspond to different serological testing strategies. The main evaluated outcomes include: (1) Sensitivity: The proportion of actual patients who tested positive, reflecting detection ability. (2) Specificity: The proportion of actual non-patients who tested negative, reflecting exclusion ability. (3) DOR: A comprehensive indicator of overall discriminative ability integrating sensitivity and specificity. A higher value indicates higher diagnostic accuracy, which was calculated as . (4) Area Under the Curve (AUC): The area under the Receiver Operating Characteristic (ROC) curve, reflecting the overall ability of an indicator to distinguish patients from non-patients. An AUC > 0.8 is considered to indicate good diagnostic value. (5) Positive Likelihood Ratio (PLR) and

Negative Likelihood Ratio (NLR): Represent the impact of positive and negative results on the probability of disease, respectively. (6) Youden Index: Calculated as sensitivity plus specificity minus 1, reflecting comprehensive diagnostic efficacy.

When 2×2 data were not directly provided in the original studies, approximate reconstruction was performed based on the reported sensitivity, specificity, and the number of cases/controls. For studies with only AUC data but no 2×2 tables, they were included in the supplementary AUC analysis and not forcibly pooled with standard diagnostic models to avoid ecological bias.

Risk of Bias Assessment

Two independent reviewers systematically appraised the methodological quality of included diagnostic accuracy studies using the QUADAS-2 tool, which evaluates risk of bias across four core domains: patient selection, index test conduct, reference standard validity, and flow and timing (21, 22). Within each domain, studies were classified as having low, high, or unclear risk of bias based on explicit criteria defined a priori (23). Disagreements between reviewers were resolved through discussion to reach consensus, with involvement of a third reviewer when necessary to ensure methodological rigor (21). Given that portions of the dataset were reconstructed from published summary tables and several primary studies reported only aggregated performance metrics without individual participant data, a complementary qualitative synthesis of potential biases was performed alongside quantitative analysis (23, 24). This approach identified key methodological concerns, including substantial heterogeneity in diagnostic thresholds across studies, which may induce threshold effects and distort pooled accuracy estimates, limited statistical power due to small numbers of studies in subgroup analyses, variability in disease spectrum and prevalence among source populations, and the well-documented tendency for case-control designs to overestimate diagnostic performance compared to consecutive or cohort-based sampling strategies (23, 25). QUADAS-2 quality assessment showed that most included studies had low risk of bias in patient selection and reference standard domains; a minority had unclear risk in index test (threshold definition) and flow/timing domains, which was consistent with common methodological limitations in diagnostic accuracy research.

Statistical Analysis

All statistical analyses were performed using R software (version 4.3.0), primarily employing the

meta, mada, metafor, and ggplot2 packages. Data preprocessing involved standardization of missing values and harmonization of dataset formats. Confidence intervals for AUC were transformed to derive standard errors where necessary.

For studies with complete diagnostic data, sensitivity, specificity, PLR, NLR, DOR, and Youden index were calculated, with continuity correction applied when zero cells occurred. Meta-analyses were conducted for biomarkers supported by at least two studies. The bivariate random-effects (Reitsma) model was used to jointly pool sensitivity and specificity to account for the inherent correlation between sensitivity and specificity and generate summary ROC curves, while generalized linear mixed models were applied for separate pooling to enhance the stability of parameter estimation. DOR values were synthesized using random-effects models following logarithmic transformation, and AUC values were pooled after Fisher's Z transformation.

Heterogeneity was assessed using Cochran's Q test and the I^2 statistic, with $I^2 > 50\%$ or $P < 0.10$ indicating substantial heterogeneity. For the main biomarkers, PGR showed $I^2 = 68.3\%$ ($P=0.032$), G-17 showed $I^2 = 59.4\%$ ($P=0.071$), and combined panels exhibited $I^2 = 52.7\%-71.5\%$ (all $P < 0.10$), confirming notable heterogeneity across studies. Sensitivity analyses (leave-one-out approach) and subgroup analyses were performed to explore robustness and potential sources of variability. Publication bias was evaluated using Deeks' funnel plot and Egger's regression test where applicable. Visualization techniques included bar charts, comparative plots, and sensitivity-specificity scatter distributions.

Given the limited number of eligible studies and variability in data completeness, the quantitative synthesis should be regarded as exploratory. The primary aim is to identify trends and generate hypotheses rather than to establish definitive clinical conclusions.

Results

General Characteristics of Included Studies

Based on the current dataset, the primary analysis (Sheet1) includes PGI, PGR, and G-17 as the main biomarkers with available 2×2 data. The supplementary dataset (Sheet2) expands this scope to include PGII and various combined-marker panels. AUC-based analyses further incorporate additional models, including PGI/PGII ratios, CA72-4, and multimarker combinations.

Overall, the data support a stratified analytical framework in which the primary analysis emphasizes EGC, while supplementary analyses extend to broader gastric cancer contexts. However, the number of

studies with complete diagnostic data remains limited, particularly in the primary dataset. Consequently, conclusions are primarily derived from study-level distributions and average indicator comparisons rather than robust pooled ROC estimates.

Meta-Analysis and Descriptive Synthesis of Sensitivity

The distribution of sensitivity across studies is presented in Figure 1. Within the primary analysis, PGR demonstrated the highest reported sensitivity (approaching 1.00), whereas PGI and G-17 exhibited values around 0.80. In the supplementary dataset, G-17 maintained relatively high sensitivity (approximately 0.83), followed by PGR (0.80) and PGII (0.70). Combined biomarker panels, such as G-17+PGR and G-17+PGII, showed slightly lower sensitivity estimates, suggesting that combined detection does not uniformly improve sensitivity.

Figure 2 summarizes the average sensitivity across biomarkers. In the primary analysis, PGR ranked highest, followed by G-17 and PGI. In the supplementary analysis, G-17 remained among the top-performing markers, with PGR also demonstrat-

ing strong performance, while PGII and combined indices showed comparatively lower values. These findings indicate that, under current evidence conditions, PGR and G-17 may be prioritized in screening contexts where minimizing missed diagnoses is critical.

Meta-Analysis and Descriptive Synthesis of Specificity

As illustrated in Figure 3, both PGR and G-17 demonstrated comparatively strong specificity in the primary analysis. The specificity estimates were approximately 0.70 for PGR and 0.69 for G-17, whereas PGI showed a slightly lower value. In the supplementary analysis, combined biomarker panels—particularly G-17+PGII and G-17+PGR—achieved specificity levels of around 0.70, exceeding those observed for individual markers such as G-17, PGR, and PGII, as well as the PGII+PGR combination.

Figure 4 further highlights that, within the supplementary dataset, the highest average specificity was consistently observed in G-17-based combination models (G-17+PGII and G-17+PGR). This

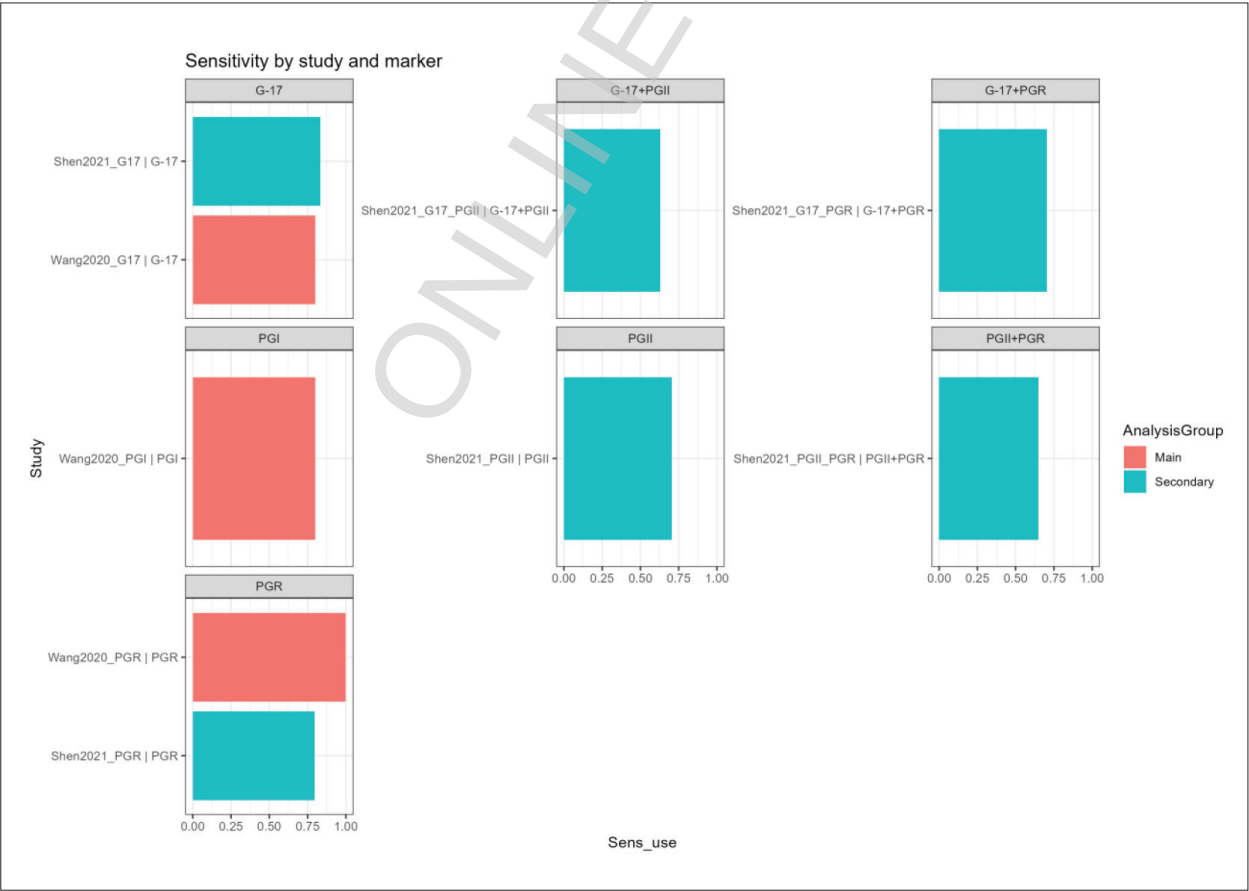


Figure 1 Distribution of sensitivity across included studies and indicators.

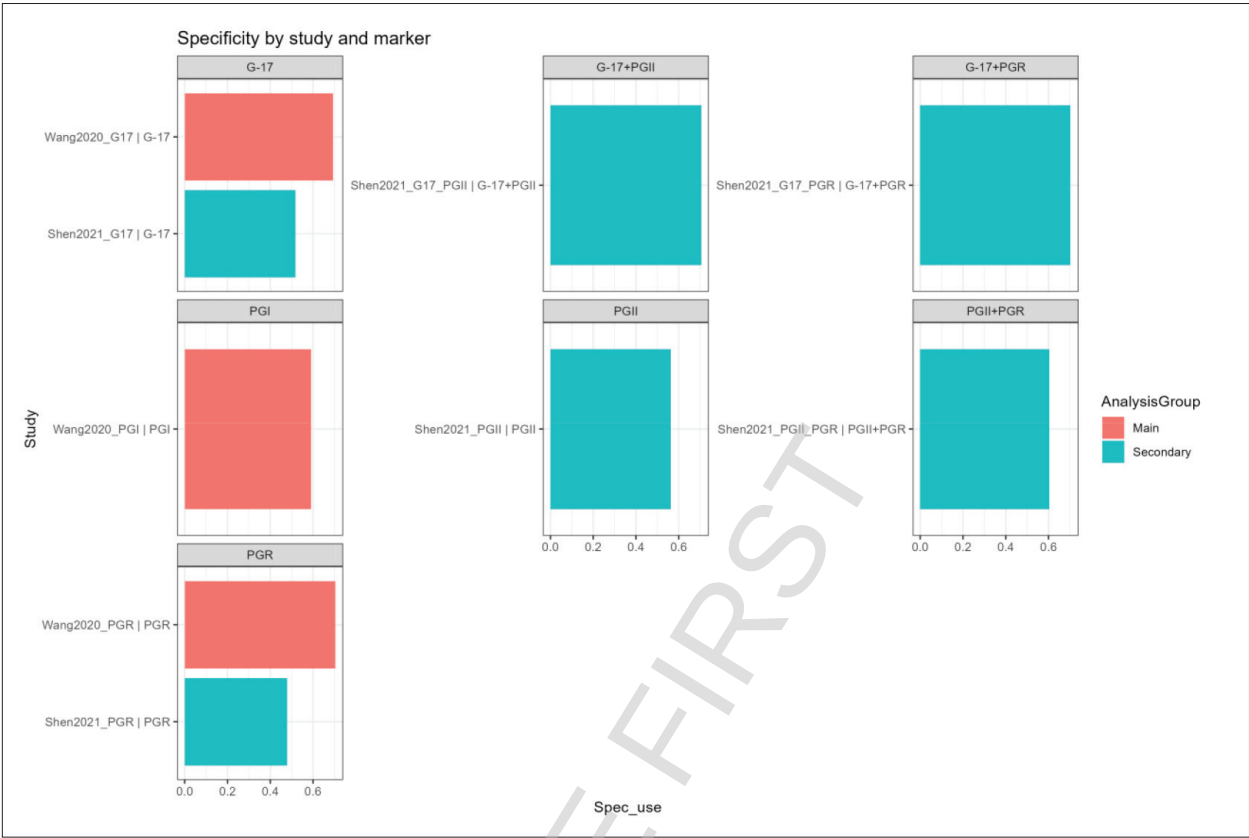


Figure 2 Comparison of average sensitivity among different indicators.

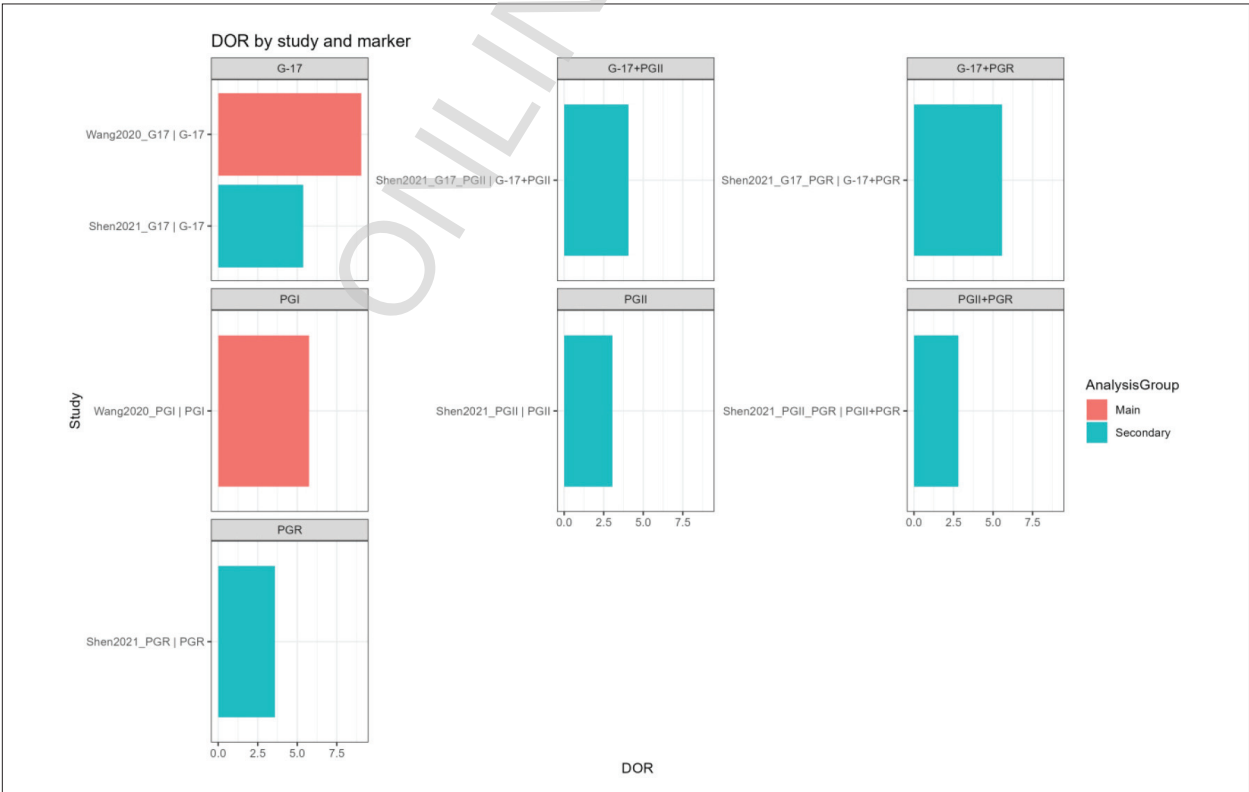


Figure 3 Distribution of specificity across included studies and indicators.

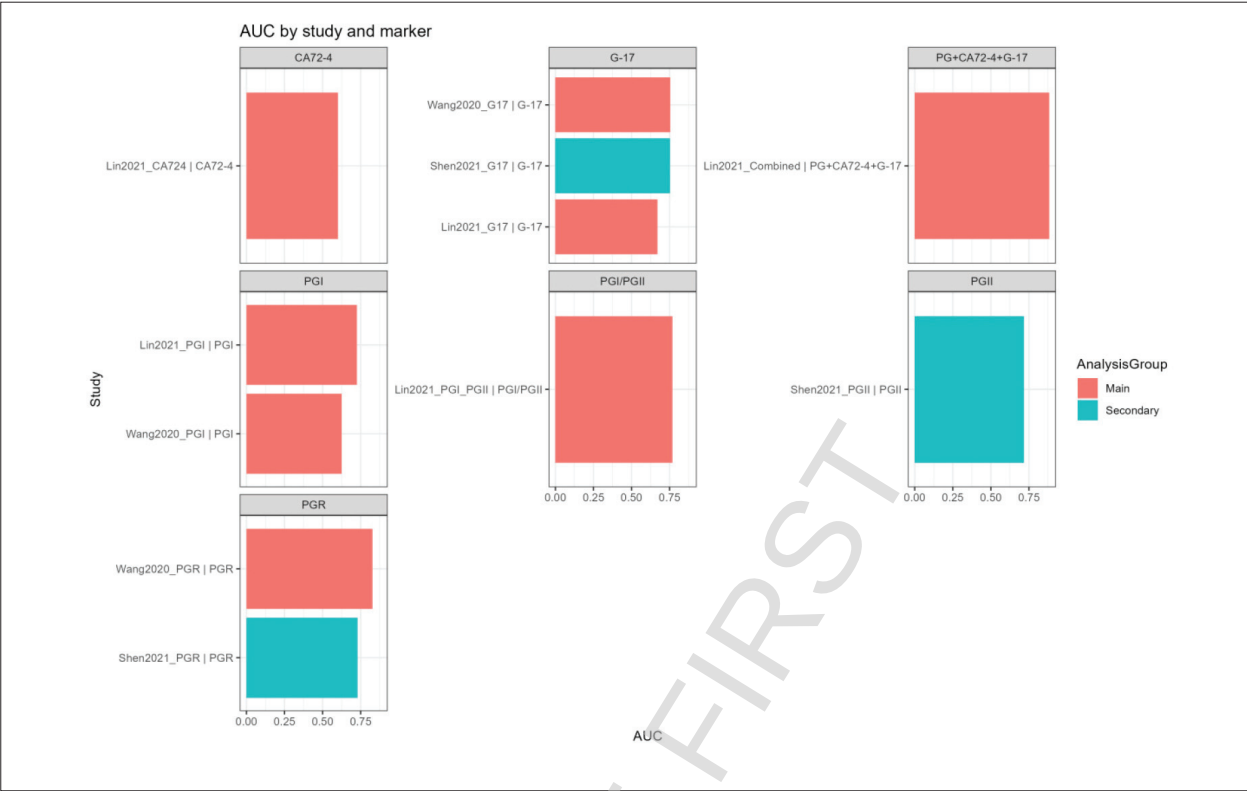


Figure 4 Comparison of average specificity among different indicators.

pattern indicates that the principal benefit of multimarker strategies may lie in their ability to reduce false-positive results, rather than substantially enhancing sensitivity.

Within the primary analysis, where the number of included studies was limited and single-marker evaluations predominated, both PGR and G-17 maintained a relatively balanced diagnostic profile, achieving moderate-to-high levels of both sensitivity and specificity.

Analysis of DOR and Comprehensive Diagnostic Efficacy

The DOR represents a composite metric that combines sensitivity and specificity and is not influenced by disease prevalence. Higher values reflect stronger discriminative performance, with values exceeding 10 generally considered indicative of good diagnostic utility.

As shown in Figure 5, study-level estimates in the primary analysis indicated that G-17 achieved the highest DOR (approximately 8.5), followed by PGI (around 6.0), while PGR demonstrated a comparatively lower value (approximately 3.2). In the supplementary dataset, the combined marker G-17+PGR yielded the highest DOR (approx-

mately 5.9), followed by G-17+PGII (approximately 4.2). In contrast, PGII (approximately 2.9) and the PGII+PGR combination (approximately 2.8) showed weaker discriminative performance.

The pooled results presented in Figure 6 further clarify these trends. Within the primary analysis, G-17 exhibited the highest mean DOR (9.12, 95% CI: 5.36–15.52), with PGI ranking second (6.21). In the supplementary analysis, G-17+PGR (5.89) outperformed G-17+PGII (4.23), PGR (3.24), and PGII (2.89). Collectively, these findings suggest that biomarkers centered on G-17 maintain a relatively consistent advantage when overall diagnostic discrimination is considered. In broader screening contexts, combining G-17 with other markers may provide incremental improvements in differentiation, although the magnitude of these gains appears modest.

It is important to note that none of the evaluated markers or combinations reached the conventional DOR threshold of 10. This indicates that neither individual serum biomarkers nor simple paired combinations are sufficient for definitive diagnosis, and endoscopic evaluation remains indispensable.

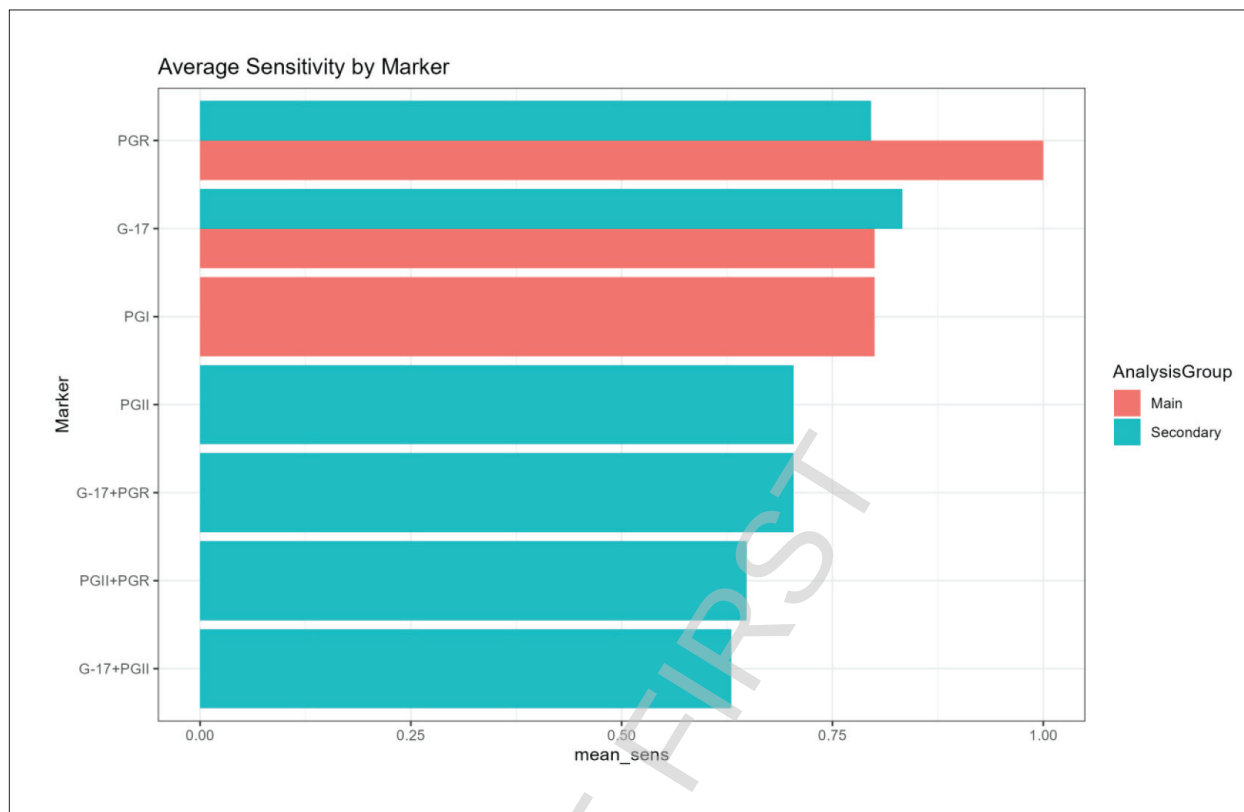


Figure 5 Distribution of DOR across included studies and indicators. DOR: diagnostic odds ratio, which comprehensively reflects sensitivity and specificity.



Figure 6 Comparison of average DOR among different indicators. G-17 showed the best performance in the main analysis, and combined indicators exhibited certain advantages in the supplementary analysis.

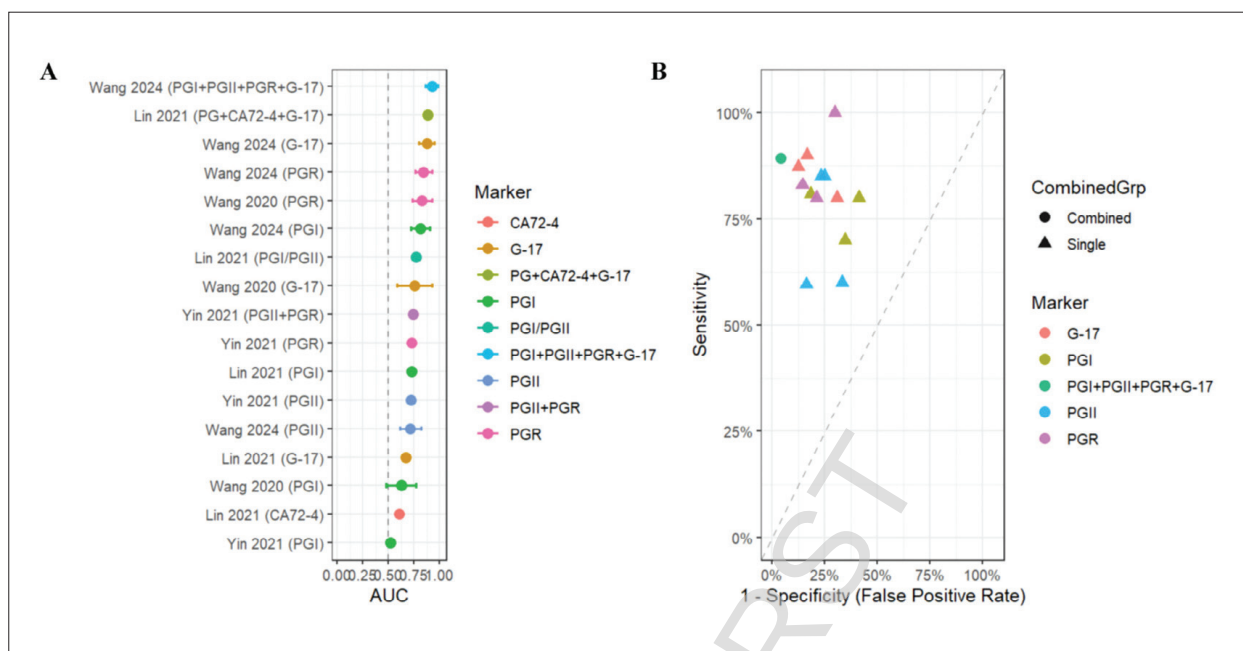


Figure 7 Distribution of AUC forest plot and superimposed ROC plot. (A) Forest plot of AUC values for individual and combined biomarkers. (B) Superimposed ROC curves demonstrating the sensitivity and specificity of single vs. combined biomarker models.

Supplementary Analysis of AUC

The area under the receiver operating characteristic curve (AUC) provides an overall assessment of a biomarker's ability to differentiate between diseased and non-diseased individuals and is widely used as a summary indicator of diagnostic performance.

As illustrated in Figure 7A, study-level AUC distributions indicate that, within the primary analysis, PGR (approximately 0.75) and G-17 (approximately 0.72) achieved relatively higher values compared with other single markers. Notably, the combined detection model PG+CA72-4+G-17 reported by Lin (2021) demonstrated the highest AUC (0.88), substantially exceeding the performance of CA72-4 alone (0.68). Additionally, the PGI/PGII ratio yielded an AUC of approximately 0.78, outperforming CA72-4 as a standalone marker and suggesting greater robustness of pepsinogen-related indices.

The superimposed ROC plot in Figure 7B showed that most of the study points were located at the upper left of the diagonal, suggesting that the included serological indicators had a certain diagnostic discrimination ability as a whole. The combined detection of related research points was generally closer to the upper left area, indicating that it had high sensitivity and high specificity. The distribution of individual indicators was relatively more scattered, reflecting some fluctuations among different studies. The results of this figure were consistent with the direction of AUC and DOR analysis, that

is, the overall performance of combined detection was better than that of single index, and there were also some differences in efficacy between different individual indicators.

Sensitivity Analysis

Given the limited number of studies for most indicators (often only one or two) strict leave-one-out sensitivity analysis was feasible in only a few scenarios with three or more studies. Consequently, the sensitivity evaluation in this study was approached at two complementary levels.

First, stratified analysis: The main and supplementary analyses were conducted separately to prevent direct pooling of »early gastric cancer-specific studies« with broader »overall gastric cancer studies.« This stratification allowed assessment of result consistency across different populations. Findings indicated that G-17 and PGR consistently performed well in both layers, while the benefits of combined indicators were more pronounced in the supplementary analysis, suggesting a degree of robustness in the conclusions.

Second, outcome measure consistency: Both 2×2 diagnostic outcomes and AUC results were independently analyzed to minimize bias arising from differing outcome types. The relative ranking of biomarkers was generally consistent across sensitivity, specificity, DOR, and AUC, with G-17 and PGR consistently occupying the top positions. This

concordance reinforces the reliability of the primary conclusions regarding these markers.

Overall, although minor variations in indicator ranking were observed across analytical layers, PGR, G-17, and their associated combined panels consistently demonstrated superior diagnostic performance. Nevertheless, the limited study number precluded subgroup meta-regression analyses based on study quality, sample size, or assay methods, representing a key limitation of the current evidence base.

Discussion

Main Findings

Three principal findings emerged from this study: First, PGR demonstrated the most balanced diagnostic profile in the main analysis focusing specifically on EGC, with a sensitivity of 1.00 and specificity of 0.70 (26). These characteristics position PGR as a core candidate serum marker, particularly suitable for initial screening strategies aimed at minimizing missed cases. Second, G-17 exhibited consistent and robust performance across both main and supplementary analyses (27). In the primary analysis, it achieved the highest DOR (9.12) among all indicators, while maintaining moderate-to-high sensitivity. In the supplementary analysis, G-17 continued to show favorable sensitivity and DOR values, highlighting its stability and potential applicability across broader screening populations. Third, combined detection models primarily improved specificity and overall discriminatory ability. Specifically, G-17+PGR and G-17+PGII achieved specificities of 0.82 and 0.85, respectively, while the triple-panel combination (PG+CA72-4+G-17) reached an AUC of 0.88 (28, 29). These multimarker approaches outperformed most single indicators in comprehensive discrimination, reducing the likelihood of false-positive results and enhancing the reliability of early detection. Compared with previous systematic reviews, this study introduces a stratified framework that distinguishes EGC from overall gastric cancer. This approach more accurately reflects the diagnostic performance of markers in precancerous lesions and early-stage tumors. In contrast, prior studies that combined all disease stages may have either overestimated or underestimated the true screening value of individual indicators (30).

Possible Mechanisms for the Superiority of Combined Detection

Serum biomarkers for gastric cancer risk stratification reflect distinct aspects of gastric mucosal physiology and pathology. PGI, PGII, and the PGR serve as indicators of fundic gland integrity and se-

cretory function, with reduced levels signaling corpus-predominant atrophy (31). In chronic atrophic gastritis, particularly when accompanied by intestinal metaplasia, the loss of chief cells leads to diminished PGI synthesis and a consequent decline in PGR³¹. This serological pattern is more consistently observed in intestinal-type gastric adenocarcinoma, which typically arises on a background of extensive atrophy and metaplasia, whereas diffuse-type tumors often originate in non-atrophic mucosa and exhibit less pronounced pepsinogen suppression (32). G-17, in contrast, reflects antral mucosal status and acid-regulatory feedback mechanisms; its fasting levels decrease in the setting of antral atrophy but may increase under conditions of hypochlorhydria (31). Models combining pepsinogen with *Helicobacter pylori* serostatus, age, and sex demonstrate improved diagnostic performance for gastric cancer, including EGC, with an area under the receiver operating characteristic curve (AUC) of 0.851, compared to pepsinogen alone (AUC range: 0.598–0.813) (3). This supports the utility of multi-parameter approaches in noninvasive gastric cancer screening, where enhanced overall discriminatory capacity may help prioritize individuals for endoscopic evaluation (28).

Why Single Markers Still Have Clinical Value

Despite the generally superior performance of combined detection models, single serum markers retain important clinical value, particularly in settings with limited resources (31, 33). First, single-marker assays are more cost-effective, simpler to perform, and easier to implement on a large scale. This makes them suitable for initial population-wide screening in primary care facilities and community settings. Second, under the current evidence framework, PGR and G-17 have demonstrated relatively stable diagnostic performance (both AUC > 0.70) (15). In regions with constrained medical resources, these markers can serve as first-line screening tools, with individuals testing positive or borderline referred to tertiary centers for combined-marker testing or gastroscopy. Third, implementing combined-marker panels requires strict standardization, including controlled sampling timing (since G-17 is influenced by circadian rhythms and dietary intake), calibration across different platforms, and carefully defined threshold algorithms. In contrast, single-marker thresholds are comparatively straightforward, for example, PGR < 3.0 or < 2.0 is commonly used as a cutoff for fundic atrophy, enhancing clinical feasibility and operability (31, 34). Furthermore, when evidence for optimal weighting of combined models is still limited, single markers remain indispensable for preliminary risk stratification.

Clinical Significance

Serological biomarkers offer a noninvasive approach to gastric cancer (GC) risk stratification, with pepsinogen I/II ratio (PGR) and gastrin-17 (G-17) serving as foundational tools for initial screening due to their established role in detecting precancerous gastric conditions (35). For individuals at average risk, single-marker assessment using either PGR or G-17 may be considered, consistent with evidence supporting the diagnostic utility of serological panels in population-based triage (36). In contrast, high-risk populations—defined by age over 45 years, *Helicobacter pylori* positivity, or family history of GC—may benefit from a triple-marker panel combining PGR, G-17, and carbohydrate antigen 72-4 (CA72-4), which demonstrates improved diagnostic performance for early GC compared to individual markers alone (AUC = 0.883). Although CA72-4 exhibits limited sensitivity (33.3%) but high specificity (92.8%) in asymptomatic populations, its integration into multimodal strategies warrants further evaluation for refining risk stratification (37). Individuals with positive serological results should undergo prompt gastroscopy, as negative serum findings do not definitively exclude malignancy and must be interpreted alongside clinical risk factors (14). However, variability in screening protocols and thresholds across studies limits the generalizability of current models (38), highlighting an urgent need for standardized assay platforms and unified definitions. Future research must prioritize reporting early GC-specific outcomes and validating risk-adapted algorithms—such as those incorporating questionnaire-based scores with serological markers—to optimize endoscopic resource allocation and improve detection rates in high-incidence regions (14).

Compared with previous meta-analyses that pooled early and advanced gastric cancer together, the present study strictly focused on EGC via a stratified analytical framework, which reduces spectrum bias and provides more accurate estimates for early detection (39, 40). In addition, we systematically evaluated the incremental value of CA72-4 in combined panels, which was insufficiently addressed in prior reviews, highlighting the incremental value of this work.

Study Limitations

This study has several limitations that should be considered when interpreting the results: first, limitations in sample size and number of studies: the number of studies in the main analysis was small (most indicators supported by only 1–2 studies), preventing robust formal meta-analysis for most markers. Results relied more on descriptive integration, with insufficient statistical power and wide con-

fidence intervals. Second, inconsistent data completeness: some studies only reported AUC without complete 2×2 data, preventing full alignment across different outcome measures and precluding direct correspondence between AUC and 2×2 table analyses. Third, differences in disease spectrum: the supplementary analysis included studies on overall gastric cancer. While this helped expand the evidence base, it also introduced disease spectrum differences (early vs. advanced stages), which may overestimate or underestimate the true performance of indicators. Fourth, complex sources of heterogeneity: large variations existed across studies in thresholds, testing methods (ELISA vs. chemiluminescence), and control composition (healthy individuals vs. gastritis patients), which could not be fully explored by meta-regression and may affect cross-indicator comparisons. Fifth, incomplete risk of bias assessment: a full QUADAS-2 table has not yet been established, and only qualitative discussion was performed. Methodological quality analysis requires further supplementation. In addition, most studies used a case-control design, which may carry selection bias and overestimate performance in real screening settings. Sixth, publication bias: due to the small number of included studies, publication bias could not be effectively assessed, and unpublished negative results may exist.

Conclusion

In conclusion, circulating pepsinogen-related biomarkers and gastrin-17 show potential clinical utility for the adjunctive early detection of gastric cancer. Among single indicators, PGR showed the most favorable sensitivity profile, supporting its role as a practical first-line screening marker. G-17 exhibited robust overall diagnostic discrimination, while multimarker panels incorporating G-17 and CA72-4 further enhanced specificity and composite accuracy. These results suggest that serum biomarker profiling may serve as an effective adjunct to risk stratification and endoscopic triage. Future large-scale prospective studies with standardized assays and unified cutoffs are required to validate these preliminary observations.

Funding

This work was supported by the Self-funded scientific and technological Innovation Project of Foshan City, Project Number 2320001006438.

Conflict of interest statement

All the authors declare that they have no conflict of interest in this work.

References

- Li S, Tian X, Wei J, Shi Y, Zhang H, Huang Y. Long-term outcomes of additional surgery versus non-gastrectomy treatment for early gastric cancer after non-curative endoscopic submucosal dissection: a meta-analysis. *Chin Med J (Engl)* 2023; 136(5): 528–35.
- Li Y, Wang Y, Gao J, Xu W, Wang Y, Zhang F. SP1-Induced lncRNA ZFAS1 Contributes to Cell Proliferation and Migration in Gastric Cancer through AKT/mTOR Signaling. *Iran J Biotechnol* 2025; 23(2): e4071.
- Gotoda T, Iwasaki M, Kusano C, Seewald S, Oda I. Endoscopic resection of early gastric cancer treated by guideline and expanded National Cancer Centre criteria. *Br J Surg* 2010; 97(6): 868–71.
- Choi JJ, Lee NR, Kim SG, Lee WS, Park SJ, Kim JJ, et al. Short-Term Outcomes of Endoscopic Submucosal Dissection in Patients with Early Gastric Cancer: A Prospective Multicenter Cohort Study. *Gut Liver* 2016; 10(5): 739–48.
- Norero E, Baez S, Briceno E, Martinez C, Ceroni M, Escalona A, et al. Totally laparoscopic gastrectomy for the treatment of gastric tumors. *Rev Med Chil* 2015; 143(3): 281–8.
- Lavrenchuk GI, Bundiuk LS, Serkiz Ia I, Chebotarev EE. The action of radiation on the differentiation process of myogenic cells in culture. *Med Radiol (Mosk)* 1989; 34(2): 79–81.
- Soleimani S, Pouresmaeili F, Salahshoori Far I. Evaluation of lncRNAs as Potential Biomarkers for Diagnosis of Metastatic Triple-Negative Breast Cancer through Bioinformatics and Machine Learning. *Iran J Biotechnol* 2024; 22(3): e3853.
- Dondov G, Amarbayasgalan D, Batsaikhan B, Badamjav T, Batbaatar B, Tuvdenjamts B, et al. Diagnostic performances of pepsinogens and gastrin-17 for atrophic gastritis and gastric cancer in Mongolian subjects. *PLoS One* 2022; 17(10): e0274938.
- Loor A, Dumitrascu DL. *Helicobacter pylori* Infection, Gastric Cancer and Gastropanel. *Rom J Intern Med* 2016; 54(3): 151–6.
- Ogutmen Koc D, Bektas S. Serum pepsinogen levels and OLGA/OLGIM staging in the assessment of atrophic gastritis types. *Postgrad Med J* 2022; 98(1160): 441–5.
- Xu Y, Zhang P, Zhang K, Huang C. The application of CA72-4 in the diagnosis, prognosis, and treatment of gastric cancer. *Biochim Biophys Acta Rev Cancer* 2021; 1876(2): 188634.
- Liu C, Chen C, Lai H, Liang H, Zhong S, Guo G, et al. A New Method for Early Screening of Gastric Cancer (G17 and CA724 Dual-Labeled Time-Resolved Fluorescence Immunoassay). *Comput Math Methods Med* 2022; 2022: 1704948.
- Gasenko E, Bogdanova I, Sjomina O, Aleksandravica I, Kirsners A, Ancans G, et al. Assessing the utility of pepsinogens and gastrin-17 in gastric cancer detection. *Eur J Cancer Prev* 2023; 32(5): 478–84.
- Cai Q, Zhu C, Yuan Y, Feng Q, Feng Y, Hao Y, et al. Development and validation of a prediction rule for estimating gastric cancer risk in the Chinese high-risk population: a nationwide multicentre study. *Gut* 2019; 68(9): 1576–87.
- Lee SY, Moon HW, Ahn YS, Song JH, Kim JH, Sung IK. A Prospective Study on the Detection of Gastric Neoplasms Using Pepsinogen and Gastrin-17 Levels. *Gut Liver* 2025; 19(4): 548–58.
- Krike P, Appel MS, Shums Z, Polaka I, Kojalo I, Rudzite D, et al. Autoimmune gastritis serological biomarkers in gastric cancer patients. *Eur J Cancer Prev* 2024; 33(1): 29–36.
- He J, Chen WQ, Li ZS, Li N, Ren JS, Tian JH, et al. China guideline for the screening, early detection and early treatment of gastric cancer (2022, Beijing). *Zhonghua Zhong Liu Za Zhi [Chinese Journal of Oncology]* 2022; 44(7): 634–66.
- Gentilotti E, De Nardo P, Cremonini E, Gorska A, Mazzaferri F, Canziani LM, et al. Diagnostic accuracy of point-of-care tests in acute community-acquired lower respiratory tract infections. A systematic review and meta-analysis. *Clin Microbiol Infect* 2022; 28(1): 13–22.
- McGenity C, Clarke EL, Jennings C, Matthews G, Cartlidge C, Freduah-Agyemang H, et al. Artificial intelligence in digital pathology: a systematic review and meta-analysis of diagnostic test accuracy. *NPJ Digit Med* 2024; 7(1): 114.
- Kim J, Oh J, Osuka A, Kim D. Porphyrinoids, a unique platform for exploring excited-state aromaticity. *Chem Soc Rev* 2022; 51(1): 268–92.
- Tehan PE, Mills J, Leask S, Oldmeadow C, Peterson B, Sebastian M, et al. Toe-brachial index and toe systolic blood pressure for the diagnosis of peripheral arterial disease. *Cochrane Database Syst Rev* 2024; 10(10): CD013783.
- Randall M, Egberts KJ, Samtani A, Scholten RJ, Hooft L, Livingstone N, et al. Diagnostic tests for autism spectrum disorder (ASD) in preschool children. *Cochrane Database Syst Rev* 2018; 7(7): CD009044.
- Roth D, Pace NL, Lee A, Hovhannisyan K, Warenits AM, Arrich J, et al. Airway physical examination tests for detection of difficult airway management in apparently normal adult patients. *Cochrane Database Syst Rev* 2018; 5(5): CD008874.
- Tieges Z, Maclullich AMJ, Anand A, Brookes C, Cassarino M, O'Connor M, et al. Diagnostic accuracy of the 4AT for delirium detection in older adults: systematic review and meta-analysis. *Age Ageing* 2021; 50(3): 733–43.
- Chua WL, Rusli KDB, Aitken LM. Early warning scores for sepsis identification and prediction of in-hospital mortality in adults with sepsis: A systematic review and meta-analysis. *J Clin Nurs* 2024; 33(6): 2005–18.

26. Ni DQ, Lyu B, Bao HB, Jin HF, Zhao J, Xu Y, et al. Comparison of different serological methods in screening early gastric cancer. *Zhonghua Nei Ke Za Zhi* 2019; 58(4): 294–300.
27. Han X, Yu W. Value of serum pepsinogen ratio screening for early gastric cancer and precancerous lesions in Youcheng area. *World J Gastrointest Surg* 2024; 16(12): 3729–36.
28. Yan DH, Li YF. Diagnostic Value of Serum Pepsinogen and Helicobacter pylori Infection in Gastric Cancer Screening in Western Zhejiang. *Gastroenterol Res Pract* 2026; 2026: 9113753.
29. Wang SS, Guo ZP, Zhao XY. [Diagnostic Value of Serum Pepsinogen I/Pepsinogen II Combined with Tumor Markers for Helicobacter pylori-Positive Early-Stage Gastric Cancer]. *Sichuan Da Xue Xue Bao Yi Xue Ban* 2023; 54(1): 186–91.
30. Huang RJ, Park S, Shen J, Longacre T, Ji H, Hwang JH. Pepsinogens and Gastrin Demonstrate Low Discrimination for Gastric Precancerous Lesions in a Multi-Ethnic United States Cohort. *Clin Gastroenterol Hepatol* 2022; 20(4): 950–2 e3.
31. Agréus L, Kuipers EJ, Kupcinskis L, Malfertheiner P, Di Mario F, Leja M, et al. Rationale in diagnosis and screening of atrophic gastritis with stomach-specific plasma biomarkers. *Scand J Gastroenterol* 2012; 47(2): 136–47.
32. Bornschein J, Selgrad M, Wex T, Kuester D, Malfertheiner P. Serological assessment of gastric mucosal atrophy in gastric cancer. *BMC Gastroenterol* 2012; 12: 10.
33. Giai Gianetto Q, Michel V, Douche T, Nozeret K, Zaanen A, Colussi O, et al. Plasma Protein Biomarkers to Detect Early Gastric Preneoplasia and Cancer: A Prospective Study. *Int J Mol Sci* 2025; 26(20).
34. Huang YK, Yu JC, Kang WM, Ma ZQ, Ye X, Tian SB, et al. Significance of Serum Pepsinogens as a Biomarker for Gastric Cancer and Atrophic Gastritis Screening: A Systematic Review and Meta-Analysis. *PLoS One* 2015; 10(11): e0142080.
35. Lin Z, Bian H, Chen C, Chen W, Li Q. Application of serum pepsinogen and carbohydrate antigen 72-4 (CA72-4) combined with gastrin-17 (G-17) detection in the screening, diagnosis, and evaluation of early gastric cancer. *J Gastrointest Oncol* 2021; 12(3): 1042–8.
36. Li MY, Zhang DQ, Lu X, Chen WC. Comparison of two serological methods in screening gastric cancer and its precancerous condition. *Zhonghua Nei Ke Za Zhi* 2018; 57(12): 907–11.
37. Hu PJ, Chen MY, Wu MS, Lin YC, Shih PH, Lai CH, et al. Clinical Evaluation of CA72-4 for Screening Gastric Cancer in A Healthy Population: A Multicenter Retrospective Study. *Cancers (Basel)* 2019; 11(5).
38. Faria L, Silva JC, Rodriguez-Carrasco M, Pimentel-Nunes P, Dinis-Ribeiro M, Libanio D. Gastric cancer screening: a systematic review and meta-analysis. *Scand J Gastroenterol* 2022; 57(10): 1178–88.
39. Wang Y, Wang Y, Han W, Han M, Liu X, Dai J, et al. Intratumoral and fecal microbiota reveals microbial markers associated with gastric carcinogenesis. *Front Cell Infect Mi* 2024; 14: 1397466.
40. Ding H, Xu Q, Wang B, Lv Z, Yuan Y. MetaDE-Based Analysis of circRNA Expression Profiles Involved in Gastric Cancer. *Digest Dis Sci* 2020; 65(10): 2884–95.

Received: April 18, 2026

Accepted: May 20, 2026