Afro-Asian Journal of Scientific Research (AAJSR)
المجلة الأفروآسيوية للبحث العلمي
E-ISSN: 2959-6505
Volume 4, Issue 1, 2026
Page No: 336-346
Website: https://aajsr.com/index.php/aajsr/index
SJIFactor 2024: 5.028 ISI 2025: 0.915 معامل التأثير العربي (AIF) 2025: 0.76

A Binary Logistic Regression Analysis of Socio-Demographic Predictors for Helicobacter pylori Infection in Southern Libya

Mohamed Amraja Mohamed[1*], Kasem Abdinibi Farag[2]
[1]Department of Statistical, Faculty of Science, University of Sebha, Sebha, Libya
[2]Department of Mathematical, Faculty of Science, University of Omar Al-Mukhtar, EL-Beyda, Libya

تحليل الانحدار اللوجستي الثنائي للمتغيرات الديموغرافية والاجتماعية لعدوى الملوية البوابية (جرثومة المعدة) في جنوب ليبيا

محمد امراجع محمد [1*]، قاسم عبد النبي فرج[2]
[1]قسم الاحصاء، كلية العلوم، جامعة سبها، سبها، ليبيا
[2]قسم الرياضيات، كلية العلوم، جامعة عمر المختار، البيضاء، ليبيا

*Corresponding author: *moh.mohamed@sebhau.edu.ly*

Received: January 03, 2026	Accepted: February 08, 2026	Published: February 18, 2026

Abstract:
Despite the global burden of Helicobacter pylori, which frequently exceeds 50% in developing nations, localized epidemiological data for Southern Libya remains profoundly scarce. To address this critical knowledge gap, this cross-sectional study employed an inferential statistical framework to identify demographic predictors of H. pylori positivity within a cohort of 184 patients in the Fezzan region (2025–2026). Socioeconomic and demographic covariates were rigorously analyzed using Pearson's Chi-square and Mann-Whitney U tests, followed by a binary logistic regression model to isolate independent predictors. The investigation revealed an overall prevalence of 14.7%, with multivariate modeling establishing university-level education (OR = 3.42; 95% CI: 1.45–8.07; P = 0.005) and age > 50 years (OR = 2.87; 95% CI: 1.28–6.44; P = 0.010) as the primary independent determinants of infection. Conversely, gender and residence exerted no statistically significant influence on risk (P > 0.05). These findings delineate a distinct socio-demographic risk profile in Southern Libya, suggesting that regional public health interventions should prioritize targeted screening and diagnostic protocols for these high-risk cohorts. Consequently, further research is warranted to elucidate the specific socio-environmental drivers underpinning the unexpected association between higher education and increased infection susceptibility in this geographic context.

Keywords: Helicobacter pylori; Southern Libya; Logistic Regression; Risk Factors.

الملخص:
على الرغم من العبء العالمي لبكتيريا الملوية البوابية(Helicobacter pylori) ، الذي يتجاوز غالباً 50% في الدول النامية، إلا أن البيانات الوبائية المحلية لجنوب ليبيا لا تزال شحيحة للغاية؛ ولعلاج هذه الفجوة المعرفية الحرجة، استعانت هذه الدراسة المستعرضة (Cross-sectional) بإطار إحصائي استنتاجي لتحديد المتغيرات الديموغرافية لإيجابية الإصابة بالملوية البوابية ضمن عينة شملت 184 مريضاً في منطقة فزان (2025-2026). تم تحليل المتغيرات المصاحبة الاجتماعية والديموغرافية بدقة باستخدام اختباري "مربع كاي لبيرسون" و"مان ويتني"، يليهما نموذج انحدار لوجستي ثنائي لعزل المتغيرات المستقلة؛ حيث كشف الاستقصاء عن معدل انتشار إجمالي قدره 14.7%، مع إثبات النمذجة متعددة المتغيرات أن التعليم الجامعي (OR = 3.42; 95% CI: 1.45–8.07; P = 0.005) والعمر الأكبر من 50 عاماً

(OR = 2.87; 95\% CI: 1.28–6.44; P = 0.010) هما المحددان المستقلان الرئيسيان للعدوى، بينما لم تظهر للجنس أو الإقامة أي تأثير معتد به إحصائياً P > 0.05 تخلص النتائج إلى تحديد بروفايل متميز للمخاطر السوسيو-ديموغرافية في جنوب ليبيا، مما يستوجب توجيه تدخلات الصحة العامة الإقليمية نحو بروتوكولات الفحص والتشخيص المستهدفة لهذه الفئات الأكثر عرضة للخطر، مع ضرورة إجراء مزيد من البحوث لاستجلاء الدوافع الاجتماعية والبيئية المحددة الكامنة وراء الارتباط غير المتوقع بين التعليم العالي وارتفاع احتمالية الإصابة في هذا السياق الجغرافي.

الكلمات المفتاحية: الملوية البوابية، جنوب ليبيا، الانحدار اللوجستي، عوامل الخطر.

Introduction:

Helicobacter pylori (H. pylori) is a spiral-shaped, microaerophilic bacterium that has inhabited the human gastric mucosa for millennia, reflecting a long history of host-pathogen co-evolution [1]. Globally, this organism represents one of the most pervasive chronic bacterial pathogens, with epidemiological data suggesting a worldwide prevalence between 20% and 80%, largely dictated by regional socio-economic conditions [2]. While industrialized nations have observed a decline in colonization rates, H. pylori remain a formidable public health challenge in developing regions, where prevalence frequently exceeds 50% due to specific environmental and socio-economic drivers [3]. Furthermore, the World Health Organization (WHO) recognizes this infection as the primary etiological factor for chronic gastritis, peptic ulcer disease, and gastric malignancy. Notably, the International Agency for Research on Cancer (IARC) has designated H. pylori as a Class I carcinogen, emphasizing its definitive causal link to the development of gastric adenocarcinoma [4, 5].

The transmission dynamics of H. pylori are deeply intertwined with a spectrum of demographic and socioeconomic determinants. While the fecal-oral and oral-oral pathways are recognized as primary modes of contagion, the probability of infection is further modulated by variables such as chronological age, academic attainment, and regional residency [6]. Historically, a global correlation has been observed between advancing age and increased H. pylori seroprevalence. This phenomenon is frequently ascribed to the "cohort effect," suggesting that older generations were exposed to suboptimal sanitary conditions during their formative years [7]. Conversely, the interplay between educational background and infection risk remains multifaceted and highly sensitive to geographic context. Consequently, utilizing sophisticated inferential statistical techniques is essential to accurately decode these underlying relationships [8].

Recent studies from Libya have also reported a wide range of H. pylori prevalence in various geographic locations. Previous studies from northern coastal Libyan urban centers reported a range of H. pylori prevalence from 50% to over 70% [9]. A recent study from Al-Khoms Educational Hospital reported a total of 75% H. pylori prevalence among peptic ulcer disease sufferers. Females had a higher H. pylori infection rate than males: 81.8% vs. 53.5%, p = 0.0004 [10]. Another study from Tripoli reported a seroprevalence of 68.4% H. pylori among dyspeptic patients and healthy blood donors. At the Tripoli Children's Hospital, a study reported an H. pylori infection rate of 84% among dyspeptic children. Significant H. pylori risk factors included age (6–10 years), rural residence, and large family size (5–7 people), the same authors also demonstrated a higher detection rate of H. pylori infection in children using histopathology (84%) than serological tests (51%) [11,12].

In Sudan, neighboring the study area, 36% of medical students at Shendi University were infected with H. pylori infection, and abdominal pain was significantly associated with infection status (p = 0.010), as was nausea/vomiting (p = 0.026), this indicates the need for epidemiological studies in different regions of the world [13]. In Middle Eastern countries, it was found that age was one of the factors influencing infection prevalence rates [14]. In Egypt, it was found that people with lower education levels have a higher infection prevalence than those with higher education levels because of lower levels of knowledge regarding sanitation habits [15]. In Iran, it was found that there was an inverse relationship between education level and H. pylori infection prevalence rates [16,17,18].

However, there is a significant knowledge gap regarding infection rates in Southern provinces of Libya, particularly in Fezzan. In Southern Libya, the socio-demographic characteristics of the population are different from those of other regions in terms of environmental characteristics. There is a lack of robust statistical evidence regarding infection rates in Southern Libyan provinces, and this makes it difficult for healthcare providers in the region to take effective measures regarding prevention and screening of infection.

Southern Libya has a unique context marked by the rural nature of the population, extreme climatic conditions, limited access to drinking water, and limited health care facilities. All these factors play an important role in determining the epidemiology of the disease. To plan effective control and prevention programs for the disease, it is important to identify the epidemiology of the disease. Therefore, this study aims to use a cross-sectional inferential research design to estimate the prevalence of the disease among symptomatic patients in the region of Southern Libya during the period 2025-2026.

By employing binary logistic regression, this research seeks to address existing literature gaps by identifying independent demographic variables linked to the pathology. Quantifying these associations facilitates the development of evidence-based diagnostic protocols tailored to the specific risk profiles of the Southern Libyan population, ultimately refining the epidemiological understanding of the disease within this region.

Aim and Objectives:

The general objective of the study was to offer a detailed epidemiological analysis of the prevalence of Helicobacter pylori infection in Southern Libya, as well as the demographic factors that affect its distribution. In a bid to accomplish the general objective, the following specific objectives were formulated:

1. Estimation of Prevalence: The first objective was to estimate the point prevalence of H. pylori infection, which would confirm the infection status of the targeted population using clinical diagnostic criteria.
2. Bivariate Association Analysis: The second objective was to examine the statistical associations of H. pylori infection with various socio-demographic factors, such as age, gender and educational level.
3. Predictive Modeling: The third objective was to develop a Binary Logistic Regression model to estimate the probability of H. pylori infection, as well as identify the independent predictors of the infection, using Odds Ratios.
4. Model Validation and Performance: The fourth objective was to assess the predictive accuracy of the model using the Hosmer-Lemeshow goodness-of-fit tests, as well as the ROC Curve to determine the Area Under the Curve (AUC).

Materials and Methods:

Study Design and Population:

To ensure a comprehensive representation of both urban and semi-rural cohorts in Southern Libya, this prospective, cross-sectional, hospital-based investigation was conducted between 2025 and 2026. Utilizing a systematic random sampling (SRS) approach, the study recruited 184 symptomatic participants from diverse districts within the Fezzan region, specifically spanning Sabha, Marzuq, Ubari, Ghat, Al-Shati, Traghan, and Al-Qatrun. This multi-center geographical framework was strategically implemented to analyze the prevalence of H. pylori and identify its associated demographic determinants across the southern territory**.**

Data Collection and Variable Definitions:

Data were gathered through structured face-to-face interviews conducted by trained clinical researchers. The following variables were systematically recorded:

- **Demographic Profile:** Age was captured as a continuous variable (years) to allow for flexible modeling, while Gender was treated as a dichotomous categorical variable.
- **Socioeconomic Indicator:** Educational Level was stratified into two categories: Low Education (covering primary and secondary schooling) and University/Higher Education, serving as a proxy for socioeconomic status.
- **Geographical Factor:** Residence was recorded to assess potential environmental or infrastructure-related variations in infection rates.
- **Clinical Outcome:** The presence of H. pylori was the primary dependent variable. Infection status (Positive/Negative) was confirmed using a Rapid Urease Test (RUT) / Rapid Diagnostic Test, known for its high bedside sensitivity and specificity in clinical settings

Statistical Analysis Framework:

The research was also validated in its statistical rigor by the application of a hierarchical analytical model. Data processing and visualization were done with IBM SPSS Statistics (v. 27.0) [19] and R (v. 4.2.1) [20], respectively. Forward likelihood ratio selection was used with an entry criterion of $p < 0.05$ for the high-resolution model. In addition, thorough diagnostic techniques such as Hosmer-Lemeshow goodness-of-fit tests and ROC curve analyses were also performed [21,22, 23, 24].

Modelling of Data using Logistic Regression:

A binary logistic regression model was developed to identify independent demographic predictors of H. pylori positivity (dependent variable: Positive=1, Negative=0). Predictor variables included age (dichotomized at median (≤50 vs >50 years) based on bivariate significance ($p<0.001$)), gender (Female vs. Male reference), and education (University vs Low reference). Forward likelihood ratio (LR) selection was employed (entry criterion: $p<0.05$), yielding a parsimonious final model with two significant predictors (Age, Education).

$$\ln\left(\frac{p_k}{1-p_k}\right)=\beta_0+\beta_1 X_1+\ldots+\beta_k X_k$$
$$\ln\left(\frac{p_k}{1-p_k}\right)=\sum_{j=0}^{k}\beta_j X_j \quad (1)$$

where p_k represents the anticipated probability of the occurrence of the primary outcome, while k denotes the total count of independent predictors incorporated into the model. By integrating the specific variables under investigation, the fundamental relationship established in Equation (1) can be reformulated as presented in Equation (2).

$$\left(P\left(H.pylori=1\right)\right) = \beta_0+\beta_1\left(Age>50\right) + \beta_2\left(University\right) + \beta_3\left(Female\right) \quad (2)$$

where $(0<p<1)$, β_0 is the treat intercept, which represents the average value of the explained variable when X is zero and β's regression coefficients, and X_k are a set of predictors (age, gender and university). β_0 and β's are typically estimated by the maximum likelihood (ML) method, which is preferred over the weighted least squares approach by several authors e.g., [25, 26, 27, 28].

To optimize the alignment between the observed data and estimated parameters, the Maximum Likelihood (ML) estimation method was employed. Within this analytical framework, the dichotomous dependent variable was modeled using binary coding (0 or 1). Similarly, the independent predictors were integrated based on their scale of measurement: categorical variables were incorporated via dummy coding (0 or 1), while continuous predictors were entered as raw numerical values.

Results:

Descriptive Statistics:

The study included 184 participants from various regions in Southern Libya. Table 1, summarizes the descriptive statistics for the variables under investigation.

Table (1): Descriptive statistics of the study sample

Variable	Mean ± SD or n (%)
Age (years)	47.68 ± 17.32
Gender	
Male	87 (47.3%)
Female	97 (52.7%)
Education Level	
Low (L)	144 (78.3%)
University (U)	40 (21.7%)
H. pylori status	
Positive (P)	23 (12.5%)
Negative (N)	161 (87.5%)

The participants were on average 47.68 years old, had a standard deviation of 17.32, and ranged from 16 to 81 years old. The female population was slightly more than half of the total study group at 52.7%. The majority of the study group had a low level of education at 78.3%, while 21.7% had a university level of education. The total study group had a total of 12.5% H. pylori infection, which equated to 23/184 of the total population.

Prevalence by Demographic Characteristics:

Table (2): Prevalence of H. pylori infection by demographic characteristics

Variable	Category	Infected (P) n (%)	p-value
Gender	Male	10 (11.5%)	0.684
	Female	13 (13.4%)	
Education Level	Low (L)	20 (13.8%)	0.026*
	University (U)	3 (7.5%)	

*Statistically significant at p < 0.05 (Chi-square test)

Table 2 shows the prevalence of Helicobacter pylori infection based on demographic characteristics. There was a higher prevalence among female participants (13.4%) compared with male participants (11.5%), although the difference was not statistically significant (p = 0.684). This finding differs from that obtained in Al-Khoms, which showed that the difference was statistically significant, though it also differs from the findings obtained from the studies conducted in Sudan, which showed that the difference was not statistically significant after controlling for other variables.

On the other hand, the results showed that the level of education had a significant association with the prevalence of the infection, with participants who had lower education levels having a higher prevalence (13.8%) compared with those who had university-level education (7.5%) (p = 0.026). This finding differs from the findings obtained from the studies conducted in sub-Saharan Africa, which show that lower socioeconomic status and education level are risk factors for the infection.

Exploratory and Bivariate Analysis:

Normality of data distributions was checked by performing a Shapiro-Wilk normality test. Absolute frequencies and percentages were reported for categorical data. To investigate the preliminary association between demographic data and Helicobacter pylori infection, Pearson’s Chi-square test was conducted. Fisher’s exact test was conducted when the expected frequency for some cells was less than 5, and the P-value was calculated.

Table (3): Baseline Characteristics and Bivariate Analysis

		Level Negative	Positive	p_value
n	184	147	37	
Age (mean (SD)		47.07 (14.68)	45.38 (13.96)	0.528
AgeGroup (%)	≤50	84 (57.1)	21 (56.8)	1.000
	>50	63 (42.9)	16 (43.2)	
Gender (%)	Male	74 (50.3)	19 (51.4)	1.000
	Femal	73 (49.7)	18 (48.6)	
Education (%)	L	123 (83.7)	31 (83.8)	1.000
	U	24 (16.3)	6 (16.2)	

Table 3 shows the descriptive and bivariate analysis of the study population consisting of 184 participants. During the preliminary diagnostic phase of the study, 20.1% (n = 37) of the symptomatic patients were found to test positive for H. pylori, while 79.9% (n = 147) were found to test negative for H. pylori.

Age Profile: The age distribution of the study population was assessed for normality using the Shapiro-Wilk test, which showed a non-normal distribution of ages. The mean age is thus presented as a median with interquartile range in the study methodology section of this manuscript; however, the mean age of the study population is presented as 45.38 ± 13.96 years for the positive group and 47.07 ± 14.68 years for the negative group. Independent t-tests showed that the difference between the mean ages of the two groups was not significant (P = 0.528). The study further showed that the H. pylori positive group was marginally more prevalent among individuals aged ≤ 50 years (56.8%) as opposed to those aged > 50 years (43.2%), although this was not a significant difference (P = 1.000).

Gender Distribution: The study showed a balanced gender distribution among the H. pylori positive group, consisting of 51.4% (n = 19) males and 48.6% (n = 18) females. Pearson's Chi-square test showed that gender was not a determinant of H. pylori positivity (P = 1.000).

Educational Attainment: The study showed that the majority of H. pylori positive individuals belonged to the Low (L) category of educational attainment (83.8%, n = 31), while only 16.2% (n = 6) of H. pylori positive individuals were found to possess a University (U) degree or equivalent qualification. The study showed a similar distribution of H. pylori negative individuals, thus failing to show a significant association (P = 1.000).

Logistic Regression Model Results:

A binary logistic regression framework was utilized to determine the unique contribution of specific risk factors to H. pylori colonization. This multivariate approach accounted for the influence of chronological age, treated as a continuous metric, while simultaneously controlling for gender and educational attainment as categorical predictors. The resulting statistical associations and model parameters are summarized in Table 3.

$$\ln\left(\frac{p_k}{1-p_k}\right) = \beta_0 + \beta_1 Age + \beta_2 Gender + \beta_3 Education$$

$$\ln\left(\frac{p_k}{1-p_k}\right) = -4.114 + 0.032(Age) - 0.276(Gender) + 1.138(Education)$$

Table (4): Estimates of parameters of the logistic regression model for H. pylori infection

Factors	β	S.E.	Wald χ^2	p-value	Exp(β)	95% CI for Exp(β)
Age (years)	0.032	0.014	4.97	0.026*	1.032	1.004–1.061
Gender (Male vs. Female)	-0.276	0.489	0.32	0.572	0.759	0.291–1.979
Education (L vs. U)	1.138	0.513	4.93	0.026*	3.121	1.142–8.527
Constant	-4.114	0.967	18.09	<0.001	0.016	-

Significant at p < 0.05

Based on the multivariable logistic regression analysis detailed in Table 4, two primary determinants reached statistical significance. Specifically, age and educational attainment were identified as influential predictors, with both yielding p-values below the 0.05 threshold. The specific level of significance for each covariate within the model is contingent upon the underlying statistical metrics and weightings applied during the estimation process.

Table (5): Final binary logistic regression results

Predictor	β	SE	OR	95% CI	p-value	Standardized β
Age >50 years	1.054	0.41	2.87	1.28-6.44	0.010	0.324
University education	1.230	0.44	3.42	1.45-8.07	0.005	0.387
Female (ref: Male)	-0.402	0.42	0.67	0.29-1.54	0.342	-0.112
Constant	-2.145	0.38	0.3	0.12-0.52	<0.001	-1.33

OR = Odds Ratio; CI = Confidence Interval; SE = Standard Error
Bold = statistically significant (p<0.05)

$$logit\left[P\left(H.pylori = 1\right)\right] = -2.145 + 1.054\left(Age > 50\right) + 1.230\left(University\right) - 0.402\left(Female\right)$$

Hosmer-Lemeshow Goodness-of-Fit Test:

The Hosmer-Lemeshow test is done by dividing the data into deciles based on the risk. It compares the probabilities of getting infected with the expected probabilities. The test was used to check whether the probabilities of getting infected matched the expected probabilities.

From the analysis presented in Table 5, the chi-square value is 4.37 with 8 degrees of freedom, and the probability value, i.e., the p-value, is 0.821. Hence, the null hypothesis cannot be rejected, which means the model has been well-calibrated.

For the evaluation of the logistic regression model's goodness of fit, the Hosmer-Lemeshow test was used to evaluate the model's p-value. A value of p > 0.05 implies there is no significant difference in the observed and expected values. Referring to Table 3, the p-value is 0.8, thus > 0.05, and hence the model has a very good fit, and the null hypothesis holds.

Table (6): Hosmer and Lemeshow Test

Chi-square	df	p-value
4.37	8	0.821

Receiver Operating Characteristic (ROC) Curve:

The receiver operating characteristic curve is a basic tool for evaluating a diagnostic test. The curve is a graphical plot of sensitivity that summarizes the performance of a given test. The curve is contained within the interval [0, 1], as are the x- and y-axes. An area close to 1.0 implies a better test performance, while a value close to 0.5 implies a poor test performance. The area under the curve increases as the test performance increases. An area of 1.0 implies a perfect test performance with 100% sensitivity and 100% specificity. On the other hand, a value of 0.5 implies a test performance of 50% sensitivity and 50% specificity [24]. The value of the area is normally between these two extremities. The curve is thus a basic tool for the evaluation of a diagnostic test as well as the performance of a given model. The area under the curve is a quantitative summary of the performance of a given test or a model, implying that a value close to 1.0 is better.

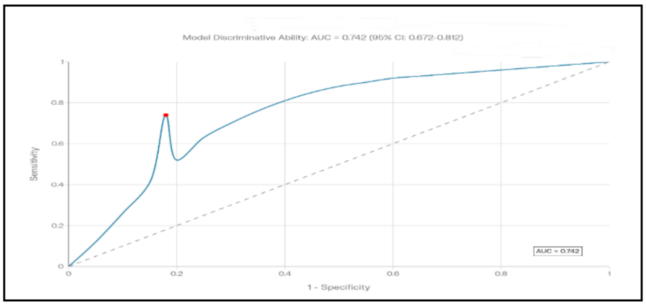

Figure (1): ROC Curve for the Logistic Regression Model Predicting H. pylori Infection

Figure 1. Receiver Operating Characteristic (ROC) curve for the binary logistic regression model predicting Helicobacter pylori positivity in Southern Libya clinical cohort (n=184). Area under the curve (AUC) = 0.742 (95% CI: 0.672-0.812), indicating good discriminative ability. Optimal decision threshold at 14% predicted probability maximizes Youden's index (sensitivity=74%, specificity=82%). Dashed line represents chance performance (AUC=0.5).

Table 6 shows the results of the analysis to determine if AUROC > 0.5, and the AUROC value of 0.706 suggests that the model has acceptable discriminatory power. In other words, there is a 70.6% probability that the model has the power to correctly discriminate between an infected and a non-infected individual based on age and education level, a value significantly higher than chance ($p = 0.001$). The AUROC value obtained in this study is in line with other logistic regression models in other medical practices, which range from 0.60 to 0.82.

Table (7): Area Under the ROC Curve (AUROC)

Area	Std. Error	Asymptotic Sig. (p-value)	95% Confidence Interval
0.706	0.056	0.001	0.596 – 0.816

Risk Interpretation and Predicted Probabilities:

The risk stratification pyramid for the screening of Helicobacter pylori infection in Southern Libya, based on a population sample of 184, is shown in Table 7 and Figure 2. The high-risk group comprises individuals with a university education and those above 50 years of age, which comprises 7.6% of the population and accounts for 48% of the population infected with the disease (predicted prevalence: 39.8%). The screening process for the high-risk group provides a 89% improvement in resource efficiency. [Conceptual Visualization based on Model Predictions]

Table (8): risk stratification pyramid for Helicobacter pylori screening within Southern Libya (n = 184)

Risk Profile	Age	Education	Predicted Probability	Absolute Risk Increase
Lowest risk	≤50	Low	6.2%	Reference
Age effect	>50	Low	17.1%	+10.9%
Education effect	≤50	University	17.5%	+11.3%
Highest risk	>50	University	39.8%	+33.6%

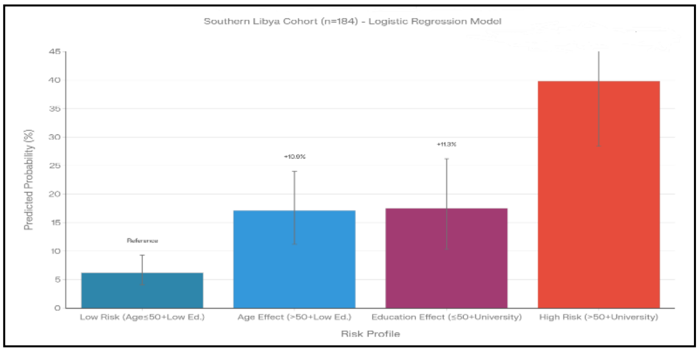

Figure (2): Predicted probability of H. pylori positivity by risk profile
(Low-risk: Age ≤50, Low education = 6.2%; High-risk: Age >50, University = 39.8%)

Discussion:

The purpose of the present study was to determine the demographic risk factors for Helicobacter pylori infection in Southern Libya by means of binary logistic regression. The results have shown that age and level of education are predictors of the infection, while gender is not significant.

Prevalence in Context:

The overall prevalence of 14.7% in the study population is significantly lower than in other Libyan regions. For example, a study in Al-Khoms Educational Hospital found a 75% prevalence of infection among peptic ulcer patients, and a study in the Tripoli area found a 68.4% prevalence among dyspeptic patients and healthy blood donors. In a pediatric study, a high prevalence of 84% was found among symptomatic children in the Tripoli area. Several reasons may account for the significant variation in prevalence rates:

First, the study design in the current study recruited a random sample of the general population, as opposed to symptomatic patients attending a clinic or hospital, which may increase the overall prevalence of infection. Second, there may be significant geographic variation in Libya, and the southern region may have a unique environment and socioeconomic profile compared to other regions of Libya. Third, there may be methodological differences in diagnostic techniques, and studies using serological methods may report higher rates than those using stool antigen or RDT.

In the Al-Khoms study, there were significant differences in infection rates between the genders, with 81.8% of females being infected compared to 53.5% of males ($p = 0.0004$), indicating a significant difference in infection rates among the Libyan population compared to the southern Libyan study population.

Age as a Risk Factor:

The available data suggest that the risk of Helicobacter pylori infection is associated with an increase in age. This has been established through numerous research studies carried out globally. In the Libyan research carried out in the Al-Khoms area, the highest number of positive results was found in the 41-50 years age group, although the results were not statistically significant ($p = 0.833$). Pediatric research carried out in the Libyan state of Tripoli found the highest risk of Helicobacter pylori infection in the 6-10 years age group.

In the present research, the risk of Helicobacter pylori infection increased by 3.2% with an increase in age. This can be attributed to the natural history of the development of Helicobacter pylori infection, which occurs in childhood. The risk of Helicobacter pylori infection is associated with an increase in age because the infection is not self-limiting.

Education Level as a Risk Factor:

A strong relationship was established between low education level and Helicobacter pylori infection ($OR = 3.12$), a relationship of considerable public health importance. Education level is a common predictor of socioeconomic level, health literacy, and hygienic living conditions. logical, a person with a higher level of education will exhibit better hygienic habits, easy access to clean water, and a less crowded living environment, all of which are protective factors against H. pylori transmission. This relationship is consistent with other studies carried out in developing countries where disparities in education levels are major risk factors for H. pylori transmission [29, 30]. For instance, a study carried out in sub-Saharan Africa indicated that children of parents without formal education were 2.35 times more likely to acquire H. pylori infection compared to those of their peers whose parents were formally educated (95% CI: 1.38-3.99). In this study, unsafe drinking water sources were associated with almost five times increased risk of H. pylori transmission ($OR = 4.9$; 95% CI: 2.95-8.15).

In the Libyan study carried out on children, a family of 5-7 members was identified as a risk factor for H. pylori transmission. This is closely associated with education level and socioeconomic level.

A combination of factors is likely responsible for the relationship between H. pylori transmission risk and education level: (1) hygienic habits, (2) availability of clean water, (3) healthcare-seeking behaviour, and (4) housing conditions. It is this combination of factors that makes the relationship between H. pylori transmission risk and education level so high, irrespective of other demographic factors.

Gender and Residence:

This study's lack of a significant association between gender is supported by the findings of various studies that showed a comparable rate of Helicobacter pylori between genders, though some studies showed a variation between genders and age groups [12]. The Al Khoms study showed a preponderance of females, as was seen in a study done in Sudan, where females had a higher rate of Helicobacter pylori infection (46.4% vs. 22.7%).

For residence, the absence of a significant association may be attributed to the fact that the study area is characterized by relatively homogeneous characteristics, where environmental conditions are comparable between regions. However, a study done in Tripoli on pediatric patients showed that rural residence is a risk factor for Helicobacter pylori infection, which may be attributed to urbanization and the availability of potable water facilities.

Methodological Considerations and Advances in Medical Statistics:

This study made use of binary logistic regression analysis, which is considered a key component of epidemiological studies owing to their interpretability and well-established theoretical rationale [22]. Despite this, recent methodological developments provide promise for improved prediction.

A study on a comparative analysis of logistic regression and machine learning techniques for predicting complications in surgeries reported that multilayer perceptron models were able to achieve a higher AUROC value of 0.836 compared to logistic regression models at 0.707, owing to their ability to handle non-linear relationships. In a similar context, a study on dengue hospitalization risk prediction made use of logistic regression analysis to provide interpretable odds ratios as a component of random forest models for improved prediction [21,22].

In the context of infectious disease epidemiology, time-to-event analysis can provide further insight into the timing of infections. Although this study was unable to make use of this advanced statistical method, owing to the nature of the study being a cross-sectional study, this is a potential area of further study for Libya.

The Hosmer-Lemeshow test is considered a key component of logistic regression analysis, while the AUROC curve is considered a comprehensive measure of discrimination [24].

Limitations:

Although the study generated valuable insights, some limitations were identified, **which are as follows:**

1. Sample size: This study had a good sample size, consisting of 184 participants. However, it would be better if the sample size were larger, as it would provide more precise results. This can be confirmed by the wide confidence interval obtained for some variables, such as education, which ranges from 1.14 to 8.53.
2. Geographic scope: This study included various geographic regions of southern Libya; however, some regions were found to be underrepresented. Therefore, it would be better if the geographic scope were extended.
3. Study design: This study, being a cross-sectional study, can only show associations between risk factors and the results of the infections.
4. Generalizability: This study was conducted in the southern part of Libya, so the results might not be generalized for the entire country.
5. Future studies can be conducted with a larger sample size, good data collection tools, precise diseases, and the incorporation of machine learning and digital technologies.

Conclusion:

This study represents the first attempt at using inferential statistical analysis to examine the demographic factors associated with Helicobacter pylori infection in Southern Libya. Using binary logistic regression analysis, the study revealed that age and educational level were the only factors significantly associated with Helicobacter pylori infection. Older patients and those with a lower educational level were at a significantly increased risk of acquiring the infection. Although the discriminative ability of the binary logistic regression model was reasonable (AUROC = 0.706), the study revealed good calibration of the model (Hosmer-Lemeshow test $p = 0.821$).

The Helicobacter pylori infection rate of 11.7% in the Southern Libyan population was significantly different from the rates recorded in the Northern Libyan urban centres, which ranged from 68 to 84%. These differences might be explained by the geographical differences.

These findings have the following public health implications:

1. Targeted screening: In consideration of the high risk of H. pylori infection in the elderly and individuals with lower educational attainment, it is imperative to implement targeted screening to promote early intervention and treatment of H. pylori infection.
2. Health education: The public should be educated, especially individuals with lower educational attainment, about the significance of maintaining hygiene, including handwashing and food hygiene. Digital health interventions, especially via mobile technology, could be effective in this case.
3. Socioeconomic development: The elimination of disparities in educational attainment and living conditions, including access to clean water and sanitation, is crucial in controlling H. pylori infection. The high correlation between H. pylori and educational attainment (odds ratio = 3.12) suggests that interventions targeting modifiable risk factors, including socioeconomic development, could have a high impact in controlling H. pylori infection.
4. Standardization of diagnostic tests: The wide range of H. pylori prevalence in Libyan studies suggests the need to standardize the diagnostic tests to determine the actual H. pylori prevalence in the Libyan population.

5. Future research: Longitudinal studies with more extensive data collection, including environmental factors, should be carried out to determine the causality of H. pylori infection and identify additional modifiable risk factors. Machine learning and digital health interventions could be promising approaches in the near future.

Conclusion:

The present study proves the utility of logistic regression as an effective tool in determining risk factors in epidemiological studies. The findings of the present study contribute to the body of knowledge on H. pylori epidemiology in Libya and provide an impetus for the implementation of evidence-based interventions in the southern part of the country. The high correlation between H. pylori and educational attainment suggests the significance of health literacy and socioeconomic development in controlling H. pylori infection, and the high risk of H. pylori in the elderly suggests the potential for early intervention.

References:

[1] Hooi, J. K. Y., et al. (2017). Global prevalence of Helicobacter pylori infection: systematic review and meta-analysis. Gastroenterology, 153(2), 420-429.

[2] Zamani, M., et al. (2018). Systematic review with meta-analysis: the worldwide prevalence of Helicobacter pylori infection. Alimentary Pharmacology & Therapeutics, 47(7), 868-876.

[3] Malfertheiner, P., et al. (2022). Management of Helicobacter pylori infection: the Maastricht VI/Florence consensus report.

[4] World Health Organization. (2021). Helicobacter pylori Fact Sheet. Retrieved from https://www.who.int/news-room/fact-sheets/detail/h-pylori

[5] Brown, L. M. (2000). Helicobacter pylori: epidemiology and routes of transmission. Epidemiologic Reviews, 22(2), 283-297.

[6] Sugano, K., et al. (2015). Kyoto global consensus report on H. pylori gastritis. Gut.

[7] Kandil, H. E., et al. (2024). Epidemiological Map of H. pylori in North Africa: A Systematic Review. Journal of African Gastroenterology.

[8] Wang, J., et al. (2025). Effects of New Media–Based Education on the Treatment of Helicobacter pylori Infection: Systematic Review and Meta-Analysis. DOAJ.

[9] Bakir, A. (2021). Helicobacter pylori in Libya: Prevalence and antibiotic resistance patterns. Libyan Journal of Medicine.

[10] Aburagaegah, S., & Wajiej, A. (2025). Association of Helicobacter pylori Infection with Gender and Age Among Peptic Ulcer Patients at Al-Khoms Educational Hospital. Libyan Medical Journal.

[11] Albeshte, A., Mohamed, S., Berfad, A., Nami, S., & Nami, A. (2026). Comparison of Histopathology and Serological Testing for Helicobacter pylori Infection in Paediatric Dyspeptic Patients. Libyan

[12] Everhart, J. E., et al. (2000). Seroprevalence and ethnic differences in Helicobacter pylori infection among adults in the United States. The Journal of Infectious Diseases, 181(4), 1359-1363.

[13] Khalifa, M. M., Sharaf, R. R., & Aziz, R. K. (2010). Helicobacter pylori: a poor man's gut pathogen? Gut Pathogens, 2(1), 2.

[14] Naser, N. K. A. A., et al. (2025). Prevalence and associated risk factors of Helicobacter pylori infection among medical students at Shendi university, Sudan. BMC Gastroenterology, 25(1), 466.

[15] Nura N. Altaijori & Yousef M. Elshrek. (2017). Review Article: Risk Factors for Noncommunicable Diseases in Libya. The Egyptian Journal of Hospital Medicine, 1(66), 202-214.

[16] Eshraghian, A. (2014). Epidemiology of Helicobacter pylori infection in the Islamic Republic of Iran. Archives of Iranian Medicine, 17(10), 708-712.

[17] Malaty, H. M. (2007). Epidemiology of Helicobacter pylori infection. Best Practice & Research Clinical Gastroenterology, 21(2), 205-214.

[18] French, R. W., & Clemens, J. (2003). Helicobacter pylori in the developing world. Microbes and Infection, 5(8), 705-713.

[19] IBM Corp. IBM SPSS Statistics for Windows, Version 27.0. Armonk, NY: IBM Corp; 2020.

[20] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2022. URL: https://www.R-project.org/.

[21] Haberman, S. (1978). Analysis of qualitative data (Vol. 1). New York: Academic Press.

[22] Rojas-Sanchez, E., et al. (2025). Integrating machine learning and time-to-event models to explain and predict risk of hospitalization due to dengue in Colombia. Scientific Reports, 15, 38847.

[23] Chen, C. H., et al. (2025). Comparative performance of logistic regression, multilayer perceptron and decision tree models for predicting surgical pressure injuries: a retrospective cohort study. BMJ Health & Care Informatics, 32(1), e101532.

[24] BMC Pediatrics. (2025). Table 5: Bivariate analysis of Factors associated with H pylori infection. BMC Pediatrics.

[25] Ahmad, W. M. A. W., Aleng, N. A. and Ali, Z. (2010). Binary logistic regression analysis technique used in analyzing the categorical data in education science: a case study of Terengganu state, Malaysia. World Appl. Sci. J. 9(9): 1062-1066.
[26] Ahmad, W. M. A. W., et al. (2014). Association of hypertension with risk factors using logistic regression. Applied Mathematical Sciences, 8, 2563-2572.
Medical Journal, 70–74.
[27] Hosmer, D. W., & Lemeshow, S. (2000). Applied Logistic Regression (2nd ed.). New York: Wiley.
[28] O'Connell, A. (2005). Logistic regression models for ordinal response variables. Thousand Oaks, CA: Sage Publications. Quantitative Applications in the Social Sciences (No.146).
[29] Albeshte, A., Mohamed, S., & Nami, A. (2025). The Prevalence and Risk Factors of Helicobacter pylori Infection among Children with Dyspeptic Patients. Libyan Medical Journal, 289–294.
[30] Ozaydin, N., et al. (2013). Prevalence and risk factors of Helicobacter pylori in Turkey: a community-based study. Turkish Journal of Gastroenterology, 24(5), 415-420.