gastroenterology insights

MDPI

Article

Helicobacter pylori Colonization and Intestinal Metaplasia Risk in an Andean Population: A Cross-Sectional Study from Peru

Yeny Uscamayta, Mercy Carolina Merejildo Vera and Jesús D. Rojas *

Programa Académico de Tecnología Médica en Laboratorio Clínico y Anatomía Patológica, Facultad de Ciencias de la Salud, Universidad Privada Norbert Wiener, Lima 15046, Peru
* Correspondence: jesus.rojas@uwiener.edu.pe

Abstract

Background: *Helicobacter pylori* infection is recognized as a Group I carcinogen and a primary driver of the gastric precancerous cascade. Andean populations in Peru experience disproportionately high rates of gastric cancer, yet data characterizing the association between *H. pylori* colonization and intestinal metaplasia remain limited in this setting. **Methods:** A cross-sectional analytical study was conducted among 169 patients undergoing upper gastrointestinal endoscopy at Hospital II EsSalud Abancay, Apurímac, Peru, between January and August 2025. Histopathological assessment determined *H. pylori* status, colonization intensity (negative, +, ++, +++) and intestinal metaplasia presence using the Updated Sydney System. Binary logistic regression estimated adjusted odds ratios (aORs) with 95% confidence intervals. **Results:** *H. pylori* infection was detected in 71 patients (42.0%), and intestinal metaplasia was present in 78 (46.2%). A dose–response pattern was observed: metaplasia prevalence increased from 18.4% in uninfected patients to 70.0%, 87.0%, and 92.9% across increasing colonization intensities (Cochran–Armitage $p < 0.001$). In multivariable analysis, colonization intensity (aOR = 10.19; 95% CI: 4.38–23.71), smoking (aOR = 11.70; 95% CI: 1.98–69.03), and alcohol consumption (aOR = 6.32; 95% CI: 1.74–22.95) remained independently associated with metaplasia. The primary binary logistic regression model demonstrated excellent discrimination (AUC = 0.921). **Conclusions:** *H. pylori* colonization intensity is strongly associated with intestinal metaplasia in this Andean population, with a clear dose–response gradient. These findings support prioritizing *H. pylori* screening and eradication strategies for gastric cancer prevention in highland Peru.

Keywords: *Helicobacter pylori*; precancerous conditions; stomach neoplasms; dose–response relationship; risk factors; alcohol drinking

Check for updates

Received: 17 December 2025
Revised: 2 February 2026
Accepted: 7 February 2026
Published: 13 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Gastric cancer remains a major global health burden, ranking as the fifth-most common malignancy and the fourth leading cause of cancer-related mortality worldwide, responsible for approximately 668,000 deaths annually [1,2]. The International Agency for Research on Cancer has classified *Helicobacter pylori* as a Group I carcinogen, recognizing its fundamental role in gastric carcinogenesis [3]. The distribution of this disease is notably uneven, with particularly elevated incidence and mortality rates observed across Latin America, with approximately 97,000 deaths annually across the region [4]. Within Peru, highland regions exhibit some of the highest gastric cancer rates in the Americas, a pattern that has persisted for decades despite advances in diagnosis and treatment [5]. The prevalence of *H. pylori* infection in Peruvian populations is among the highest globally, ranging from 70% to

90% in urban and rural settings, with indigenous Andean communities demonstrating near-universal infection rates [6,7]. This geographic disparity suggests that regional factors, whether environmental, infectious, or genetic, may contribute to the heightened risk observed in Andean populations.

The pathogenesis of gastric adenocarcinoma of the intestinal type follows the well-characterized Correa cascade, progressing sequentially from chronic non-atrophic gastritis through atrophic gastritis, intestinal metaplasia, dysplasia, and ultimately invasive carcinoma [8,9]. Within this multistep model, intestinal metaplasia represents a critical inflection point where reversibility diminishes substantially and malignant transformation risk increases significantly [10]. The annual progression rate from intestinal metaplasia to gastric cancer has been estimated at 0.0121 per 100 person-years, with incomplete metaplasia subtypes conferring a markedly elevated risk compared to complete metaplasia [11,12]. Furthermore, a dose–response relationship between *H. pylori* colonization density and histological severity has been documented, whereby higher bacterial burden correlates with more extensive mucosal damage and accelerated progression through the precancerous cascade [13,14].

H. pylori exhibits remarkable genomic diversity, which substantially modulates its pathogenic potential and clinical outcomes [15]. This bacterium possesses several well-characterized virulence factors that contribute to gastric epithelial damage and influence progression through the precancerous cascade. The cytotoxin-associated gene A (*cagA*), located within the *cag* pathogenicity island (*cag* PAI), encodes an immunodominant oncoprotein that is translocated into host gastric epithelial cells via a bacterial type IV secretion system [16]. Following intracellular delivery, CagA undergoes tyrosine phosphorylation at specific EPIYA motifs and interacts with multiple host signaling molecules, including SHP-2 phosphatase, thereby disrupting cellular proliferation, polarity, and intercellular adhesion [17]. The CagA protein has been demonstrated to upregulate β-catenin-dependent *cdx1* transcription and intestinal-differentiation markers such as MUC2, both of which are characteristic of gastric intestinal metaplasia [18].

The vacuolating cytotoxin A (*vacA*) gene exhibits significant allelic diversity, with polymorphisms in the signal (s1/s2), middle (m1/m2), and intermediate (i1/i2) regions determining cytotoxic activity [15]. Strains harboring the *vacA* s1m1 genotype demonstrate enhanced vacuolating activity and are associated with more severe mucosal inflammation, epithelial damage, gastric atrophy, and intestinal metaplasia compared to s2m2 variants [19]. Additional virulence determinants, including the blood group antigen-binding adhesin BabA2, outer inflammatory protein OipA, and duodenal ulcer-promoting gene DupA, further modulate host–pathogen interactions and disease outcomes [15,20]. The interplay between these virulence factors and host immune responses creates a complex pathogenic milieu that determines whether infection results in asymptomatic carriage, peptic ulceration, or neoplastic progression. Notably, the geographic distribution of virulent genotypes varies considerably, with *cagA*-positive and *vacA* s1m1 strains predominating in East Asian and Latin American populations characterized by elevated gastric cancer burden [21–24].

Despite substantial advances in understanding *H. pylori*-associated gastric pathology, significant knowledge gaps persist regarding disease dynamics in Andean populations. The “altitude enigma” describes the paradoxical concentration of gastric cancer burden in mountainous regions of Pacific Latin America, suggesting that factors beyond infection prevalence alone modulate carcinogenic risk [25,26]. Recent evidence indicates that disrupted host–*H. pylori* coevolution, wherein European-lineage bacterial strains infect hosts with predominantly Amerindian genetic ancestry, may contribute to the exceptional disease burden observed in highland populations [27,28]. The Apurímac region of Peru, situated at 2378 m altitude with socioeconomically disadvantaged communities, presents a

unique epidemiological context requiring locally derived evidence to inform prevention strategies [5,29].

The present study was therefore designed to address three objectives. First, to determine the prevalence of *H. pylori* infection and intestinal metaplasia among patients undergoing upper gastrointestinal endoscopy at a regional hospital in the Peruvian highlands. Second, to evaluate whether a dose–response relationship exists between colonization intensity and metaplasia. Third, to identify sociodemographic and behavioral factors independently associate with metaplasia after adjustment for colonization status. These findings are intended to inform gastric cancer prevention strategies tailored to this vulnerable population.

2. Materials and Methods

2.1. Study Design and Setting

A cross-sectional analytical study was conducted at Hospital II EsSalud Abancay, a secondary referral center located in the Apurímac region of the Peruvian Andes at an altitude of 2378 m above sea level. This facility serves as the primary gastroenterology reference center for the region, attending patients from predominantly rural, Quechua-speaking communities. Data collection occurred between January and August 2025. The study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for cross-sectional studies.

2.2. Study Population

The source population comprised 300 adult patients who underwent upper gastrointestinal endoscopy with gastric biopsy during the study period. Sample size was calculated using the standard formula for estimating a proportion in a finite population:

$$n = \frac{N \times Z^2 \times p \times (1-p)}{d^2 \times (N-1) + Z^2 \times p \times (1-p)}$$

where N = 300 (source population), Z = 1.96 (95% confidence), d = 0.05 (margin of error), and p = 0.50 (expected prevalence).

The use of 50% expected prevalence was adopted because no prior local data existed on intestinal metaplasia prevalence in the Apurímac region of Peru; this conservative assumption maximizes the variance term $p(1-p)$ = 0.25, yielding the largest required sample size for any true prevalence value [30–32]. Based on these parameters, the minimum required sample size was 169 participants.

Inclusion criteria were (1) age of 18–80 years old; (2) clinical indication for upper endoscopy; (3) availability of histopathological results from gastric biopsy; and (4) complete medical records. Exclusion criteria included the following: (1) prior *H. pylori* eradication therapy; (2) history of gastric surgery; (3) confirmed gastric malignancy; (4) use of proton pump inhibitors, antibiotics, or bismuth compounds within 4 weeks prior to endoscopy; and (5) incomplete or illegible clinical documentation. Consecutive patients meeting eligibility criteria were enrolled using simple random sampling from the gastroenterology service registry.

Gastric biopsies were obtained exclusively from the antral mucosa during upper gastrointestinal endoscopy. Two biopsy specimens were collected from the gastric antrum (lesser and greater curvature) following the Updated Sydney System protocol [33,34]. The antrum was selected as the target sampling site based on established evidence that this region represents the preferential site for *H. pylori* colonization and the initial development of intestinal metaplasia in the context of antral-predominant gastritis, which character-

izes most of the *H. pylori*-associated pathology in high-prevalence populations [35,36] (Figure S1).

2.3. Variables and Measurements

The primary outcome was intestinal metaplasia, defined as presence or absence based on histopathological examination. A secondary outcome classified the metaplasia subtype as absent, complete, or incomplete. The main predictor variable was *H. pylori* colonization intensity, measured as an ordinal variable with four categories: negative (0), mild (+), moderate (++), and severe (+++).

Covariates included sociodemographic and lifestyle factors extracted from medical records: age (continuous, years), sex (male/female). Family history of cancer was recorded as the presence of any malignancy in first-degree relatives (parents, siblings, or children) as documented in the patient's medical record. This variable was captured dichotomously (yes/no) without specification of cancer type or anatomical site, reflecting the manner in which family history data were routinely collected in clinical documentation at the institution. Current smoking status (yes/no) and alcohol consumption (yes/no) were also included. All variables were operationalized according to standardized definitions prior to data extraction.

2.4. Histopathological Evaluation

Gastric biopsies were obtained during upper endoscopy following the Updated Sydney System protocol, with specimens collected from standardized anatomical sites including antrum and corpus. Tissue samples were fixed in 10% buffered formalin, processed using standard techniques, and stained with hematoxylin–eosin for morphological assessment and Giemsa for *H. pylori* identification.

H. pylori colonization density was graded semi-quantitatively according to Sydney System criteria: negative (no organisms identified), mild (+; scattered organisms), moderate (++; clusters of organisms in multiple high-power fields), and severe (+++; confluent dense colonies). *H. pylori* colonization density was graded semi-quantitatively according to the Updated Sydney System criteria as described by Dixon et al. [33] and standardized by Stolte and Meining [34]. Using the combination of hematoxylin–eosin (H&E) and modified Giemsa staining, colonization intensity was classified into four categories based on examination at 400× magnification:

- Negative (Grade 0): No organisms identified after examination of all available tissue
- Mild (+; Grade 1): Scattered organisms, difficult to find, requiring careful searching across multiple high-power fields
- Moderate (++; Grade 2): Organisms readily identifiable, dispersed in several areas, with small clusters on surface epithelium
- Severe (+++; Grade 3): Large clusters or confluent layers of organisms covering extensive portions of the mucosal surface and within gastric pits

These grading criteria correspond to the visual analog scale anchors provided in the Updated Sydney System, which has demonstrated acceptable interobserver reproducibility in validation studies [33,34]. Representative histopathological images illustrating each colonization grade are provided in Supplementary Figure S2.

Intestinal metaplasia was classified histologically as complete (type I), characterized by absorptive enterocytes with brush border and goblet cells secreting sialomucins, or incomplete (types II/III), featuring intermediate or colonic-type epithelium with sulfomucin-secreting goblet cells. All histopathological evaluations were performed by experienced pathologists blinded to clinical and endoscopic findings (Supplementary Figure S2).

Intestinal metaplasia subtype classification was based on morphological criteria using H&E staining, following established histological patterns [37]. Complete (type I) intestinal metaplasia was identified by the presence of mature absorptive enterocytes with well-defined brush border and Paneth cells at the crypt base, in addition to goblet cells. Incomplete intestinal metaplasia (types II/III) was characterized by the presence of irregular goblet cells distributed throughout the epithelium, absence of Paneth cells, lack of mature absorptive cells, and distorted glandular architecture with intermediate features between gastric and colonic epithelium.

All histopathological assessments were performed by a single board-certified pathologist with more than 15 years of experience in gastrointestinal pathology, who was blinded to clinical and endoscopic findings at the time of microscopic evaluation. Standardized Updated Sydney System criteria were applied to minimize subjectivity in grading.

2.5. Statistical Analysis

Descriptive statistics were calculated for all variables. Continuous variables were expressed as mean (standard deviation) or median [interquartile range] according to distribution normality assessed by the Shapiro–Wilk test. Categorical variables were presented as absolute frequencies and percentages.

Bivariate associations between *H. pylori* status and intestinal metaplasia were evaluated using Pearson's chi-square test. Crude odds ratios (ORs) with 95% confidence intervals (CIs) were calculated for each potential risk factor. The dose–response relationship between colonization intensity and metaplasia prevalence was assessed using the Cochran–Armitage test for trend.

The primary analysis employed binary logistic regression with intestinal metaplasia (present vs. absent) as the outcome. The full model included colonization intensity (coded 0–3), sex, age, family history, smoking, and alcohol consumption. Model selection was performed using backward elimination based on Bayesian Information Criterion (BIC). A parsimonious model retained only variables with $p < 0.05$ in the primary model.

As a secondary exploratory analysis, multinomial logistic regression was conducted with metaplasia subtype (absent/complete/incomplete) as the outcome. Given the low events per variable (EPV = 3.5) for the incomplete metaplasia category, these results are reported with appropriate caution regarding precision and stability of estimates.

As a sensitivity analysis, Firth's penalized likelihood regression was applied to address potential bias from quasi-complete separation observed in the data (e.g., zero cases of complete metaplasia among patients with + colonization intensity). This method adds a penalty term to the likelihood function that reduces small-sample bias in maximum likelihood estimates. For the multinomial subtype analysis, separate penalized binary logistic regressions were fitted for each comparison (complete vs. absent and incomplete vs. absent), as recommended when standard multinomial models cannot accommodate Firth's correction [38].

Pre-model assumptions were verified: linearity in the logit was assessed using the Box–Tidwell test, and multicollinearity was evaluated using variance inflation factors (VIFs), with values < 5 considered acceptable. Model fit was assessed using the Hosmer–Lemeshow goodness-of-fit test and discrimination was quantified by the area under the receiver operating characteristic curve (AUC). Influential observations were identified using Cook's distance, and sensitivity analyses excluding outliers were performed to assess estimate stability.

Post hoc statistical power was calculated based on observed effect sizes. All analyses were conducted using R software version 4.5.2 (R Foundation for Statistical Computing, Vienna, Austria). A two-sided p-value < 0.05 was considered statistically significant.

2.6. Ethical Considerations

This study was approved by the Institutional Committee of Ethics and Scientific Integrity of Universidad Privada Norbert Wiener (Protocol No. 1324-202, approved on 10 September 2025) prior to data collection. Given the retrospective nature of the study utilizing de-identified medical records, a waiver of individual informed consent was granted by the ethics committee.

Patient confidentiality was rigorously maintained throughout the study. All participants were assigned unique alphanumeric codes, and personal identifiers were removed from the dataset prior to analysis. Data were stored securely in password-protected files with access restricted exclusively to the study investigators. Participant selection followed principles of equity, ensuring fair inclusion without discrimination or exploitation.

The research was conducted in accordance with the Declaration of Helsinki and applicable Peruvian regulations governing human subject research. Scientific integrity was upheld through rigorous data collection and analysis procedures, ensuring that all findings accurately reflect the observed data.

3. Results

3.1. Participant Characteristics

A total of 169 patients met the inclusion criteria and were enrolled in the study. The mean age was 55.9 years (SD = 9.9), with 49.7% males. *H. pylori* infection was detected in 71 patients (42.0%), while 98 (58.0%) were negative. Table 1 presents baseline characteristics stratified by *H. pylori* status. Patients with *H. pylori* infection showed significantly higher rates of smoking (52.1% vs. 8.2%, $p < 0.001$) and alcohol consumption (69.0% vs. 30.6%, $p < 0.001$) compared to *H. pylori*-negative individuals. Male sex was more prevalent among *H. pylori*-positive patients (63.4% vs. 39.8%, $p = 0.003$). No significant differences were observed in age ($p = 0.217$) or family history of gastric cancer ($p = 0.132$) between groups.

Table 1. Baseline characteristics of the study population ($n = 169$).

Characteristic	Overall ($n = 169$) [1]	*H. pylori* Negative ($n = 98$) [1]	*H. pylori* Positive ($n = 71$) [1]	p-Value [2]
Age, years	55.6 (9.9)	57.6 (9.7)	52.9 (8.8)	0.001
Male sex				<0.001
Female	85 (50.3%)	62 (63.3%)	23 (32.4%)	
Male	84 (49.7%)	36 (36.7%)	48 (67.6%)	
Family history of cancer	59 (34.9%)	37 (37.8%)	22 (31.0%)	0.455
Current or former smoking	45 (26.6%)	21 (21.4%)	24 (33.8%)	0.105
Alcohol consumption	79 (46.7%)	50 (51.0%)	29 (40.8%)	0.249

[1] Mean (SD); *n* (%); [2] Wilcoxon rank sum test; Pearson's chi-squared test.

Regarding *H. pylori* status, 71 patients (42.0%) tested positive on histopathological examination, while 98 (58.0%) were negative. Among *H. pylori*-positive cases, colonization intensity was graded as mild (+) in 20 patients (11.8%), moderate (++) in 23 patients (13.6%), and severe (+++) in 28 patients (16.6%).

Intestinal metaplasia was identified in 78 patients (46.2%), while 91 patients (53.8%) had no metaplastic changes. Among those with metaplasia, 57 (33.7% of total) had complete metaplasia (type I) and 21 (12.4% of total) had incomplete metaplasia (types II/III).

Established risk factors were prevalent in this population: 59 patients (34.9%) reported a family history of cancer, 45 (26.6%) were current or former smokers, and 79 (46.7%) reported alcohol consumption. The complete demographic and clinical characteristics of the study population are presented in Table 1.

3.2. Bivariate Associations

Bivariate analysis revealed strong associations between *H. pylori* infection and the presence of intestinal metaplasia (Table 2). Binary *H. pylori* status was profoundly associated with metaplasia, with infected patients demonstrating 24.24-fold higher odds of metaplasia compared to uninfected patients (OR = 24.24; 95% CI: 10.66–55.12; $p < 0.001$). Similarly, when analyzed as an ordinal variable, each incremental increase in colonization intensity level was associated with a 4.90-fold increase in odds of metaplasia (OR = 4.90; 95% CI: 3.17–8.30; $p < 0.001$).

Table 2. Bivariate associations between potential risk factors and intestinal metaplasia ($n = 169$).

Variable	OR	95% CI	*p*-Value
H. pylori (Positive vs. Negative)	24.24	10.66–55.12	<0.001 *
Colonization intensity (per level)	4.90	3.17–8.30	<0.001 *
Age (per year)	0.98	0.949–1.011	0.203
Sex (Male vs. Female)	2.44	1.31–4.54	0.004 *
Family history (Yes vs. No)	2.53	1.32–4.85	0.005 *
Smoking (Yes vs. No)	3.63	1.75–7.51	<0.001 *
Alcohol consumption (Yes vs. No)	2.80	1.50–5.22	0.001 *

OR: odds ratio; CI: confidence interval. * statistically significant at $p < 0.05$. Colonization intensity coded as ordinal (0 = negative, 1 = +, 2 = ++, 3 = +++) OR for colonization intensity represents odds increase per one-level increment.

Among sociodemographic factors, male sex was significantly associated with metaplasia (OR = 2.44; 95% CI: 1.31–4.54; $p = 0.004$). A positive family history of cancer also conferred elevated risk (OR = 2.53; 95% CI: 1.32–4.85; $p = 0.005$). Lifestyle factors demonstrated notable associations: smoking was strongly associated with metaplasia (OR = 3.63; 95% CI: 1.75–7.51; $p < 0.001$), as was alcohol consumption (OR = 2.80; 95% CI: 1.50–5.22; $p = 0.001$).

Notably, age as a continuous variable did not demonstrate a significant association with metaplasia in bivariate analysis (OR = 0.98; 95% CI: 0.949–1.011; $p = 0.203$), suggesting that within this population age range, infection status and lifestyle factors may be more influential determinants.

3.3. Sex-Stratified Analysis

Intestinal metaplasia prevalence differed significantly by sex, with higher rates observed in males (57.1%, 48/84) compared to females (35.3%, 30/85; crude OR = 2.44, 95% CI: 1.31–4.54, $p = 0.004$). This difference was consistent across age strata: in patients aged <50 years, metaplasia prevalence was 47.8% in males vs. 22.7% in females; in those aged 50–59 years, 61.3% vs. 37.0%; and in those ≥60 years, 61.3% vs. 48.0% (Figure S3).

Sex differences in lifestyle factors were marked. Smoking prevalence was substantially higher among males (45.2%) than females (8.2%, $p < 0.001$), as was alcohol consumption (67.9% vs. 25.9%, $p < 0.001$). *H. pylori* positivity also differed by sex (52.4% in males vs. 31.8% in females, $p = 0.006$). These differences in exposure profiles likely explain the attenuation of sex as an independent predictor in the adjusted model (aOR = 0.22, 95% CI: 0.04–1.20, $p = 0.081$), suggesting that the crude association between male sex and metaplasia is largely mediated through lifestyle-related confounders.

Formal interaction analysis between sex and colonization intensity was not performed due to limited statistical power with the observed number of events in each subgroup (30 in females, 48 in males). The stratified dose–response relationship appeared consistent in direction for both sexes, though sample sizes preclude definitive conclusions about effect modification.

3.4. Dose–Response Relationship

A striking dose–response relationship was observed between *H. pylori* colonization intensity and metaplasia prevalence (Table 3, Figure 1). Among patients without *H. pylori* infection (negative), only 18.4% (18/98) had metaplasia. This prevalence increased sharply with increasing colonization intensity: 70.0% (14/20) in the mild (+) colonization group, 87.0% (20/23) in the moderate (++) colonization group, and 92.9% (26/28) in the severe (+++) colonization group.

Table 3. Dose–response relationship between *H. pylori* colonization intensity and intestinal metaplasia prevalence.

Colonization Intensity	Total (*n*)	Metaplasia (*n*)	Prevalence (%)
Negative	98	18	18.4
+ (Mild)	20	14	70.0
++ (Moderate)	23	20	87.0
+++ (Severe)	28	26	92.9

Cochran–Armitage test for trend: $\chi^2 = 68.67$, $p < 0.001$.

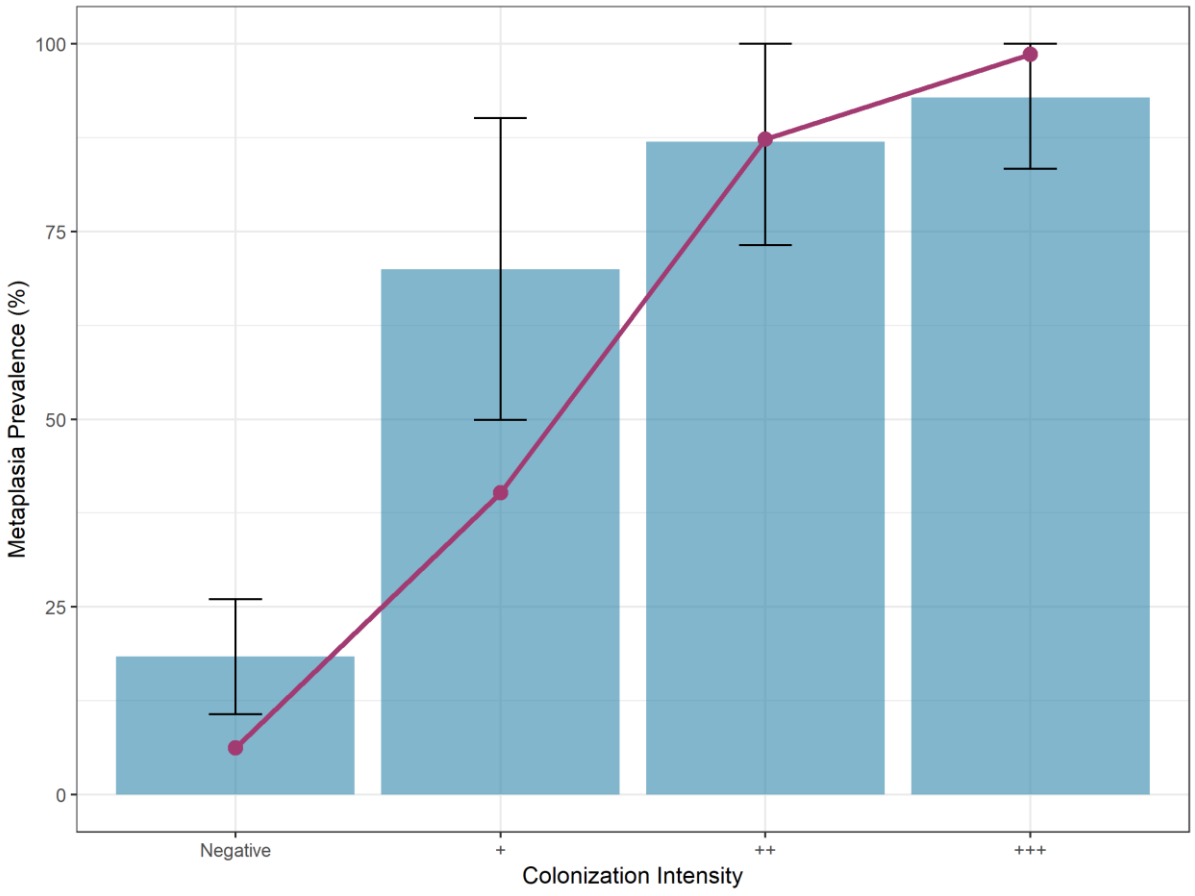


Figure 1. Dose–response relationship between *H. pylori* colonization intensity and intestinal metaplasia prevalence. Metaplasia prevalence increased progressively from 18.4% in *H. pylori*-negative patients to 70.0%, 87.0%, and 92.9% in patients with mild (+), moderate (++), and severe (+++) colonization, respectively. The Cochran–Armitage test for trend: $\chi^2 = 68.67$, $p < 0.001$.

The Cochran–Armitage test for trend confirmed that this progressive increase in metaplasia prevalence across colonization intensity categories was statistically significant ($\chi^2 = 68.67$; $p < 0.001$). This robust dose–response gradient strongly supports a causal

relationship between bacterial colonization burden and the development of intestinal metaplasia, consistent with the chronic inflammatory cascade proposed by Correa [8].

3.5. Multivariate Analysis (Primary Model—Binary Logistic Regression)

To identify independent predictors of intestinal metaplasia while controlling for potential confounders, we constructed a multiple logistic regression model including colonization intensity as the primary exposure variable, along with sex, age, family history of cancer, smoking, and alcohol consumption as covariates. The primary binary logistic regression model included 169 patients with 78 events (metaplasia present), yielding an adequate events-per-variable ratio (EPV = 13) (Table 4, Figure 2).

Table 4. Multivariate binary logistic regression analysis: independent predictors of intestinal metaplasia (n = 169).

Variable	aOR	95% CI	p-Value	Sig.
Colonization intensity (per level)	10.19	4.38–23.71	<0.001	*
Sex (Male)	0.22	0.04–1.20	0.081	
Age (per year)	1.00	0.96–1.05	0.839	
Family history (Yes)	1.18	0.36–3.86	0.778	
Smoking (Yes)	11.70	1.98–69.03	0.007	*
Alcohol consumption (Yes)	6.32	1.74–22.95	0.005	*

aOR: adjusted odds ratio; CI: confidence interval; Sig.: statistical significance. * $p < 0.05$. Model information: outcome = metaplasia (present vs. absent); events = 78; AUC = 0.921; BIC = 157.46; Hosmer–Lemeshow $p = 0.21$.

After adjustment for all covariates, three factors remained independently and significantly associated with intestinal metaplasia. *H. pylori* colonization intensity demonstrated the strongest association, with each incremental increase in colonization level associated with a 10.19-fold increase in the odds of metaplasia (aOR = 10.19; 95% CI: 4.38–23.71; $p < 0.001$). Smoking was independently associated with an 11.70-fold increase in odds (aOR = 11.70; 95% CI: 1.98–69.03; $p = 0.007$). Alcohol consumption conferred a 6.32-fold increase in odds of metaplasia (aOR = 6.32; 95% CI: 1.74–22.95; $p = 0.005$).

Notably, several factors that demonstrated significant bivariate associations lost statistical significance after multivariate adjustment. Male sex, which showed a crude OR of 2.44, was no longer significant in the adjusted model (aOR = 0.22; 95% CI: 0.04–1.20; $p = 0.081$). Age remained non-significant (aOR = 1.00; 95% CI: 0.96–1.05; $p = 0.839$), and family history of cancer also lost significance (aOR = 1.18; 95% CI: 0.36–3.86; $p = 0.778$). This attenuation suggests that the apparent associations observed in bivariate analysis were confounded by the strong effects of colonization intensity and lifestyle factors.

Events per variable (EPV) is a critical metric for assessing the reliability of logistic regression estimates, calculated as the number of outcome events in the smallest category divided by the number of predictors in the model [39]. For the binary primary model, EPV was 13.0 (78 events/6 predictors), exceeding the conventional threshold of 10 required for stable coefficient estimation. However, for the multinomial model analyzing metaplasia subtypes, the EPV differed substantially between categories: 9.5 for complete metaplasia (57 events/six predictors) and 3.5 for incomplete metaplasia (21 events/six predictors).

The low EPV for the incomplete metaplasia category carries important implications for result interpretation. Simulation studies have demonstrated that EPV below 10 can lead to (1) biased regression coefficients, typically with inflation of effect estimates; (2) inflated standard errors and unreliable confidence intervals; (3) failure of the Wald test and likelihood ratio test maintaining nominal Type I error rates; and (4) poor model convergence, particularly in the presence of sparse data patterns [39,40].

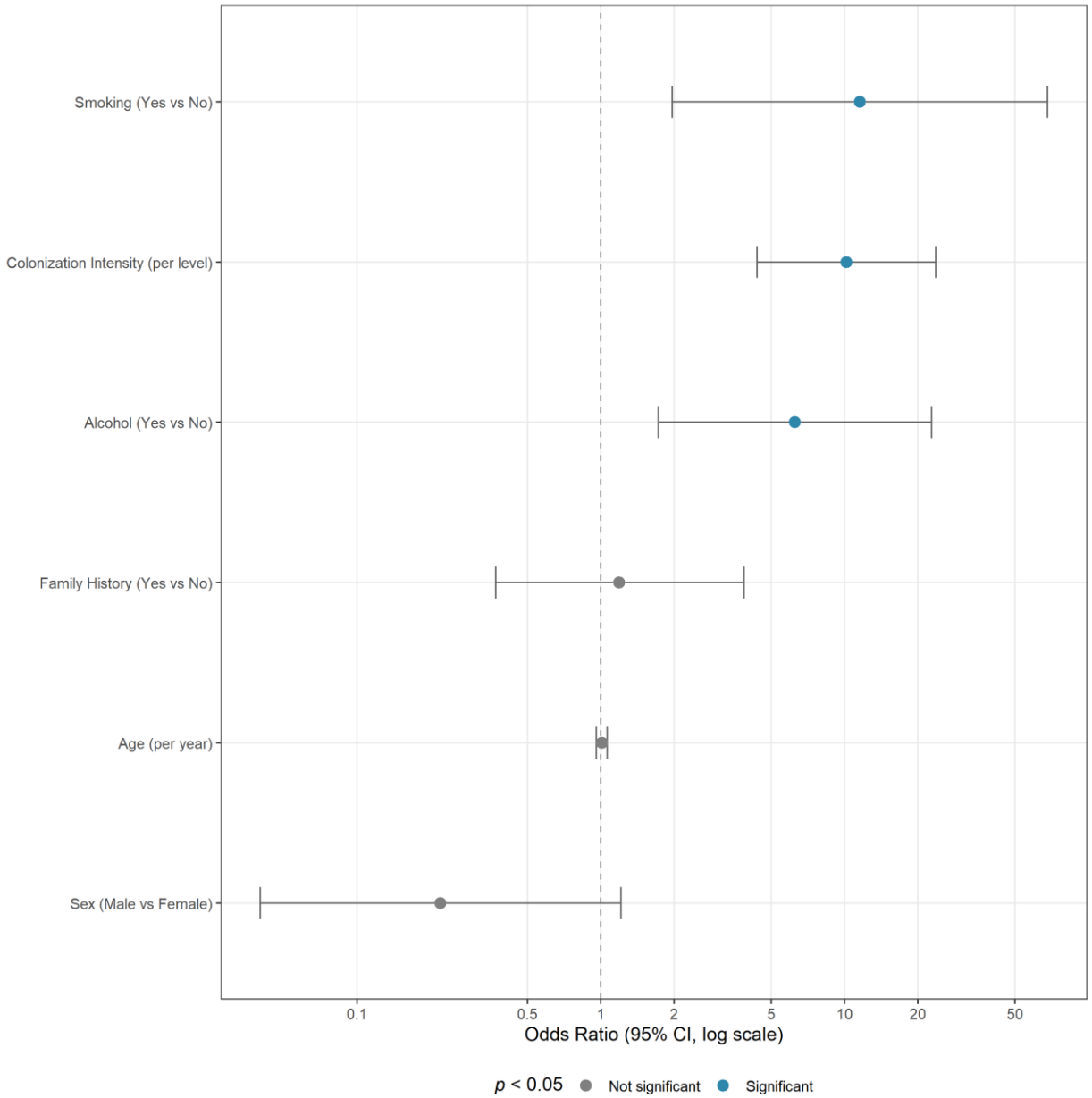


Figure 2. Forest plot depicting adjusted odds ratios and 95% confidence intervals from the primary multivariate binary logistic regression model. Colonization intensity (aOR = 10.19), smoking (aOR = 11.70), and alcohol consumption (aOR = 6.32) emerged as independent predictors of intestinal metaplasia. Sex, age, and family history were not statistically significant after adjustment. Horizontal lines represent 95% confidence intervals; squares represent point estimates, with square size proportional to statistical weight. aOR, adjusted odds ratio; CI, confidence interval.

These limitations are evident in our multinomial results. The confidence interval for alcohol consumption in the complete metaplasia comparison spanned nearly two orders of magnitude (OR = 65.27; 95% CI: 7.28–585.07), reflecting substantial uncertainty in the point estimate. Similarly, the OR for smoking in complete metaplasia (11.80; 95% CI: 1.37–101.85) exhibited wide confidence bounds characteristic of sparse data. These patterns confirm that the multinomial subtype analysis should be interpreted as hypothesis-generating rather than confirmatory, with validation required in larger cohorts before definitive conclusions can be drawn about differential risk factor associations by metaplasia subtype.

Firth's penalized regression was applied as sensitivity analysis to assess the impact of sparse data patterns on model estimates. For the primary binary model, Firth-corrected OR

estimates showed changes ranging from 0% to 23% compared to standard ML estimates, with the largest changes observed for colonization intensity (−21.6%), smoking (−22.2%), and sex (+23.1%). Importantly, the direction and statistical significance of all associations remained unchanged, supporting the robustness of the main findings. For the complete metaplasia subtype, the standard ML model showed evidence of quasi-complete separation (extremely large OR estimates for colonization intensity and alcohol consumption), which was substantially corrected by Firth's penalization. For incomplete metaplasia (EPV = 3.5), the penalized estimates demonstrated greater divergence from ML estimates (up to 41% change for smoking), supporting the interpretation of this subtype analysis as exploratory (Supplementary Table S1).

3.6. Metaplasia Subtype Analysis

To explore differential associations with metaplasia subtypes, a multinomial logistic regression model was constructed with metaplasia subtype (absent/complete/incomplete) as the outcome variable. This analysis should be interpreted as exploratory due to limited events per variable, particularly for the incomplete metaplasia category (EPV = 3.5, below the recommended minimum of 10).

For complete metaplasia versus absent (n = 57 events), colonization intensity demonstrated an exceptionally strong association (OR = 24.55; 95% CI: 8.01–75.27; $p < 0.001$). Alcohol consumption showed the most pronounced effect among lifestyle factors (OR = 65.27; 95% CI: 7.28–585.07; $p < 0.001$), and smoking was also significantly associated (OR = 11.80; 95% CI: 1.37–101.85; $p = 0.025$).

For incomplete metaplasia versus absent (n = 21 events), colonization intensity remained significantly associated but with attenuated magnitude (OR = 6.35; 95% CI: 2.49–16.19; $p < 0.001$). Smoking showed a significant association (OR = 11.10; 95% CI: 1.48–83.23; $p = 0.019$), while alcohol consumption was not significantly associated with incomplete metaplasia (OR = 2.54; 95% CI: 0.57–11.43; $p = 0.224$).

The differential effect of alcohol consumption across metaplasia subtypes is noteworthy: alcohol demonstrated a strong, highly significant association with complete metaplasia (OR = 65.27) but showed no significant association with incomplete metaplasia (OR = 2.54, $p = 0.224$). This divergent pattern suggests potential subtype-specific pathogenic mechanisms that warrant further investigation in adequately powered studies.

3.7. Model Performance and Validation

The primary binary logistic regression model demonstrated excellent discriminative ability, with an area under the receiver operating characteristic curve (AUC) of 0.921 (95% CI: 0.883–0.960), indicating outstanding capacity to distinguish patients with metaplasia from those without (Figure 3). Model calibration was adequate, as assessed by the Hosmer–Lemeshow goodness-of-fit test ($p = 0.21$), indicating no significant departure from observed outcomes.

Additional model fit indices supported the adequacy of the primary model. The McFadden pseudo-R^2 was 0.479 and the Nagelkerke pseudo-R^2 was 0.646, indicating substantial explanatory power. The Brier score of 0.1168 confirmed good calibration and discrimination combined. Model assumptions were verified: the Box–Tidwell test confirmed linearity in the logit ($p > 0.05$), and variance inflation factors (maximum VIF = 3.63) indicated acceptable multicollinearity.

Sensitivity analysis excluding an influential outlier (patient with age 89 years, *H. pylori* negative, no metaplasia) demonstrated model stability, with all coefficient changes below 10% (maximum relative change: 5.6% for the intercept). This confirms that model estimates are not unduly influenced by extreme observations.

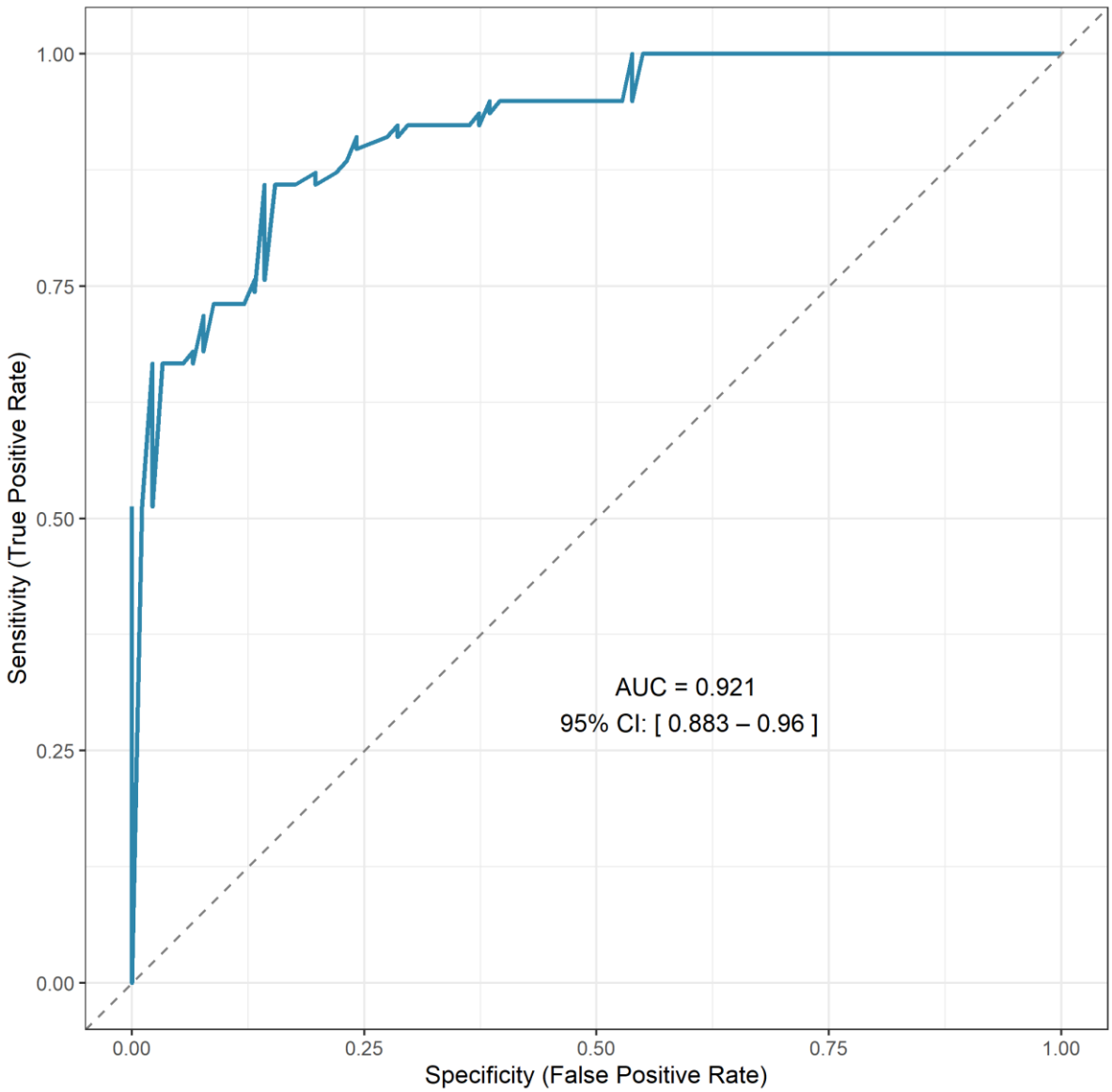


Figure 3. Receiver operating characteristic (ROC) curve for the primary binary logistic regression model predicting intestinal metaplasia. The model demonstrated excellent discriminative ability with an AUC of 0.921 (95% CI: 0.883–0.960). Model calibration was adequate as assessed by the Hosmer–Lemeshow goodness-of-fit test ($p = 0.21$). AUC, area under the curve; CI, confidence interval.

The multinomial model for metaplasia subtype analysis achieved an overall accuracy of 81.7%, with AUC values of 0.955 for complete metaplasia versus absent and 0.771 for incomplete metaplasia versus absent. The lower discriminative performance for incomplete metaplasia is consistent with the limited events in this category and reinforces the exploratory nature of subtype-specific findings.

Statistical power analysis confirmed that the study was adequately powered (calculated power > 99%) to detect the observed effect sizes, with an event rate of 46.2% (78/169) and a Cohen's h effect size of 1.287.

4. Discussion

The results of this study demonstrate a strong and statistically significant association between *H. pylori* infection and intestinal metaplasia in an Andean population from Apurímac, Peru. The primary multivariate logistic regression analysis revealed that colonization intensity was the strongest independent predictor of metaplasia, with an adjusted odds

ratio of 10.19 (95% CI: 4.38–23.71; $p < 0.001$) per incremental grade increase. This indicates that each stepwise elevation in bacterial density—from negative to mild, mild to moderate, or moderate to severe—confers approximately a ten-fold increase in metaplasia risk. The dose–response gradient was particularly striking: metaplasia prevalence escalated from 18.4% among uninfected individuals to 70.0%, 87.0%, and 92.9% in patients with mild (+), moderate (++), and severe (+++) colonization, respectively (Cochran–Armitage trend test $\chi^2 = 68.67$; $p < 0.001$). Furthermore, smoking (aOR = 11.70; 95% CI: 1.98–69.03) and alcohol consumption (aOR = 6.32; 95% CI: 1.74–22.95) emerged as significant independent modifiable risk factors, while demographic variables including sex, age, and family history lost statistical significance following multivariate adjustment, suggesting these associations were substantially confounded by lifestyle exposures.

The magnitude of association observed in our Andean cohort markedly exceeds that documented in most published literature from other geographic regions. The ORs of 24.24 (95% CI: 10.66–55.12) for *H. pylori* positivity are substantially higher than the pooled estimate of OR 3.65 (95% CI: 2.79–4.55) reported in comprehensive meta-analyses encompassing diverse global populations [20], and considerably surpass the OR of 11.5 documented in large Japanese multicenter studies [21]. Our colonization intensity-specific adjusted risk (aOR = 10.19 per grade) also exceeds findings from Middle Eastern and North African studies; Basir et al. reported significant but more modest correlations between bacterial density and histological severity in Iranian patients [13], while Souissi et al. documented associations with inflammation ($p < 0.001$) and atrophy ($p = 0.049$) but noted that metaplasia frequency remained relatively stable at 10–15% regardless of bacterial load [14]. In contrast, our population demonstrated a near-linear escalation from 18% to 93% across colonization grades. This exceptional effect magnitude likely reflects the convergence of multiple risk-amplifying characteristics inherent to Andean populations: ultra-high endemic *H. pylori* prevalence exceeding 80% with infection acquisition during early childhood [6], predominance of highly virulent bacterial genotypes with *cagA* positivity and *vacA* s1m1 alleles documented in 67–89% of highland isolates [22,23], and the well-characterized host–pathogen evolutionary "mismatch" wherein European-lineage (hpEurope) bacterial strains infect hosts with predominantly Amerindian genetic ancestry [16,17]. This phylogeographic discordance has been demonstrated to amplify inflammatory responses and epithelial DNA damage compared to ancestrally matched host–pathogen combinations [18].

We did not assess *H. pylori* virulence factors—including *cagA* genotype, *vacA* allelic variants (s1/s2, m1/m2, i1/i2), and *babA2* adhesin status—which substantially modulate pathogenic potential and may confound the observed colonization–metaplasia relationship [21–23]. This limitation merits particular emphasis given the established mechanistic pathways by which these factors influence gastric carcinogenesis. CagA-positive strains amplify pro-inflammatory responses and induce epithelial DNA damage through sustained activation of NF-κB and β-catenin signaling pathways [16,41]; the *vacA* s1m1 genotype is associated with enhanced cytotoxicity, more severe mucosal inflammation, and accelerated progression through the precancerous cascade compared to less virulent allelic combinations [19]. Critically, these virulence factors may confound our findings through correlation with colonization density: highly virulent strains potentially achieve greater bacterial loads due to enhanced mucosal adherence and immune evasion capacity, meaning that the observed effect of colonization intensity on metaplasia risk could partially reflect unmeasured virulence profiles rather than bacterial burden per se. Prior studies in Peruvian populations have documented high prevalence of virulent genotypes, with *cagA* positivity ranging from 67% to 89% and *vacA* s1m1 alleles predominating among highland isolates [21]. Without virulence factor characterization, we cannot determine whether the dose–response relationship observed in our study is independent of the pathogenic genotype profile, nor can

we assess potential effect modification by virulence status. Future studies incorporating molecular characterization of infecting strains would enhance mechanistic interpretation and clarify whether colonization intensity exerts effects on metaplasia risk independent of virulence factor expression.

The clear dose–response relationship between colonization intensity and metaplasia prevalence constitutes compelling evidence supporting a causal association, fulfilling the biological gradient criterion—one of the most persuasive elements—of the Bradford–Hill framework for causal inference [24]. This gradient is biologically plausible given established pathogenic mechanisms: higher bacterial burden amplifies pro-inflammatory cytokine cascades (IL-1β, IL-8, TNF-α), increases reactive oxygen species generation and oxidative DNA damage, accelerates epithelial cell turnover rates, and enhances delivery of bacterial virulence factors including the CagA oncoprotein through the type IV secretion system [25,26]. Chuang et al. demonstrated that *H. pylori* strains with stronger CagA phosphorylation intensity—a marker of effector protein delivery—conferred substantially elevated metaplasia risk with odds ratios reaching 15.26 [27]. Our findings extend previous observations by demonstrating that the dose–response relationship persists robustly through the metaplasia stage of the Correa cascade, suggesting that bacterial load remains pathogenically consequential even after substantial histological transformation has occurred. This challenges the traditional view that bacterial density diminishes in relevance to metaplastic epithelium, which lacks native *H. pylori* receptors, progressively replacing gastric mucosa [28].

The identification of smoking and alcohol consumption as independent risk factors after adjustment for colonization intensity carries substantial clinical and public health implications. The synergistic interaction between smoking and *H. pylori* infection has been extensively documented in European and Asian populations. Peleteiro et al. demonstrated that the combination of high-virulence bacterial strains and active smoking produced multiplicative effects on incomplete metaplasia risk, with odds ratios reaching 9.81 in Portuguese males [29]. Our finding of aOR = 11.70 for smoking, while characterized by a wide confidence interval (1.98–69.03) reflecting the limited number of smoking events in this cohort, aligns with established pathophysiological mechanisms whereby tobacco constituents compound *H. pylori*-induced oxidative stress, impair mucosal prostaglandin synthesis and bicarbonate secretion, reduce gastric blood flow, and compromise epithelial regenerative capacity [30]. The persistent significance of smoking after adjusting for colonization intensity suggests that tobacco exerts effects on metaplasia risk through pathways at least partially independent of bacterial density modulation.

Alcohol consumption demonstrated particularly intriguing patterns in our exploratory multinomial analysis distinguishing metaplasia subtypes. While alcohol was strongly associated with complete (type I) metaplasia (OR = 65.27; 95% CI: 7.28–585.07; $p < 0.001$), no statistically significant association was observed for incomplete (types II/III) metaplasia (OR = 2.54; 95% CI: 0.57–11.43; $p = 0.224$). This differential effect warrants careful interpretation given the wide confidence intervals and low events per variable for the incomplete subtype (EPV = 3.5) yet remains biologically intriguing. Incomplete metaplasia, characterized by colonic-type mucin expression and substantially higher malignant transformation potential [12,31], may arise through distinct molecular pathways less susceptible to alcohol-mediated promotion. Alternatively, the progression from complete to incomplete metaplasia may involve factors beyond those driving initial metaplastic transformation. These observations merit investigation in larger cohorts adequately powered for subtype-specific analyses.

The attenuation of associations for sex, age, and family history following multivariate adjustment merits mechanistic consideration. In bivariate analysis, male sex (OR = 2.44; 95%

CI: 1.31–4.54), family history of cancer (OR = 2.53; 95% CI: 1.32–4.85), and age demonstrated statistically significant associations with metaplasia presence. However, these variables lost significance ($p > 0.05$) after inclusion of colonization intensity and lifestyle factors in the multivariate model. This pattern suggests that the observed bivariate associations were substantially confounded by or mediated through modifiable exposures. Males in Andean Peruvian communities typically exhibit higher prevalence of both smoking and alcohol consumption compared to females, potentially explaining the sex differential through behavioral rather than purely biological mechanisms. Similarly, family history may capture shared household environmental exposures—including dietary patterns, water sources facilitating *H. pylori* transmission, and intrafamilial lifestyle behaviors—rather than exclusively genetic susceptibility. The attenuation of the family history association may be partially attributable to our variable definition, which captured any malignancy in first-degree relatives rather than gastric cancer specifically. A more targeted assessment of gastric cancer family history might have yielded different results, given established evidence that gastric cancer aggregates in families through both genetic and shared environmental mechanisms [42,43]. The variance inflation factors remained within acceptable limits (maximum VIF = 3.63), confirming that multicollinearity did not compromise the validity of adjusted estimates.

These findings carry significant implications for gastric cancer prevention strategies in high-risk Andean populations. Current Maastricht VI/Florence consensus guidelines recommend population-based *H. pylori* screening and treatment in regions with gastric cancer incidence exceeding 20 per 100,000 person-years [32]. The Apurímac region, characterized by endemic *H. pylori* infection exceeding 80% prevalence, exceptional precursor lesion burden, and mortality patterns placing gastric cancer among leading causes of cancer death, unequivocally meets criteria for high-risk designation. Our results specifically support targeted surveillance, prioritizing individuals demonstrating heavy bacterial colonization on index endoscopy, active smoking history, and regular alcohol consumption. The excellent discriminative capacity of our primary model (AUC = 0.921) suggests potential clinical utility for risk stratification, though external validation would be required before implementation. Furthermore, the modifiable nature of both smoking and alcohol consumption presents opportunities for behavioral interventions as cost-effective adjuncts to pharmacological *H. pylori* eradication. Population-based cost-effectiveness analyses consistently demonstrate that screen-and-treat strategies are economically favorable in endemic settings, with incremental cost-effectiveness ratios well below standard willingness-to-pay thresholds [33,34]. Integration of smoking cessation and alcohol reduction counseling into *H. pylori* eradication programs may enhance preventive impact while leveraging existing healthcare infrastructure.

Several methodological strengths support the internal validity and robustness of our findings. The sample size of 169 patients with 78 metaplasia events provided 100% statistical power for detecting the primary association, eliminating concerns regarding type II error. The primary binary logistic model demonstrated excellent discrimination (AUC = 0.921; 95% CI: 0.883–0.960), indicating strong capacity to distinguish patients with and without metaplasia based on the included predictors. Model calibration was adequate as evidenced by a non-significant Hosmer–Lemeshow goodness-of-fit test ($p = 0.21$), suggesting predicted probabilities aligned well with observed outcomes across risk strata. Events per variable for the primary model (EPV = 13) met conventional adequacy thresholds, supporting reliable coefficient estimation. Sensitivity analysis excluding an identified age outlier (89-year-old patient) confirmed remarkable estimate stability, with maximum coefficient change of only 5.6%—well below the 10% threshold typically considered indicative of influential observations. Furthermore, convergent results across multiple analytical approaches—including

the primary full model, backward BIC-selected model, and parsimonious model retaining only significant predictors—strengthen confidence in the robustness and reproducibility of the identified associations. Multicollinearity diagnostics confirmed acceptable variance inflation factors for all predictors.

Despite these strengths, several limitations should be acknowledged. The cross-sectional design fundamentally precludes causal inference regarding the temporal sequence of colonization intensity escalation and metaplasia development; prospective cohort studies would be required to establish whether high bacterial density precedes and predicts subsequent metaplastic transformation. Hospital-based sampling from a single EsSalud facility may introduce selection bias, as patients referred for upper gastrointestinal endoscopy represent a symptomatic population that differs systematically from the broader Apurímac community. Specifically, this referral-based cohort likely exhibits higher prevalence of both *H. pylori* infection and intestinal metaplasia compared to the general asymptomatic population, potentially inflating our point estimates. Patients undergoing endoscopy for dyspeptic symptoms, alarm features, or surveillance indications typically harbor greater disease burden, more advanced histological changes, and higher cumulative exposure to risk factors than population-based samples would capture. Consequently, our findings regarding the colonization–metaplasia association, while internally valid for risk stratification among symptomatic patients undergoing diagnostic evaluation, may not directly generalize to asymptomatic individuals who might be identified through population-based *H. pylori* screening programs. This limitation is inherent to endoscopy-based studies and consistent with comparable literature examining precancerous gastric lesions [44,45]. The clinical relevance of our findings for secondary prevention strategies—identifying high-risk individuals among those already presenting for evaluation—remains undiminished, though extrapolation to primary screening contexts should be approached with appropriate caution. We did not assess *H. pylori* virulence factors—including *cagA* genotype, *vacA* allelic variants (s1/s2, m1/m2, i1/i2), and babA2 adhesin status—which substantially modulate pathogenic potential and may confound or mediate the observed colonization–metaplasia relationship [21–23]. Given the documented predominance of high-virulence genotypes in Peruvian highland populations, virulence factor assessment would enhance mechanistic interpretation. Furthermore, formal analyses of statistical interactions (for example, between colonization intensity and smoking status, or between colonization and unmeasured virulence profiles) were not performed due to insufficient statistical power; detecting multiplicative interactions typically requires sample sizes substantially larger than our cohort of 169 patients, with approximately 650–700 subjects needed for the observed effect sizes.

Additionally, the distinction between complete and incomplete intestinal metaplasia was based on morphological criteria using H&E staining without Alcian blue pH 2.5 histochemical confirmation. Although Alcian blue staining provides more precise mucin characterization by differentiating sialomucins from sulfomucins, this technique was not available at our institution during the study period. This limitation may have resulted in some degree of misclassification between metaplasia subtypes. However, the primary outcome of our analysis—presence versus absence of intestinal metaplasia—remains unaffected by this limitation, and the morphological criteria employed have been validated in previous studies and are commonly used in clinical practice settings where specialized histochemical staining is unavailable [46].

The family history variable captured any malignancy in first-degree relatives rather than gastric cancer specifically. This broad definition may have attenuated the potential association with intestinal metaplasia, as only gastric cancer family history would be expected to confer increased risk through shared genetic susceptibility (such as *CDH1* germline mutations), common *H. pylori* strain transmission within households, or similar

environmental exposures [42,43]. The lack of cancer type specification precluded analysis of site-specific familial aggregation and may partially explain the loss of statistical significance for this variable in the adjusted model (aOR = 1.18; $p = 0.778$).

Molecular quantification of *H. pylori* density by quantitative PCR was not performed; colonization intensity was assessed by histological grading according to the Updated Sydney System, which remains the internationally accepted standard for this purpose [33,34]. While qPCR would have provided continuous rather than ordinal bacterial load data [47], our histological approach aligns with current diagnostic practice and enables direct comparison with the majority of published studies utilizing Sydney System grading.

Histopathological assessment was performed by a single pathologist, precluding formal evaluation of interobserver reliability. While this represents a methodological limitation, the use of standardized Updated Sydney System criteria with explicitly defined visual analog anchors minimizes subjective variation in grading [33,34]. Furthermore, the pathologist was blinded to clinical findings, reducing potential bias in histological interpretation. Future studies should consider dual independent review with formal kappa agreement assessment.

The single-center design limits generalizability to other Andean communities, coastal Peruvian populations, or highland regions of other Latin American countries, each characterized by distinct genetic, environmental, and microbiological profiles. For the exploratory multinomial regression distinguishing complete from incomplete metaplasia subtypes, events per variable were critically low (EPV = 3.5 for incomplete metaplasia), necessitating designation of these findings as hypothesis-generating rather than confirmatory. The ordinal proportional odds model was invalidated due to violation of the proportional odds assumption, precluding its use for ordered categorical inference; this suggests that predictors exert differential effects across outcome categories rather than uniform effects amenable to ordinal modeling. Finally, residual confounding from unmeasured variables—including dietary patterns (salt intake, fresh vegetable consumption), specific socioeconomic indicators, water source quality, and host genetic polymorphisms (IL-1β, IL-1RN variants)—cannot be excluded. Despite these constraints, the present study provides valuable epidemiological data for an understudied Andean population.

5. Conclusions

This study demonstrates a strong association between *H. pylori* colonization intensity and intestinal metaplasia in a highland Peruvian population, characterized by a clear dose–response relationship. The magnitude of the association observed here exceeds that reported in most comparable studies, a finding that may reflect the convergence of hyperendemic infection, virulent circulating strains, and prolonged mucosal exposure characteristic of this Andean setting. Smoking and alcohol consumption were identified as additional independent risk factors, highlighting the contribution of modifiable behaviors to gastric precancer risk.

From a public health perspective, these findings support the prioritization of *H. pylori* screening and eradication programs in high-risk Andean communities as a strategy for gastric cancer prevention. Future research should evaluate the cost-effectiveness of such interventions, assess the role of bacterial virulence determinants and host genetic factors in mediating disease progression, and employ longitudinal designs to characterize the natural history of preneoplastic lesions in this population.

Additionally, future studies incorporating molecular techniques such as quantitative PCR may provide additional insights into the relationship between precise bacterial load and metaplasia risk, potentially revealing threshold effects or non-linear relationships not captured by semi-quantitative histological grading.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/gastroent17010015/s1, Figure S1: Study flowchart depicting patient selection and final sample composition; Figure S2: Representative histopathological images demonstrating Helicobacter pylori colonization intensity and intestinal metaplasia subtypes in antral biopsies; Figure S3: Sex-stratified analysis of intestinal metaplasia prevalence and risk factor distribution; Table S1: Comparison of maximum likelihood (ML) and Firth's penalized likelihood regression estimates.

Author Contributions: Conceptualization, Y.U. and M.C.M.V.; methodology, Y.U.; formal analysis, J.D.R.; investigation, Y.U.; resources, Y.U.; data curation, Y.U. and J.D.R.; writing—original draft preparation, J.D.R.; writing—review and editing, Y.U., M.C.M.V. and J.D.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Committee of Ethics and Scientific Integrity of Universidad Privada Norbert Wiener (Protocol No. 1324-202, approved on 10 September 2025).

Informed Consent Statement: Patient consent was waived due to the retrospective nature of the study, which utilized de-identified medical records. Patient confidentiality was maintained through assignment of unique study codes, with all personal identifiers removed prior to analysis.

Data Availability Statement: The data presented in this study are available on reasonable request from the corresponding author. The data are not publicly available due to privacy and ethical restrictions governing patient medical records at Hospital II EsSalud Abancay.

Acknowledgments: During the preparation of this manuscript, the authors used Claude AI, Model Opus 4.5 for the purposes of grammar editing and correction. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation	Definition
aOR	Adjusted Odds Ratio
AUC	Area Under the Receiver Operating Characteristic Curve
BIC	Bayesian Information Criterion
CI	Confidence Interval
EPV	Events Per Variable
IQR	Interquartile Range
OR	Odds Ratio
SD	Standard Deviation
VIF	Variance Inflation Factor

References

1. Bray, F.; Laversanne, M.; Sung, H.; Ferlay, J.; Siegel, R.L.; Soerjomataram, I.; Jemal, A. Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2024**, *74*, 229–263. [CrossRef]
2. *Stomach Cancer Statistics*; World Cancer Research Fund: London, UK. Available online: https://www.wcrf.org/preventing-cancer/cancer-statistics/stomach-cancer-statistics/ (accessed on 16 December 2025).
3. International Agency for Research on Cancer. Schistosomes, Liver Flukes and *Helicobacter pylori*. In *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*; International Agency for Research on Cancer: Lyon, France, 1994; Volume 61, pp. 177–241.

4. Morgan, D.R.; Torres, J.; Sexton, R.; Herrero, R.; Salazar-Martínez, E.; Greenberg, E.R.; Bravo, L.E.; Dominguez, R.L.; Ferreccio, C.; Lazcano-Ponce, E.C.; et al. Risk of Recurrent *Helicobacter pylori* Infection 1 Year After Initial Eradication Therapy in 7 Latin American Communities. *JAMA* **2013**, *309*, 578–586. [CrossRef] [PubMed]
5. Santos, E. Current Approaches to Gastric Cancer in Peru and Mexico. *Transl. Gastroenterol. Hepatol.* **2017**, *2*, 55. [CrossRef] [PubMed]
6. Ramírez Ramos, A.; Mendoza Requena, D.; Leey Casella, J.; Guerra Valencia, J. Estudio Del *Helicobacter pylori* En El Perú. *Rev. Peru. Med. Exp. Salud Pública* **2002**, *19*, 209–214.
7. Boehnke, K.F.; Valdivieso, M.; Bussalleu, A.; Sexton, R.; Thompson, K.C.; Osorio, S.; Reyes, I.N.; Crowley, J.J.; Baker, L.H.; Xi, C. Antibiotic Resistance among *Helicobacter pylori* Clinical Isolates in Lima, Peru. *Infect. Drug Resist.* **2017**, *10*, 85–90. [CrossRef]
8. 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*, 6735–6740.
9. Conteduca, V.; Sansonno, D.; Lauletta, G.; Russi, S.; Ingravallo, G.; Dammacco, F.H. *pylori* Infection and Gastric Cancer: State of the Art (Review). *Int. J. Oncol.* **2013**, *42*, 5–18. [CrossRef]
10. Sugano, K.; Moss, S.F.; Kuipers, E.J. Gastric Intestinal Metaplasia: Real Culprit or Innocent Bystander as a Precancerous Condition for Gastric Cancer? *Gastroenterology* **2023**, *165*, 1352–1366.e1. [CrossRef]
11. Song, H.; Ekheden, I.G.; Zheng, Z.; Ericsson, J.; Nyrén, 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. [CrossRef]
12. Mera, R.M.; Bravo, L.E.; Camargo, M.C.; Bravo, J.C.; Delgado, A.G.; Romero-Gallo, J.; Yepez, M.C.; Realpe, J.L.; Schneider, B.G.; Morgan, D.R.; et al. Dynamics of *Helicobacter pylori* Infection as a Determinant of Progression of Gastric Precancerous Lesions: 16-Year Follow-up of an Eradication Trial. *Gut* **2018**, *67*, 1239–1246. [CrossRef] [PubMed]
13. Ghasemi Basir, H.R.; Ghobakhlou, M.; Akbari, P.; Dehghan, A.; Seif Rabiei, M.A. Correlation between the Intensity of *Helicobacter pylori* Colonization and Severity of Gastritis. *Gastroenterol. Res. Pract.* **2017**, *2017*, 8320496. [CrossRef]
14. Souissi, S.; Makni, C.; Belhadj Ammar, L.; Bousnina, O.; Kallel, L. Correlation between the Intensity of *Helicobacter pylori* Colonization and Severity of Gastritis: Results of a Prospective Study. *Helicobacter* **2022**, *27*, e12910. [CrossRef]
15. Ranjbar, R.; Behzadi, P.; Farshad, S. Advances in Diagnosis and Treatment of *Helicobacter pylori* Infection. *Acta Microbiol. Immunol. Hung.* **2017**, *64*, 273–292. [CrossRef]
16. Kusters, J.G.; van Vliet, A.H.M.; Kuipers, E.J. Pathogenesis of *Helicobacter pylori* Infection. *Clin. Microbiol. Rev.* **2006**, *19*, 449–490. [CrossRef] [PubMed]
17. Hatakeyama, M. *Helicobacter pylori* CagA and Gastric Cancer: A Paradigm for Hit-and-Run Carcinogenesis. *Cell Host Microbe* **2014**, *15*, 306–316. [CrossRef] [PubMed]
18. Murata-Kamiya, N.; Kurashima, Y.; Teishikata, Y.; Yamahashi, Y.; Saito, Y.; Higashi, H.; Aburatani, H.; Akiyama, T.; Peek, R.M.; Azuma, T.; et al. *Helicobacter pylori* CagA Interacts with E-Cadherin and Deregulates the β-Catenin Signal That Promotes Intestinal Transdifferentiation in Gastric Epithelial Cells. *Oncogene* **2007**, *26*, 4617–4626. [CrossRef] [PubMed]
19. Figueiredo, C.; Machado, J.C.; Yamaoka, Y. Pathogenesis of *Helicobacter pylori* Infection. *Helicobacter* **2005**, *10*, 14–20. [CrossRef]
20. Cover, T.L.; Blanke, S.R. *Helicobacter pylori* VacA, a Paradigm for Toxin Multifunctionality. *Nat. Rev. Microbiol.* **2005**, *3*, 320–332. [CrossRef]
21. Guzmán, J.; Castillo, D.; González-Siccha, A.D.; Bussalleu, A.; Trespalacios-Rangel, A.A.; Lescano, A.G.; Sauvain, M. *Helicobacter pylori cagA*, *vacA*, *iceA* and *babA* Genotypes from Peruvian Patients with Gastric Intestinal Metaplasia. *Cancers* **2024**, *16*, 1476. [CrossRef]
22. González-Vázquez, R.; Córdova-Espinoza, M.G.; Escamilla-Gutiérrez, A.; Morales-Méndez, I.; Ochoa-Pérez, S.A.; Armendáriz-Toledano, F.; Fonseca-Nájera, J.; Giono-Cerezo, S. Frequency of Virulence Genes in Mixed Infections with *Helicobacter pylori* Strains from a Mexican Population. *Rev. Gastroenterol. México* **2016**, *81*, 11–20. [CrossRef]
23. González, I.; Romero, J.; Rodríguez, B.; Llanos, J.; Morales, E.; Figueroa, H.; Perez-Castro, R.; Valdés, E.; Cofre, C.; Rojas, A. High Prevalence of Virulence-Associated Genotypes in *Helicobacter pylori* Clinical Isolates in the Region Del Maule, Chile. *Scand. J. Infect. Dis.* **2011**, *43*, 652–655. [CrossRef]
24. Hu, Y.; Zhu, Y.; Lu, N.-H. The Management of *Helicobacter pylori* Infection and Prevention and Control of Gastric Cancer in China. *Front. Cell. Infect. Microbiol.* **2022**, *12*, 1049279. [CrossRef]
25. Torres, J.; Correa, P.; Ferreccio, C.; Hernandez-Suarez, G.; Herrero, R.; Cavazza-Porro, M.; Dominguez, R.; Morgan, D. Gastric Cancer Incidence and Mortality Is Associated with Altitude in the Mountainous Regions of Pacific Latin America. *Cancer Causes Control* **2013**, *24*, 249–256. [CrossRef]
26. Kodaman, N.; Pazos, A.; Schneider, B.G.; Piazuelo, M.B.; Mera, R.; Sobota, R.S.; Sicinschi, L.A.; Shaffer, C.L.; Romero-Gallo, J.; de Sablet, T.; et al. Human and *Helicobacter pylori* Coevolution Shapes the Risk of Gastric Disease. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 1455–1460. [CrossRef]

27. de Sablet, T.; Piazuelo, M.B.; Shaffer, C.L.; Schneider, B.G.; Asim, M.; Chaturvedi, R.; Bravo, L.E.; Sicinschi, L.A.; Delgado, A.G.; Mera, R.M.; et al. Phylogeographic Origin of *Helicobacter pylori* Is a Determinant of Gastric Cancer Risk. *Gut* **2011**, *60*, 1189–1195. [CrossRef] [PubMed]
28. Muñoz-Ramírez, Z.Y.; Mendez-Tenorio, A.; Kato, I.; Bravo, M.M.; Rizzato, C.; Thorell, K.; Torres, R.; Aviles-Jimenez, F.; Camorlinga, M.; Canzian, F.; et al. Whole Genome Sequence and Phylogenetic Analysis Show *Helicobacter pylori* Strains from Latin America Have Followed a Unique Evolution Pathway. *Front. Cell. Infect. Microbiol.* **2017**, *7*, 50. [CrossRef]
29. Instituto Nacional de Estadística e Informática (INEI) Perú: Perfil Sociodemográfico. Available online: https://www.gob.pe/institucion/inei/noticias/535452-inei-difunde-base-de-datos-de-los-censos-nacionales-2017-y-el-perfil-sociodemografico-del-peru (accessed on 16 December 2025).
30. Lwanga, S.K.; Lemeshow, S. *Sample Size Determination in Health Studies: A Practical Manual*; World Health Organization: Geneva, Switzerland, 1991.
31. Charan, J.; Biswas, T. How to Calculate Sample Size for Different Study Designs in Medical Research? *Indian J. Psychol. Med.* **2013**, *35*, 121–126. [CrossRef] [PubMed]
32. Pourhoseingholi, M.A.; Vahedi, M.; Rahimzadeh, M. Sample Size Calculation in Medical Studies. *Gastroenterol. Hepatol. Bed Bench* **2013**, *6*, 14–17. [PubMed]
33. Dixon, M.F.; Genta, R.M.; Yardley, J.H.; Correa, P. Classification and Grading of Gastritis: The Updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. *Am. J. Surg. Pathol.* **1996**, *20*, 1161. [CrossRef] [PubMed]
34. Stolte, M.; Meining, A. The Updated Sydney System: Classification and Grading of Gastritis as the Basis of Diagnosis and Treatment. *Can. J. Gastroenterol. Hepatol.* **2001**, *15*, 367832. [CrossRef]
35. Rugge, M.; Meggio, A.; Pennelli, G.; Piscioli, F.; Giacomelli, L.; Pretis, G.D.; Graham, D.Y. Gastritis Staging in Clinical Practice: The OLGA Staging System. *Gut* **2007**, *56*, 631–636. [CrossRef]
36. El-Zimaity, H.M.T.; Graham, D.Y. Evaluation of Gastric Mucosal Biopsy Site and Number for Identification of *Helicobacter pylori* or Intestinal Metaplasia: Role of the Sydney System. *Hum. Pathol.* **1999**, *30*, 72–77. [CrossRef]
37. Filipe, M.I.; Muñoz, N.; Matko, I.; Kato, I.; Pompe-Kirn, V.; Jutersek, A.; Teuchmann, S.; Benz, M.; Prijon, T. Intestinal Metaplasia Types and the Risk of Gastric Cancer: A Cohort Study in Slovenia. *Int. J. Cancer* **1994**, *57*, 324–329. [CrossRef] [PubMed]
38. Heinze, G.; Puhr, R. Bias-Reduced and Separation-Proof Conditional Logistic Regression with Small or Sparse Data Sets. *Stat. Med.* **2010**, *29*, 770–777. [CrossRef] [PubMed]
39. Peduzzi, P.; Concato, J.; Kemper, E.; Holford, T.R.; Feinstein, A.R. A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis. *J. Clin. Epidemiol.* **1996**, *49*, 1373–1379. [CrossRef]
40. Vittinghoff, E.; McCulloch, C.E. Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression. *Am. J. Epidemiol.* **2007**, *165*, 710–718. [CrossRef]
41. Abadi, A.T.B.; Ierardi, E.; Lee, Y.Y. Why Do We Still Have *Helicobacter pylori* in Our Stomachs. *Malays. J. Med. Sci. MJMS* **2015**, *22*, 70–75. [PubMed]
42. Yaghoobi, M.; Bijarchi, R.; Narod, S.A. Family History and the Risk of Gastric Cancer. *Br. J. Cancer* **2010**, *102*, 237–242. [CrossRef]
43. Weyermann, M.; Rothenbacher, D.; Brenner, H. Acquisition of *Helicobacter pylori* Infection in Early Childhood: Independent Contributions of Infected Mothers, Fathers, and Siblings. *Off. J. Am. Coll. Gastroenterol. ACG* **2009**, *104*, 182. [CrossRef]
44. Saidani, F.; Khayat, L.; Elleuch, N.; Jaziri, H.; Wafa, D.; Slama, A.B.; Ajmi, S.; Braham, A.; Hammami, A.; Ksiaa, M. Prevalence and Predictive Factors of Gastric Atrophy and Intestinal Metaplasia in *Helicobacter pylori*-Associated Gastritis. *Endoscopy* **2024**, *56*, eP459. [CrossRef]
45. Asaka, M.; Sugiyama, T.; Nobuta, A.; Kato, M.; Takeda, H.; Graham, D.Y. Atrophic Gastritis and Intestinal Metaplasia in Japan: Results of a Large Multicenter Study. *Helicobacter* **2001**, *6*, 294–299. [CrossRef] [PubMed]
46. Jass, J.R.; Filipe, M.I. The Mucin Profiles of Normal Gastric Mucosa, Intestinal Metaplasia and Its Variants and Gastric Carcinoma. *Histochem. J.* **1981**, *13*, 931–939. [CrossRef] [PubMed]
47. Lee, J.Y.; Kim, N. Diagnosis of *Helicobacter pylori* by Invasive Test: Histology. *Ann. Transl. Med.* **2015**, *3*, 10. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.