

RESEARCH

Open Access



Quantifying the effects of the Correa pathway from *Helicobacter pylori* infection to gastric cancer: causal inference found in 6.8 million Koreans

Hyun Jin Oh¹, Chung Ho Kim², Hyun-Wook Park² and Bomi Park^{2*} 

Abstract

Background The Correa pathway describes the sequential progression of gastric cancer through *Helicobacter pylori* (*H. pylori*) infection, chronic gastritis, atrophic gastritis, intestinal metaplasia, adenoma, and gastric cancer. Clarifying the causal relationships along this pathway can provide robust evidence for targeted gastric cancer preventive strategies.

Methods This study was conducted using data from the Korean National Health Insurance Service, including 6,863,103 individuals who participated in the National Cancer Screening Program for gastric cancer in 2018, with a 2-year follow-up. We used doubly robust targeted maximum likelihood estimation (TMLE) to quantify the total effect of *H. pylori* eradication on incident gastric cancer and applied causal mediation analysis (CMA) to evaluate indirect pathways through atrophic gastritis/intestinal metaplasia and adenoma. Analyses were adjusted for age, sex, income, smoking, alcohol use, and family history of gastric cancer.

Results TMLE revealed that *H. pylori* infection significantly increased gastric cancer risk (relative risk [RR] = 6.40; 95% confidence interval [CI]: 6.05–6.77), as well as the risk of atrophic gastritis/intestinal metaplasia (RR = 1.41; 95% CI: 1.35–1.43) and adenoma (RR = 5.81; 95% CI: 5.68–5.94). atrophic gastritis/intestinal metaplasia substantially elevated the risk of adenoma (RR = 1.72; 95% CI: 1.67–1.77) and gastric cancer (RR = 1.33; 95% CI: 1.28–1.44). CMA showed that 3% of the *H. pylori* effect on gastric cancer was mediated through atrophic gastritis/intestinal metaplasia (odds ratio [OR] = 1.03, 95% CI: 1.02–1.04), and 36% was mediated via adenoma (OR = 1.97, 95% CI: 1.94–2.01). Among individuals with atrophic gastritis/intestinal metaplasia, adenoma accounted for 44% of the subsequent gastric cancer risk (OR = 1.13, 95% CI: 1.03–1.34).

Conclusions Our findings demonstrate that *H. pylori* infection is the most important causal determinant of gastric cancer and adenoma, indicating that prevention of *H. pylori* is central to gastric cancer prevention.

Keywords Stomach neoplasms, Precancerous Conditions, Adenoma, Metaplasia, Gastritis, Atrophic

*Correspondence:
Bomi Park
bpark@cau.ac.kr

¹Division of Gastroenterology, Department of Internal Medicine, Center for Cancer Prevention and Detection, National Cancer Center, Goyang, Republic of Korea

²Department of Preventive Medicine, College of Medicine, Chung-Ang University, 84 Heukseok-Ro, Dongjak-Gu, Seoul 06974, Republic of Korea



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Gastric cancer (GC) remains a significant global health concern. In 2022, an estimated 968,784 new cases of stomach cancer were diagnosed worldwide, making it the fifth most common cancer [1]. It also accounted for approximately 659,853 deaths, making it the fourth leading cause of cancer-related mortality [1]. East Asian countries, including South Korea, report the highest incidence and mortality rates. According to the latest Korean National Cancer Registry data (2022), approximately 26,662 new cases were recorded, highlighting a substantial national burden [2]. Despite advancements in diagnostic and therapeutic strategies, prognosis remains poor in advanced stages, underscoring the importance of early detection [3, 4]. Therefore, implementing effective screening and prevention programs is essential to reducing GC incidence and mortality [5].

The Correa pathway is a widely accepted model describing the histological progression of intestinal-type GC: chronic gastritis, atrophic gastritis (AG), intestinal metaplasia (IM), dysplasia, and invasive carcinoma [6, 7]. *Helicobacter pylori* (*H. pylori*) infection is a well-established initiator of chronic gastric inflammation, promoting these premalignant stages [8, 9]. Numerous sero-epidemiological studies have robustly linked *H. pylori* to an increased GC risk [8, 9]. AG and IM, frequently resulting from chronic *H. pylori*-induced gastritis, are known to significantly elevate cancer risk [6]. However, several uncertainties remain. The molecular mechanisms by which *H. pylori* drives carcinogenesis are not fully elucidated. Moreover, while each stage in the Correa pathway is recognized as a precursor, the precise causal contributions and the extent to which each intermediate lesion mediates the progression from *H. pylori* infection to invasive cancer require further elucidation [10].

Several epidemiological studies have investigated the associations between *H. pylori* infection, AG, IM, adenoma, and GC. Both case-control and cohort studies have consistently demonstrated strong correlations between these premalignant lesions and GC risk [11–13]. For instance, an extensive Japanese cohort study reported a significantly increased risk of GC among individuals with AG or IM identified via endoscopy [13]. Similarly, European observational studies have emphasized AG or IM as precancerous conditions warranting close surveillance [12]. However, most of these studies have relied on associative analyses rather than formal causal inference methods. These limitations include potential residual confounding, absence of rigorous quantitative mediation analysis, and methodological constraints in accurately estimating causal effects [11]. To address these limitations, contemporary causal inference techniques are necessary but have seldom been applied in this context.

Therefore, this study aimed to quantitatively evaluate the causal effects of *H. pylori* infection on GC development and to elucidate the mediating roles of AG, IM, and gastric adenoma. By clarifying the causal pathways and mediation proportions within the Correa pathway, our findings aim to provide robust evidence to support targeted prevention strategies in GC control.

Methods

Study design and population

This study used data from South Korea's National Health Insurance Service (NHIS), which covers approximately 97% of the population [14]. A retrospective cohort study was conducted among individuals aged 40–74 years who underwent GC screening through the National Cancer Screening Program (NCSP) in 2018. Because endoscopic findings of AG/IM were available in the national screening database only from 2018, the analytic cohort was restricted to individuals undergoing index screening endoscopy in 2018 or later. Exclusion criteria included a prior history of any cancer (ICD-10 codes C00–C97) before the baseline screening, GC diagnosed within one year after the baseline screening to minimize reverse causation and exclude latent cases [15], death within two years of screening, and missing data on key confounders. After applying these criteria, the final cohort included 6,863,103 participants.

Exposure and mediators

The exposure of interest was *H. pylori* infection, operationalized by identifying individuals who were prescribed eradication therapy in 2018 using prescription claims data from the NHIS. Potential mediators included AG, IM, and gastric adenoma, which were identified through the 2018 NCSP database. The presence of each condition was determined based on standardized records that reflect clinicians' assessments of endoscopic findings, supported by biopsy results.

Outcome

The primary outcome was GC diagnosed within 24 months after the index endoscopy. This 24-month interval was treated as a follow-up window, consistent with the biennial screening interval of the Korean National Cancer Screening Program. Incident GC was identified from NHIS claims using ICD-10 code C16 as a primary or secondary diagnosis in combination with cancer-specific Rare and Intractable Disease (RID) exemption codes (V193/V194), which indicate official registration for confirmed cancer. This combined definition has been validated against the Korea Central Cancer Registry (KCCR), demonstrating high accuracy (sensitivity 96.0% and positive predictive value 94.1%) for identifying GC cases in Korean administrative data [16].

Covariates

Covariates were selected based on prior literature and causal assumptions depicted in a Directed Acyclic Graph (DAG) developed for this study (Fig. 1) [17]. These variables were considered potential confounders of the relationship between *H. pylori* infection and subsequent GC risk. Information on covariates was obtained from the 2018 NHIS database, specifically from the eligibility file (demographics and socioeconomic status), general health screening questionnaire (e.g., smoking and alcohol use), and NCSP questionnaire (e.g., family history of GC). Socioeconomic status was measured using national health insurance premium contribution quartiles, which are commonly used as a proxy for household income in Korean population-based studies. The final covariates included age (categorized: 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70–74 years), sex (male, female), socioeconomic status (quartiles), smoking status (never, former, current), alcohol consumption (none, ≤ 3 times/week, ≥ 4

times/week), and family history of GC. *H. pylori* bacterial genotypes were not included in the minimally sufficient adjustment set. Because CagA and vacA genotypes arise only after *H. pylori* infection, they represent strain-specific virulence characteristics of the exposure rather than independent pre-exposure confounders [18]. Adjusting for such post-exposure variables may block part of the biological pathway from infection to AG/IM, adenoma, and GC, leading to over-adjustment. Therefore, these bacterial genotypes were not included in the minimally sufficient adjustment set.

Statistical analyses

Targeted Maximum Likelihood Estimation (TMLE) was applied to estimate the causal effect of *H. pylori* infection on GC incidence, using baseline data from 2018 as time zero [19]. TMLE is a doubly robust, semi-parametric method that minimizes bias from confounding and model misspecification by combining outcome

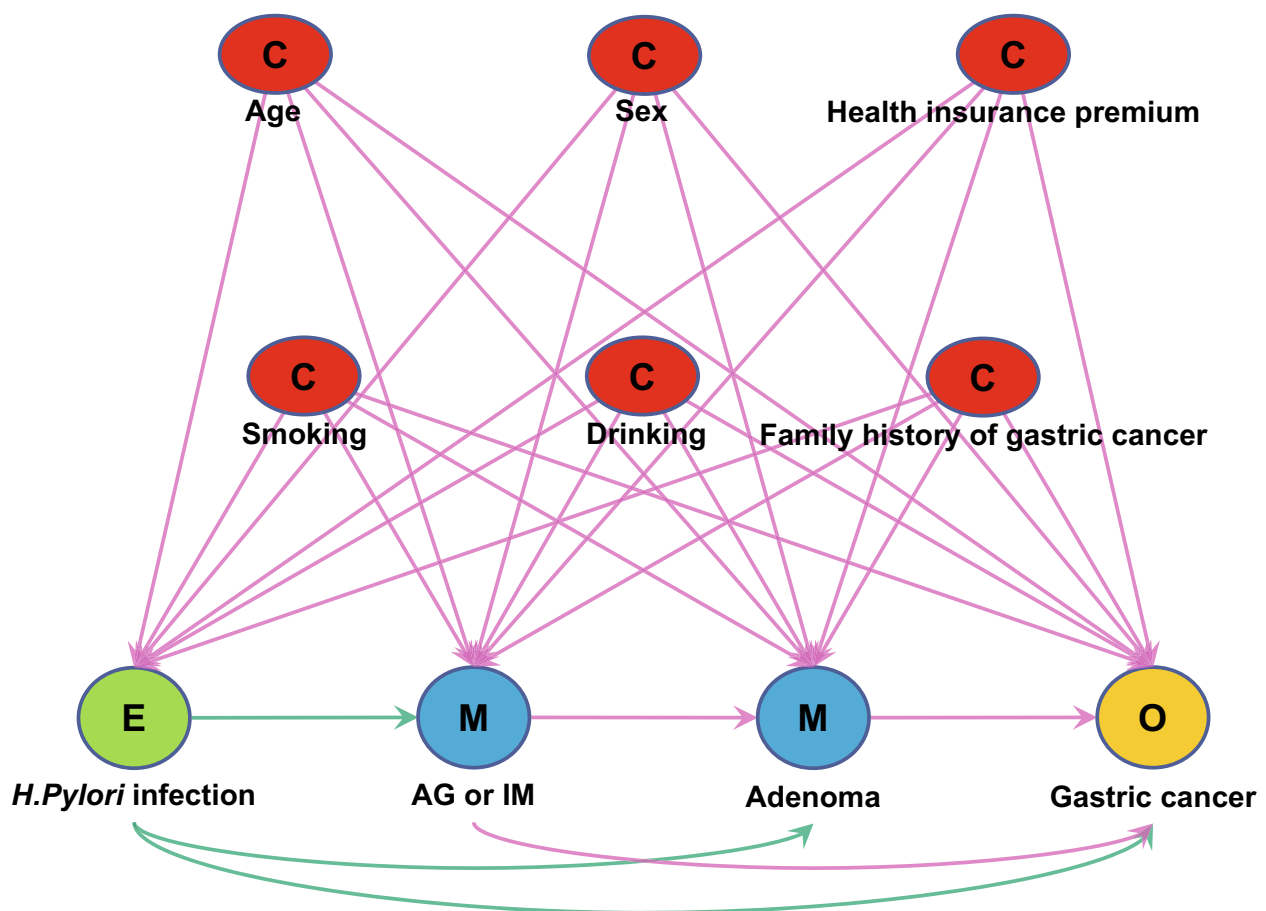


Fig. 1 Directed Acyclic Graph (DAG) illustrating the causal pathways leading to gastric cancer. Abbreviations: AG or IM, Atrophic gastritis or intestinal metaplasia. The exposure, *Helicobacter pylori* infection (shown in green, labeled as 'E'), and the primary outcome, gastric cancer (shown in yellow, labeled as 'O'), are highlighted. Covariates include age, sex, health insurance premium, smoking, alcohol consumption, family history of gastric cancer, atrophic gastritis (AG) or intestinal metaplasia (IM), and adenoma (shown in red, labeled as 'C'). Of these, age, sex, smoking, alcohol consumption, health insurance premium, and family history of gastric cancer act as confounders, while AG or IM and adenoma function as mediators (highlighted in blue). For visual clarity, arrows between confounders are omitted

regression with inverse probability weighting [20]. To flexibly model both the outcome and treatment assignment mechanisms, we applied Super Learner, an ensemble machine learning algorithm that combines multiple candidate learners through cross-validation. Our Super Learner library included generalized linear models and XGBoost, with fivefold cross-validation used to optimize predictive performance [21].

Causal Mediation Analysis (CMA) was conducted to investigate mediating pathways. Specifically, we assessed 1) whether AG or IM mediated the association between *H. pylori* infection and GC, 2) whether adenoma mediated the association between *H. pylori* infection and GC, and 3) whether adenoma mediated the association between AG or IM and GC. All mediation analyses were adjusted for the same covariates used in the TMLE analysis based on the constructed DAG. We assumed that there was no unmeasured confounding between the exposure and mediator, or between the mediator and outcome, and that no mediator–outcome confounders were themselves affected by the exposure. Mediation effects were quantified using natural direct and indirect effects. Natural direct effects represent the influence of *H. pylori* infection independent of mediators, while indirect effects capture pathways mediated by AG, IM, or adenoma [22].

All statistical analyses were conducted in R software (version 4.3.1). TMLE was implemented using the *tmle* package, and CMA was performed using the *medflex* package. Statistical significance was defined as a two-sided *P*-value < 0.05.

Results

A total of 7,102,798 individuals aged 40–74 years participated in the 2018 NCSP GC screening. The excluded individuals included 21,433 with a prior history of GC,

32,593 who died within 2 years of screening, and individuals with missing data on confounders (112,237 missing health insurance premium, 29,280 missing smoking status, 24,464 missing alcohol consumption, and 52,341 missing family history of GC). The final analytic cohort comprised 6,863,103 individuals (Fig. 2).

Among the 6,863,103 adults included, 2,624 developed GC during the two-year follow-up, while 6,860,479 did not. Individuals who developed GC were more likely to be older, male, current smokers, and to have *H. pylori* infection, AG, IM, and adenoma compared to those who remained cancer-free. Baseline characteristics of the study population are summarized in Table 1.

Individuals infected with *H. pylori* had a significantly higher risk of developing GC (adjusted relative risk [RR]=6.40; 95% confidence interval [CI]: 6.05–6.77). *H. pylori* infection also substantially increased the risk of intermediate precancerous conditions, including AG or IM (RR=1.41; 95% CI: 1.35–1.43) and adenoma (RR=5.81; 95% CI: 5.68–5.94). Moreover, AG or IM independently raised the risk of developing adenoma (RR=1.72; 95% CI: 1.67–1.77) and GC (RR=1.33; 95% CI: 1.28–1.44). Among all assessed pathways, the strongest association was observed between adenoma and GC (RR=72.1; 95% CI: 69.20–75.86). All analyses were adjusted for age, sex, health insurance premium, smoking status, alcohol consumption, and family history of GC based on the DAG presented in Fig. 1 (Table 2).

Figures 3, 4, and 5 present the CMA results for GC incidence. *H. pylori* infection had a substantial total effect on GC. The natural direct effect, not mediated by AG or IM, remained significant (odds ratio [OR]=5.58; 95% CI: 5.33–5.63), while the indirect effect via AG or IM was also significant (OR=1.03; 95% CI: 1.02–1.04; mediation proportion=3%). When adenoma was examined

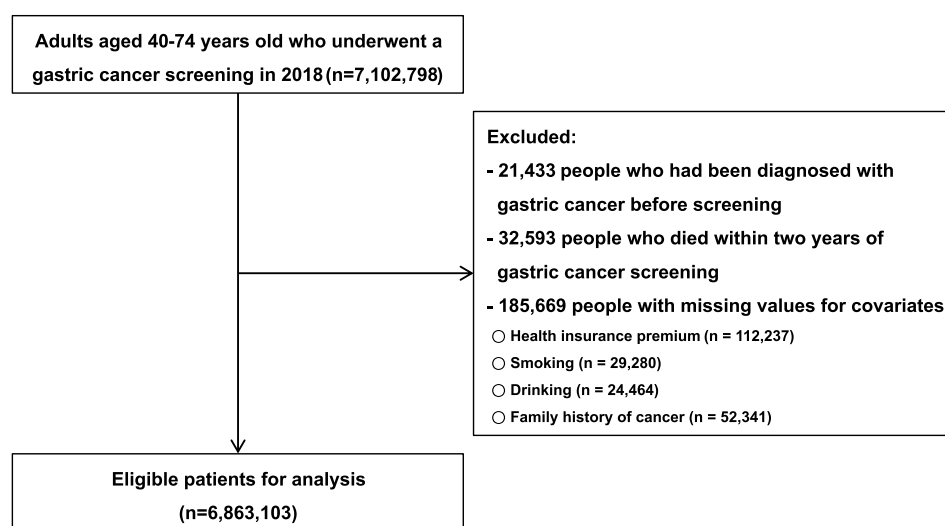


Fig. 2 Flowchart of study population selection. Covariate exclusions are not mutually exclusive

Table 1 The baseline characteristics of the study population

Characteristics	Total (N=6,863,103)	Non-gastric cancer group (N= 6,860,479)	Gastric cancer group (N= 2624)	P-value
Age				< 0.001
40 ~ 44	1,156,886 (16.8)	1,156,739 (16.9)	147 (5.6)	
45 ~ 49	923,474 (13.5)	923,282 (13.5)	192 (7.3)	
50 ~ 54	1,284,123 (18.7)	1,283,702 (18.7)	421 (16.0)	
55 ~ 59	997,762 (14.6)	997,270 (14.5)	492 (18.8)	
60 ~ 64	1,203,969 (17.6)	1,203,348 (17.5)	621 (23.7)	
65 ~ 69	650,132 (9.4)	649,770 (9.5)	362 (13.8)	
70 ~ 74	646,757 (9.4)	646,368 (9.4)	389 (14.8)	
Sex				< 0.001
Male	3,193,616 (46.6)	3,191,692 (46.5)	1924 (73.3)	
Female	3,669,487 (53.4)	3,668,787 (53.5)	700 (26.7)	
Health insurance premium				0.178
Medical Aid beneficiaries	248,804 (3.7)	248,719 (3.6)	85 (3.2)	
~ 25 percentile	1,347,019 (19.6)	1,346,512 (19.6)	507 (19.3)	
~ 50 percentile	1,322,785 (19.3)	1,322,251 (19.3)	534 (20.4)	
~ 75 percentile	1,646,985 (23.9)	1,646,324 (24.0)	661 (25.2)	
~ 100 percentile	2,297,510 (33.5)	2,296,673 (33.5)	837 (31.9)	
Smoking				< 0.001
Non-Smoker	4,357,937 (63.5)	4,356,847 (63.5)	1090 (41.5)	
Ex-Smoker	1,273,595 (18.6)	1,272,889 (18.6)	706 (26.9)	
Smoker	1,231,571 (17.9)	1,230,743 (17.9)	828 (31.6)	
Drinking				< 0.001
No drinking	4,703,278 (68.5)	4,701,446 (68.5)	1832 (69.8)	
≤ 3 times/week	1,564,668 (22.8)	1,564,049 (22.8)	619 (23.6)	
≥ 4 times/week	595,157 (8.7)	594,984 (8.7)	173 (6.6)	
Family history of gastric cancer				< 0.001
No	6,199,918 (90.3)	6,197,659 (90.3)	2259 (86.1)	
Yes	663,185 (9.7)	662,820 (9.7)	365 (13.9)	
Helicobacter Pylori infection				< 0.001
No	6,834,934 (99.6)	6,832,378 (99.6)	2556 (97.4)	
Yes	28,169 (0.4)	28,101 (0.4)	68 (2.6)	
Atrophic gastritis or Intestinal metaplasia				< 0.001
No	6,536,623 (95.2)	6,534,259 (95.3)	2364 (90.1)	
Yes	326,480 (4.8)	326,220 (4.7)	260 (9.9)	
Adenoma				< 0.001
No	6,845,012 (99.7)	6,842,962 (99.7)	2050 (78.1)	
Yes	18,091 (0.3)	17,517 (0.3)	574 (21.9)	

Values are presented as numbers (%)

as a mediator, the direct effect of *H. pylori* infection remained substantial (OR=3.32; 95% CI: 3.21–3.35), and the indirect effect through adenoma was also significant (OR=1.97; 95% CI: 1.94–2.01; mediation proportion=36%). Furthermore, AG or IM exhibited a substantial direct effect on GC (OR=1.16; 95% CI: 1.04–1.23) and an indirect effect via adenoma (OR=1.13; 95% CI: 1.03–1.34, accounting for 44% of the total effect. All mediation models were adjusted for age, sex, health insurance premium, smoking status, alcohol consumption, and family history of GC, following the DAG presented in Fig. 1.

Discussion

Our findings substantiate the Correa pathway by demonstrating that *H. pylori* infection significantly increases GC risk via defined histopathological stages, including AG, IM, and gastric adenoma. Using CMA, we quantified the extent to which each stage contributed to GC development and found that these preneoplastic and dysplastic changes collectively explain a meaningful portion of the effect of *H. pylori* infection. Among these, adenoma accounted for the largest proportion of mediation, while AG and IM also showed significant associations with subsequent adenoma or cancer development. Our findings demonstrate that *H. pylori* infection is the dominant

Table 2 Estimated causal association between variables involved in gastric carcinogenesis

Risk factor	Outcome	Risk ratio (95% CI)
HP	GC	6.40 (6.05–6.77)
HP	AG or IM	1.41 (1.35–1.43)
HP	Adenoma	5.81 (5.68–5.94)
AG or IM	Adenoma	1.72 (1.67–1.77)
AG or IM	GC	1.33 (1.28–1.44)
Adenoma	GC	72.10 (69.20–75.86)

Five-fold cross-validated targeted maximum likelihood estimation (TMLE) with super learner modeling was used to derive risk ratio

Models were adjusted for age, sex, health insurance premium, smoking, alcohol consumption, and family history of gastric cancer

Abbreviations: CI Confidence interval, HP *H. Pylori* infection, GC Gastric cancer, AG or IM Atrophic gastritis or intestinal metaplasia

risk factor for GC development in our study population. Using CMA, we quantified the extent to which AG, IM, and gastric adenoma contributed to GC development. Among these mediating pathways, adenoma accounted for the largest proportion, while AG and IM showed smaller but statistically significant associations with subsequent adenoma or cancer development. Taken together, these results indicate that while gastric carcinogenesis may proceed through stepwise histopathological stages, *H. pylori* infection itself remains the primary driver of GC risk.

By employing advanced causal inference methods, TMLE and CMA, our study provides robust causal estimates that overcome limitations inherent in traditional observational analyses. Randomized controlled trials (RCTs) evaluating the full Correa sequence are virtually infeasible due to the long latency of gastric carcinogenesis and ethical challenges in withholding treatment from high-risk individuals. As such, observational studies using rigorous analytical methods remain the most feasible and ethical approach to understanding the causal processes linking *H. pylori* infection to GC. Our use of TMLE and CMA enables precise quantification of both direct and indirect effects, minimizing bias from confounding and measurement error [23, 24]. This analytical clarity reinforces the biological plausibility of the Correa pathway and informs more refined strategies for surveillance and prevention.

Mechanistically, *H. pylori* infection induces chronic inflammation, oxidative stress, and DNA damage, triggering histological changes that contribute to GC development [8, 10, 25]. AG and IM are preneoplastic conditions marked by the replacement of gastric epithelium with intestinal-type cells, while adenomas are advanced dysplastic lesions with substantial malignant potential [26]. Our findings support this pathogenic sequence by identifying adenoma as a particularly significant mediator

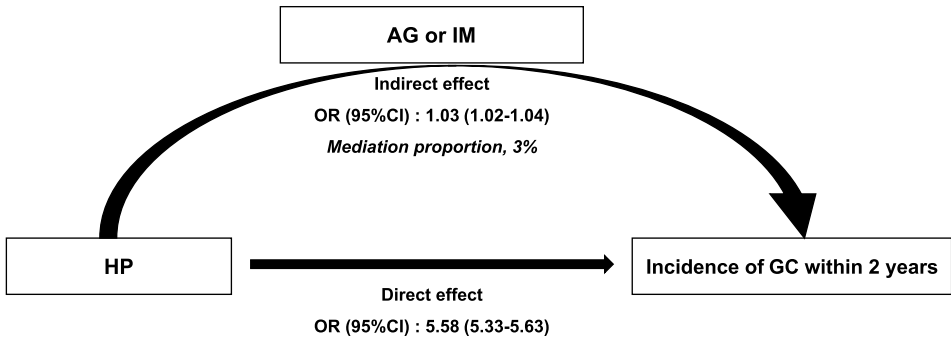


Fig. 3 Mediation effect of AG or IM on the association between HP and GC incidence within 2 years. Abbreviations: AG or IM, Atrophic gastritis or intestinal metaplasia; HP, *H. Pylori* infection; GC, Gastric cancer. Models adjusted for age, sex, health insurance premium, smoking, alcohol consumption, and family history of gastric cancer

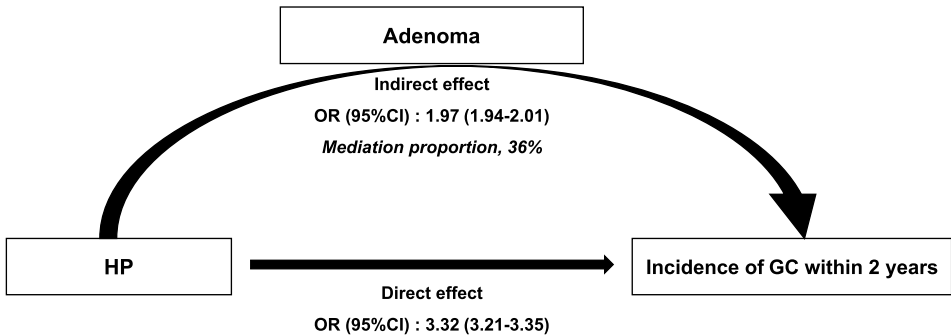


Fig. 4 Mediation effect of adenoma on the association between HP and GC incidence within 2 years. Abbreviations: HP, *H. Pylori* infection; GC, Gastric cancer. Models adjusted for age, sex, health insurance premium, smoking, alcohol consumption, and family history of gastric cancer

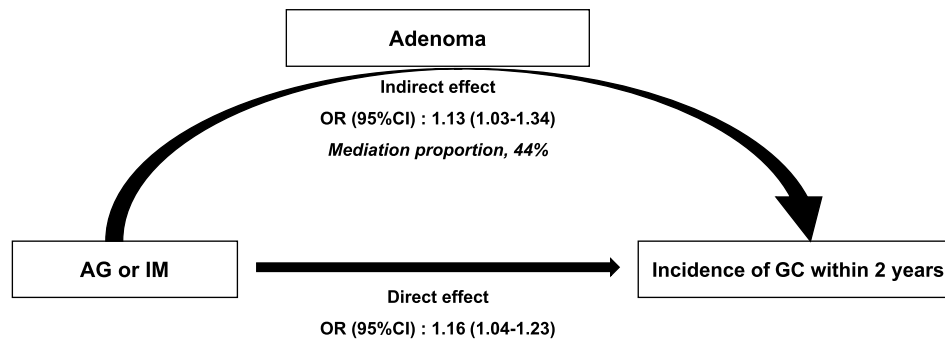


Fig. 5 Mediation effect of adenoma on the association between AG or IM and GC incidence within 2 years. *Abbreviations:* AG or IM, Atrophic gastritis or intestinal metaplasia; GC, Gastric cancer. Models adjusted for age, sex, health insurance premium, smoking, alcohol consumption, and family history of gastric cancer

in the *H. pylori*–GC pathway. The presence of adenoma accounted for 36% of the total effect of *H. pylori* on GC and mediated 44% of the effect of AG/IM on GC, underscoring its pivotal role in malignant transformation. These results highlight the importance of actively detecting and managing adenomatous lesions, in addition to *H. pylori* infection or early-stage AG/IM, particularly in high-risk populations. Effective GC prevention may require integrated approaches combining early detection of adenoma with eradication of *H. pylori* and mitigation of chronic inflammation.

The traditional model suggests sequential progression through AG and IM as key precancerous stages [6, 7]. The Correa hypothesis posits that these conditions are more likely to lead to the intestinal type of GC, highlighting the importance of their proper surveillance and management [6, 7]. Our findings align with this model: individuals with AG or IM had a significantly elevated risk of developing adenoma or GC within two years of screening endoscopy. This result supports previous evidence suggesting that the progression from IM to neoplastic lesions can occur relatively quickly. For instance, the 2020 American Gastroenterological Association (AGA) clinical guideline estimated an annual progression risk from IM to GC of 0.124% [27], and a European cohort study reported that progression from IM to adenoma or GC occurred within a median of 11–32 months—particularly shorter in cases of incomplete IM [28]. These data suggest that a uniform surveillance interval of two years, as currently applied in Korea’s NCSP, may not be optimal for individuals with AG or IM. Risk-stratified surveillance intervals should be considered for more effective prevention in this high-risk subgroup.

However, despite the strong association between AG or IM and GC, our mediation analysis revealed that these conditions accounted for only a small portion—approximately 3%—of the total effect of *H. pylori* infection on GC. This limited mediating effect suggests that while AG and IM are important indicators of gastric

carcinogenesis, additional or alternative pathways may contribute to malignant transformation. Previous studies have indicated that targeting AG or IM alone through pharmacological or lifestyle interventions has yielded only modest reductions in GC incidence [11, 12]. Emerging evidence supports the existence of *H. pylori*-induced carcinogenic mechanisms that bypass these intermediate lesions, including direct effects of chronic inflammation, oxidative stress, and genetic or epigenetic alterations [29, 30]. In this nationwide cohort, only a modest proportion of the association between *H. pylori* infection and GC was mediated through clinically detected AG or IM. Even acknowledging that our approach likely underestimates the total mediation, these findings indicate that, in current real-world practice, a substantial share of GCs may arise through pathways not fully captured by the classical Correa sequence. Thus, our data do not support the view that the Correa pathway is uniformly predominant in this population; instead, they highlight the probable importance of Correa-independent or as-yet-uncharacterized mechanisms. Given that our study population was restricted to individuals aged 40–74 years, these patterns may be more pronounced in this relatively younger cohort compared with GC populations that include older ages.”

From a prevention standpoint, these results imply that population-based strategies focusing solely on surveillance for AG/IM may have limited impact on overall GC incidence in contemporary Korea. Broader approaches—such as intensified *H. pylori* eradication, identification of molecular or genetic markers of non-Correa pathways, and risk-stratified endoscopic surveillance—are likely necessary to address the substantial burden of cancers arising outside the classical Correa sequence.

A major strength of our study lies in the application of rigorous causal inference methods—TMLE and CMA—to a large-scale, nationally representative cohort. These methods allow for doubly robust estimation and minimize bias due to confounding and model

misspecification, which are inherent challenges in traditional observational studies. By utilizing the comprehensive NHIS database, we were able to include a large number of individuals with diverse clinical and demographic backgrounds, enhancing both the internal and external validity of our findings. Furthermore, the integration of national endoscopy screening data with longitudinal claims enabled a temporally ordered analysis of *H. pylori* infection, intermediate lesions, and GC, supporting causal interpretations of complex disease pathways.

Despite its strengths, our study has several limitations. First, information on the severity and extent of AG and IM. Although AG and IM were identified using standardized endoscopic diagnoses, supported by biopsy findings, diagnosis based primarily on white-light endoscopy has limited accuracy and substantial interobserver variability, which may vary across screening centers. As a result, misclassification across a broad spectrum of AG and IM severity may have biased the estimated mediating effects toward the null, potentially leading to an underestimation of the contribution of AG and IM to gastric carcinogenesis in our analyses. Second, in the CMA, we assumed that there was no unmeasured confounding between the exposure and mediator, or between the mediator and outcome, and that the mediator–outcome confounders were unlikely to be affected by the exposure. This latter assumption is supported by prior evidence indicating that age, sex, socioeconomic status, smoking, and alcohol use function primarily as upstream determinants rather than exposure-induced variables [10, 31]. However, we cannot completely exclude the possibility of residual exposure-related confounding. Third, we excluded individuals who died within two years after baseline screening. This may have introduced selection bias, as those at highest risk of early death and GC may have been systematically excluded, potentially biasing our estimates toward the null. Moreover, this exclusion precluded proper consideration of competing risks, such as death prior to cancer diagnosis, which may have provided a more realistic depiction of risk trajectories. Fourth, participants with missing data on key confounders such as smoking and alcohol consumption were excluded from the analysis, potentially introducing additional selection bias. The use of multiple imputation or other missing data techniques may be useful to explore in future studies to assess the robustness of findings. Fifth, *H. pylori* infection was inferred from records of eradication therapy rather than confirmed by diagnostic testing. This proxy definition may have misclassified untreated infected individuals as uninfected and could not account for treatment failure or unsuccessful eradication, leading to non-differential misclassification and a potential bias toward the null. In addition, because gastric inflammation and mucosal injury may gradually regress following

successful eradication, the risk estimates derived from treated individuals may underestimate the true magnitude of risk associated with persistent active infection. Therefore, our findings likely represent a conservative estimate of the carcinogenic effect of ongoing *H. pylori* infection. In addition, *H. pylori* eradication therapy may have been under-ascertained in the claims data, because only reimbursed prescriptions are captured; eradication regimens prescribed as non-reimbursed (out-of-pocket) treatments were not available, which could have further contributed to exposure misclassification. Sixth, we lacked bacteriological data on *H. pylori* virulence genotypes. However, in Korea and East Asia, *H. pylori* strains show very high homogeneity, with near-universal CagA positivity (East-Asian EPIYA-D type) and predominant vacA s1/i1/m1 genotypes [32, 33]. Because these virulence markers vary little within this population and are post-exposure attributes of the infection, they are unlikely to introduce meaningful unmeasured confounding in our analysis. Lastly, information on histologic subtype and tumor location was not consistently available; therefore, our results reflect overall GC risk rather than subtype- or site-specific associations.

Conclusions

Using a large, nationally representative database from a population with high *H. pylori* prevalence, we found that clinically diagnosed AG/IM account for only a limited fraction of the *H. pylori*-associated risk of GC. Importantly, this finding indicates that a substantial proportion of GCs may arise from *H. pylori*-exposed stomachs even in the absence of clinically recognized intermediate lesions, underscoring that *H. pylori* infection itself is a dominant determinant of GC risk. These results indicate that, although the Correa pathway has long been regarded as the predominant route to GC, pathways beyond the classical sequence may play an important role in gastric carcinogenesis in Korea, at least among individuals aged 40–74 years. Although early detection and management of AG or IM remain relevant for individuals who follow this pathway, our results underscore that effective population-level prevention of GC and adenoma should primarily focus on preventing *H. pylori* infection itself.

Abbreviations

AG	Atrophic gastritis
AGA	American Gastroenterological Association
CI	Confidence interval
CMA	Causal mediation analysis
DAG	Directed Acyclic Graph
GC	Gastric cancer
<i>H. pylori</i>	<i>Helicobacter pylori</i>
ICD	International Classification of Diseases
IM	Intestinal metaplasia
NCSP	National Cancer Screening Program
NHIS	National Health Insurance Service

OR Odds ratio
 RCT Randomized controlled trial
 RR Relative risk
 TMLE Targeted maximum likelihood estimation

Acknowledgements

Not applicable.

Author's contributions

Conceptualization, B.P.; Data curation, C.H.K.; Formal analysis, C.H.K.; Funding acquisition, B.P.; Investigation, C.H.K., and B.P.; Methodology, H.J.O., and H-W.P.; Project administration, H.J.O.; Resources, H-W.P.; Software, C.H.K.; Supervision, B.P.; Validation, H.J.O., and H-W.P.; Visualization, H-W.P.; Writing-original draft, H.J.O., and B.P.; Writing-review & editing, C.H.K., and H-W.P. All authors have read and agreed to the final version of the manuscript.

Funding

This study was supported by the National R&D Program for Cancer Control through the National Cancer Center (NCC) funded by the Ministry of Health & Welfare, Republic of Korea (RS-2023-CC138164) and by the National Research Foundation of Korea grant funded by the Korea government Ministry of Science and ICT (RS-2025-00518173).

Data availability

The dataset supporting the conclusions of this article is available from the Korean National Health Insurance Service; however, restrictions apply to the availability of these data, which were accessed under license for the current study, and they are not publicly available.

Declarations

Ethics approval and consent to participate

The Institutional Review Board of the Chung-Ang University approved this study (IRB No. 1041078–20230201–HR-030 and 1041078–20250409–HR-115). As anonymized administrative data were used, the requirement of informed consent was waived. All datasets were stored on the servers of the Korean National Health Insurance Service and were fully anonymized prior to being accessed by the researchers. All procedures involved in this study adhered to the ethical standards of the relevant national and institutional committees on human experimentation, as well as the principles of the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 4 October 2025 / Accepted: 23 December 2025

Published online: 29 December 2025

References

1. Stomach cancer statistics. <https://www.wcrf.org/preventing-cancer/cancer-statistics/stomach-cancer-statistics/>. Accessed May 18, 2025.
2. Registry KCC. National Cancer Center: Annual report of cancer statistics in Korea in 2022. Ministry of Health and Welfare: Republic of Korea; 2024.
3. Kim GH, Liang PS, Bang SJ, Hwang JH. Screening and surveillance for gastric cancer in the United States: is it needed? *Gastrointest Endosc*. 2016;84(1):18–28.
4. Hamashima C, Shabana M, Okada K, Okamoto M, Osaki Y. Mortality reduction from gastric cancer by endoscopic and radiographic screening. *Cancer Sci*. 2015;106(12):1744–9.
5. Leung WK, Wu MS, Kakugawa Y, Kim JJ, Yeoh KG, Goh KL, et al. Screening for gastric cancer in Asia: current evidence and practice. *Lancet Oncol*. 2008;9(3):279–87.
6. Correa P, Piazuelo MB. The gastric precancerous cascade. *J Dig Dis*. 2012;13(1):2–9.
7. Correa P. Human gastric carcinogenesis: a multistep and multifactorial process—first American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. *Cancer Res*. 1992;52(24):6735–40.
8. Helicobacter, Cancer Collaborative G. Gastric cancer and *Helicobacter pylori*: a combined analysis of 12 case control studies nested within prospective cohorts. *Gut*. 2001;49(3):347–53.
9. Teh JL, Tan JR, Lau LJ, Saxena N, Salim A, Tay A, Shabbir A, Chung S, Hartman M, So JB. Longer examination time improves detection of gastric cancer during diagnostic upper gastrointestinal endoscopy. *Clin Gastroenterol Hepatol*. 2015;13(3):480–7. e482.
10. Wroblewski LE, Peek RM Jr, Wilson KT. *Helicobacter pylori* and gastric cancer: factors that modulate disease risk. *Clin Microbiol Rev*. 2010;23(4):713–39.
11. Song H, Ekheden IG, Zheng Z, Ericsson J, Nyren O, Ye W. Incidence of gastric cancer among patients with gastric precancerous lesions: observational cohort study in a low risk Western population. *BMJ*. 2015;351:h3867.
12. de Vries AC, van Grieken NC, Looman CW, Casparie MK, de Vries E, Meijer GA, et al. Gastric cancer risk in patients with premalignant gastric lesions: a nationwide cohort study in the Netherlands. *Gastroenterology*. 2008;134(4):945–52.
13. Uemura N, Okamoto S, Yamamoto S, Matsumura N, Yamaguchi S, Yamakido M, et al. *Helicobacter pylori* infection and the development of gastric cancer. *N Engl J Med*. 2001;345(11):784–9.
14. Lee J, Lee JS, Park SH, Shin SA, Kim K. Cohort profile: the National Health Insurance Service-National Sample Cohort (NHIS-NSC), South Korea. *Int J Epidemiol*. 2017;46(2):e15.
15. Hoang T, Lee J, Kim J. Comorbidity risk score in association with cancer incidence: results from a cancer screening cohort. *Cancers (Basel)*. 2020;12(7):1834.
16. Yang MS, Park M, Back JH, Lee GH, Shin JH, Kim K, et al. Validation of cancer diagnosis based on the national health insurance service database versus the national cancer registry database in Korea. *Cancer Research and Treatment: Official Journal of Korean Cancer Association*. 2022;54(2):352–61.
17. Hernán MA, Robins JM. Causal Inference: What If. 1st ed. Boca Raton: Chapman & Hall/CRC; 2020.
18. Yamaoka Y. Mechanisms of disease: *Helicobacter pylori* virulence factors. *Nat Rev Gastroenterol Hepatol*. 2010;7(11):629–41.
19. van der Laan MJ, Rose S. Targeted learning: causal inference for observational and experimental data. New York, NY: Springer; 2011.
20. Schnitzer ME, Moodie EE, Platt RW. Targeted maximum likelihood estimation for marginal time-dependent treatment effects under density misspecification. *Biostatistics*. 2013;14(1):1–14.
21. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol*. 2007;6:Article25.
22. VanderWeele T. Explanation in Causal Inference: Methods for Mediation and Interaction. Oxford University Press; 2015.
23. Díaz I, van der Laan MJ. Targeted data adaptive estimation of the causal dose-response curve. *J Causal Inference*. 2013;1(2):171–92.
24. Tchetgen EJ, Shpitser I. Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. *Ann Stat*. 2012;40(3):1816–45.
25. Wang F, Meng W, Wang B, Qiao L. *Helicobacter pylori*-induced gastric inflammation and gastric cancer. *Cancer Lett*. 2014;345(2):196–202.
26. Okamoto Y, Kanzaki H, Tanaka T, Sakae H, Abe M, Iwamuro M, et al. Gastric adenoma: a high incidence rate of developing carcinoma and risk of metachronous gastric cancer according to long-term follow-up. *Digestion*. 2021;102(6):878–86.
27. Gawron AJ, Shah SC, Altayar O, Davitkov P, Morgan D, Turner K, et al. AGA Technical Review on Gastric Intestinal Metaplasia-Natural History and Clinical Outcomes. *Gastroenterology*. 2020;158(3):705–31. e705.
28. Dinis-Ribeiro M, Lopes C, da Costa-Pereira A, Guilherme M, Barbosa J, Lomba-Viana H, et al. A follow up model for patients with atrophic chronic gastritis and intestinal metaplasia. *J Clin Pathol*. 2004;57(2):177–82.
29. Machado AM, Figueiredo C, Seruca R, Rasmussen LJ. *Helicobacter pylori* infection generates genetic instability in gastric cells. *Biochim Biophys Acta*. 2010;1806(1):58–65.
30. Polk DB, Peek RM Jr. *Helicobacter pylori*: gastric cancer and beyond. *Nat Rev Cancer*. 2010;10(6):403–14.
31. Malaty HM, Graham DY. Importance of childhood socioeconomic status on the current prevalence of *Helicobacter pylori* infection. *Gut*. 1994;35(6):742–5.
32. Kim SY, Woo CW, Lee YM, Son BR, Kim JW, Chae HB, et al. Genotyping CagA, VacA subtype, lceA1, and BabA of *Helicobacter pylori* isolates from Korean

patients, and their association with gastroduodenal diseases. *J Korean Med Sci.* 2001;16(5):579.

33. Jones KR, Joo YM, Jang S, Yoo Y-J, Lee HS, Chung I-S, et al. Polymorphism in the CagA EPIYA motif impacts development of gastric cancer. *J Clin Microbiol.* 2009;47(4):959–68.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.