Gastro Hep Advances 2024;3:78–83

ORIGINAL RESEARCH—CLINICAL

Ascertainment of *Helicobacter pylori* Infection and Eradication Treatment Using a Nationwide Electronic Health Record Database

Shailja C. Shah,[1,2] Rohan Gupta,[3] Ranier Bustamante,[1] Mark Lamm,[1,4] Hanin Yassin,[1,4] Ashley Earles,[1,4] Adriana Hung,[5,6] Alese Halvorson,[5,7] Robert Greevy,[5,7] Samir Gupta,[1,2] Joshua Demb,[1,2] Lin Liu,[1,8,*] and Christianne L. Roumie[5,6,*]

[1]*Gastroenterology Section, Jennifer Moreno VA San Diego Healthcare System, San Diego, California;* [2]*Division of Gastroenterology, University of California San Diego, San Diego, California;* [3]*College of Medicine, State University of New York Upstate Medical University, Syracuse, New York;* [4]*Veterans Medical Research Foundation, San Diego, California;* [5]*Veteran Administration Tennessee Valley VA Health Care System Geriatric Research Education Clinical Center (GRECC), Nashville, Tennessee;* [6]*Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee;* [7]*Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee; and* [8]*Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, San Diego, California*

BACKGROUND AND AIMS: There are limited contemporary population-based data on *Helicobacter pylori* epidemiology and outcomes in the United States. Our primary aim was to create a validated cohort of veterans with *H pylori* testing or treatment using Veterans Health Administration data. **METHODS:** Using Veterans Health Administration structured and unstructured data, we developed and validated 4 algorithms for *H pylori* infection (3 algorithms) and treatment status (1 algorithm). During the development phase, we iteratively modified each algorithm based on a manual review of random sets of electronic health records (reference standard). The *a priori* validation goal was to achieve a one-sided 95% confidence lower bound (LB) for positive predictive value (PPV) and/or negative predictive value (NPV) >90%. We applied the Bonferroni correction when both PPV and NPV were relevant. **RESULTS:** For *H pylori* infection, we achieved 99.0% PPV (LB = 94.6%) and 100% NPV (LB = 96.4%) for discriminating *H pylori* positive vs negative status using structured (ie, laboratory tests) and 95% PPV (LB = 90.3%) and 97.9% NPV (LB = 93.9%) using unstructured (ie, histopathology reports) data. Diagnostic codes achieved 98% PPV (LB = 93.0%) for *H pylori* diagnosis. The treatment algorithm was composed of multiple antimicrobial combinations and overall achieved ≥98% PPV (LB = 93.0%) for *H pylori* treatment, except for amoxicillin/levofloxacin (PPV<60%). Application of these algorithms yielded nearly 1.2 million veterans with *H pylori* testing and/or treatment between 1999 and 2018. **CONCLUSION:** We assembled a validated national cohort of veterans who were tested or treated for *H pylori* infection. This cohort can be used for evaluating *H pylori* epidemiology and treatment patterns, as well as complications of chronic infection.

Keywords: Gastric Neoplasm; Gastrointestinal Diseases; Peptic Ulcer Disease; Infectious Disease

Introduction

Helicobacter pylori is a gram-negative bacterium that colonizes over half of the world's population and is the most common chronic bacterial infection globally. Chronic *H pylori* infection leads to gastric inflammation. While the majority of people will be asymptomatic or have nonspecific gastrointestinal symptoms, a small percentage of individuals will experience complications such as peptic ulcer disease or, rarely, gastric cancer.[1] *H pylori* is the strongest known risk factor for gastric adenocarcinoma and is designated by the World Health Organization as a human carcinogen. Successful eradication of *H pylori* infection is associated with a reduced risk of these complications.[2] Thus, because of *H pylori*'s persistent nature, eradication treatment is typically recommended. In the United States, treatment entails a 10–14 day course of 1–3 antibiotics and a proton pump inhibitor in combination.[3,4]

H pylori epidemiology, diagnostic and treatment practices, and outcomes such as eradication treatment failure are poorly described in the United States at a population level. The last large cohort of individuals evaluated for *H pylori* infection was the National Health and Nutrition Examination Survey cross-sectional study in 1988–1991 and 1999–2000.[5,6] This cohort evaluated *H pylori* infection via serology only and did not evaluate non-serological testing or *H pylori* treatment. Notably, the first clinical consensus and guideline statements recommending *H pylori* eradication

**These authors contributed as co-senior authors.*

Abbreviations used in this paper: **EGD, esophagogastroduodenoscopy; EHR, electronic health records; ICD, International Classification of Diseases; LB, lower bound; NPV, negative predictive value; PPI, proton pump inhibitor; PPV, positive predictive value; VA, Veterans Affairs; VA-HP, VA-*Helicobacter pylori* cohort; VHA, Veterans Health Administration.**

Most current article

2772-5723
https://doi.org/10.1016/j.gastha.2023.09.005

treatment among individuals testing positive were not published until 1998.[7] Ascertainment of *H pylori* episodes of care poses a challenge due to the several modalities of testing and treatment, as well as imperfect diagnostic coding. These knowledge gaps not only impact patient outcomes related to *H pylori* management but also resource allocation for at-risk groups.

Our primary aim was to assemble and validate a national cohort of veterans who had been tested for *H pylori*, diagnosed with *H pylori* based on diagnostic billing codes, or treated for *H pylori* as ascertained by structured and unstructured data contained within the Veterans Health Administration (VHA), one of the largest integrated health systems in the United States.

Methods

This study was approved by the committee on research ethics at Veterans Affairs (VA) San Diego and VA Tennessee Valley, in accordance with the Declaration of Helsinki.

Study Data Sources

We compiled data from the VHA, which provides longitudinal care to approximately 9 million US veterans. The VHA Corporate Data Warehouse aggregates longitudinal patient data from all US veterans receiving care through any VHA facility across the United States and its territories. Individual-level data include demographics, comorbidities, laboratory testing and results, procedures and results, progress notes, diagnoses, billing codes, vital status files, and all dates. We collated all demographics, laboratory tests and results, International Classification of Diseases (ICD) codes, and procedure information related to inpatient and outpatient encounters between 1999 and 2018. These data elements include a mix of both structured and unstructured data. We linked these individual-level data to the VHA pharmacy files to identify details of all filled prescriptions, including dates of therapy, quantity dispensed, and dosing instructions (ie, dose and frequency).

Algorithm Development and Iteration

Diagnosis. Three algorithms were developed to identify individuals tested for and/or diagnosed with *H pylori* (Figure).

Algorithm 1 development: ascertainment of H pylori *status based on* H pylori *diagnostic laboratory tests.* The following elements were ascertained from inpatient and outpatient structured data: *H pylori* date of testing, testing modality (serology vs non-serology), and testing result (positive, negative, equivocal). *H pylori* stool antigen testing, Campylobacter-like organism test, rapid urease test, urea breath test, and *H pylori* culture were classified as non-serological tests. Qualitative results of each test were recorded based on the data query; result categories included positive or negative. While the majority of serological lab tests were reported qualitatively, a minority were reported quantitatively (<10%). The lab value ranges and units of measurement for these quantitative results were explored. It was identified that the threshold values for categorizing these quantitative labs as positive vs negative varied depending on the test type, assay, and unit of measurement. Because there was no way to standardize the reference ranges and units across each quantitative lab measurement, we categorized values greater than the 75th percentile as 'positive,' less than the 25th percentile as 'negative,' and between the 25th and 75th percentile as 'equivocal.' Qualitative equivocal tests were not evaluated as they comprised <3% of all lab values.

Algorithm 2 development: ascertainment of H pylori *status based on histopathology reports with gastric biopsies.* This algorithm leveraged both keyword searching and natural language processing methods applied to unstructured data. We identified individuals who had undergone an inpatient or outpatient esophagogastroduodenoscopy (EGD) procedure as identified based on the Current Procedural Terminology code (Table A1) and had a pathology report available in the ± 30 days from the date of the EGD.

We then applied the keywords for 'stomach' and different gastric locations only to the portion of the histopathology report where biopsy results are reported. All reports demonstrated the same organization of unstructured data elements; for example, all biopsy results were reported in the section following the "gross description" section. In order to prevent interpreting historical or history of *H pylori* or other gastric diseases, we developed a code such that only the text after the phrase "gross description" in the pathology report was interpreted. Thus, if the histopathology report states "history of *H pylori*" in the indication section, but gastric biopsies were not obtained during the EGD, then this report would not be selected by the algorithm since *H pylori* status cannot be ascertained without gastric sampling.

Pathology report documents were divided into separate sentences for the purpose of identifying sentences with positive or negative diagnosis of *H pylori*. The end of a sentence was defined as having at least 2 characters and could include parentheses or letters in that character string. Gastric biopsies evaluate the presence vs absence of *H pylori* infection using a stain and often include positive and negative quality control results in the report. Thus, the initial algorithm was constructed such that word pairings or phrases related to the performance of the control stains were removed from consideration; these phrases included "positive and negative control stained appropriately," "positive control," "positive and negative," "appropriately positive," and "appropriately negative." "With feature" and "with finding" were also excluded due to their high frequency and nonspecific context. Once this basic algorithm was developed, formal rounds of validation using randomly selected groups of algorithm positive vs negative individuals were completed, with algorithm modifications implemented as appropriate if threshold performance metrics were not achieved.

Algorithm 3 development: ascertainment of H pylori *diagnosis based on ICD-9 and ICD-10 codes.* We evaluated the performance of a single inpatient or outpatient ICD-9 (041.86) or ICD-10 (B96.81) code for identifying someone with a true diagnosis of *H pylori* infection, in the absence of other structured indicators of *H pylori* infection (ie, laboratory tests). A diagnosis was considered verified if a manual chart review confirmed that the individual had *H pylori* infection based on testing (either within or outside the VHA), treatment, or a clinical progress note documenting current or prior *H pylori* infection. ICD codes can only be used to identify *H pylori*-positive individuals since there are no corresponding ICD codes for *H pylori*-negative status (ie, someone who was tested for *H pylori*

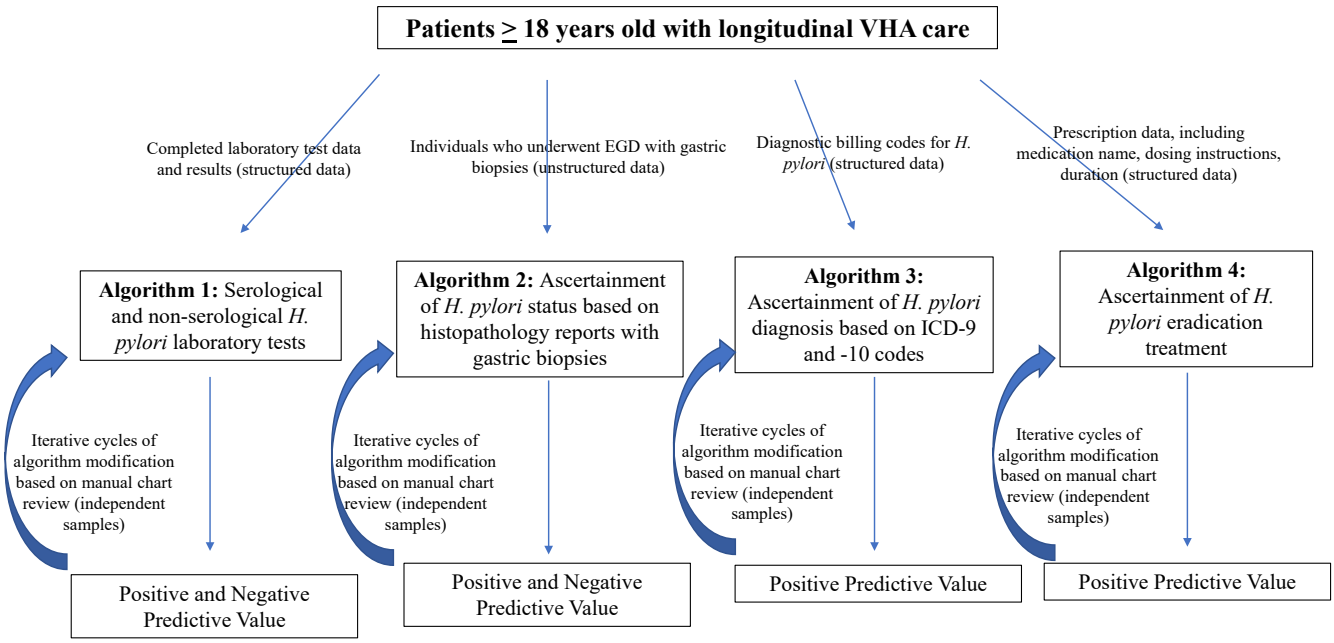

Figure. Schematic diagram of the validation approach for *H pylori* diagnosis or treatment status based on structured and unstructured Veterans Health Administration data.

but tested negative). Thus, the negative predictive value (NPV) was not applicable to this algorithm.

***H pylori* eradication treatment.** *Algorithm 4 development: ascertainment of* H pylori *eradication treatment.* We validated multiple guideline-recommended *H pylori* eradication regimens. *H pylori* eradication treatment consists of 1–3 antimicrobials ± bismuth along with twice per day proton pump inhibitor (PPI), in combination for 10–14 days. The US guideline-recommended *H pylori* eradication treatment regimens are provided in Table A2.[3] Because PPIs are also obtained over the counter, these medications were not included in the initial algorithm development. For each *H pylori* eradication regimen, we evaluated the accuracy of individual treatment regimens for *H pylori* in the absence of a concomitant ICD code or *H pylori* laboratory test.

A minimum 7-day overlap was required for each prescription, which was based on an initial round of chart review and data exploration during algorithm development. The days of medication dispensed, that is "day supply," was restricted to 5–30 days in order to exclude outliers; even though therapy is only prescribed for a maximum of 14 days, it is not uncommon for a 30-day days supply to be dispensed due to VHA benefits, which is why 30 days and not 14 days were selected as the upper limit. High-dose amoxicillin and PPI therapy (dual therapy) were not included in the query due to the nonspecific nature and due to the rarity of this regimen for *H pylori* treatment in the United States currently.[3,8] For each *H pylori* treatment regimen, we evaluated its positive predictive value (PPV) for *H pylori* indication in the absence of a concomitant ICD code or *H pylori* laboratory test. NPV was not applicable for this algorithm.

Chart Abstraction, Validation, and Analysis

Each algorithm underwent independent rounds of review of a randomly selected set of unique electronic health records (EHRs) that were categorized as algorithm "positive" or "negative." Independent samples of cases and controls were drawn for each round of review. The EHR was considered the reference standard for categorizing true vs false positives and negatives. The EHR comprised all laboratory tests, medical encounter notes, procedure notes, medications, and billing documentation. Medical charts without any evidence or mention of *H pylori* testing, diagnostic code, or treatment were not considered in the validation process. The reviewers (M.L., H.Y., S.C.S.) were blinded to the algorithm output. Discordant findings were carefully reviewed and the algorithms were modified as needed based on the etiology of the discordance.

The a *priori* goal was to achieve a one-sided 95% confidence lower bound (LB) for PPV and/or NPV >90%. Iterative modifications to each algorithm were applied if the a *priori* thresholds for PPV or NPV were not achieved (Figure). The PPV was defined as the proportion of subjects who were correctly classified by the proposed algorithm as having the outcome (ie, proportion of true positives; eg, individuals who were classified by algorithm 1 or 2 as *H pylori*-positive and who were manually confirmed to have *H pylori* based on laboratory testing or histopathology, respectively). The NPV was defined as the proportion of subjects without the outcome among those who were classified by the proposed algorithm as not having the outcome (ie, proportion of true negatives; eg, individuals who were classified by algorithm 1 or 2 as *H pylori*-negative and who were manually confirmed to be *H pylori*-negative on laboratory testing or histopathology, respectively). The prevalence estimates for each of the PPV and/or NPV performance calculations were derived by applying the respective algorithms to the full cohort of individuals receiving longitudinal VHA care.

For those algorithms in which both PPV and NPV were relevant (algorithms 1 and 2), we applied the Bonferroni multiple comparison adjustment to ensure an overall confidence level of 95% for PPV and NPV estimates. Based on these target thresholds and the estimated PPV and NPV from the

development phase, we projected that we would need to randomly sample at least 100 algorithm-positive cases and 100 algorithm-negative controls, respectively, for each iterative validation round.[9] Based on the projected sample size, if the estimated PPV and NPV achieved 95% or greater in the validation round, the 95% one-sided confidence LBs would be at least 90%, and if the estimated PPV and NPV achieved 90% or greater, then the 95% one-sided confidence LBs would be at least 84%.[9]

H Pylori *Cohort Creation*

We applied each of the 4 validated algorithms to individuals >18 years old receiving longitudinal care within the VHA to assemble a nationwide cohort of *H pylori* tested and/or treated individuals between 1999 and 2018. Longitudinal care was defined as having at least one medical encounter, laboratory test, or prescription fill. We compiled basic demographic information for the cohort.

Results

The performance of each of the validation algorithms is summarized in Table.

Algorithm 1: H Pylori *Laboratory Tests*

Among individuals with any laboratory-based *H pylori* testing, the algorithm classified the testing result as positive vs negative with 99% PPV (LB 94.6%) and 100% NPV (LB 96.4%) (Table). This algorithm performance was achieved using the threshold for classifying quantitative *H pylori* results greater than the 75th percentile as positive and less than the 5th percentile as negative; as previously noted, quantitative results comprised <10% of the *H pylori* lab tests overall. The prevalence of *H pylori* based on the laboratory testing algorithm was 26%, which is consistent with recently published data reporting *H pylori* positivity rates among the VHA population with an indication for *H pylori* testing.[10] The sensitivity of the algorithm was 100% and specificity was 99.6%.

Algorithm 2: Histopathology

In the first round of validation, the algorithm achieved 94% PPV (LB 87.4%) and 97% NPV (LB 91.5%) for *H pylori* positive and negative diagnosis, respectively (Table), which was above the a *priori* target performance threshold for PPV.

The algorithm was therefore modified based on the experience from the first round. We subsequently added variations of "*neg helicobacter found stain*" to the negative identifier for *H pylori* and removed the 4-part rule "*no strength histopath abnormal*," in favor of using various 3-part combinations of that phrase instead (eg, "*no strength histopath*," "*no histopath abnormal*"). The second independent random set of 100 algorithm-positive and 100 algorithm-negative EHRs were reviewed and the PPV improved to 94.5% (LB 88.5%), without significantly impacting the NPV (97.3%, LB 92.2%), but this was still above the target PPV.

In the third validation round, we again modified the algorithm based on the second-round experience. We adjusted the code to recognize complete sentences such that multiple punctuation characters would be accepted at the end of a sentence (eg, ".."). We also expanded the tenses for applicable positive phrases and added "none" and "non-" as synonyms for "no." A third independent random sample of 100 algorithm-positive and 100 algorithm-negative EHRs were selected to validate and achieved 95.0% PPV (LB 90%) and 97.9% NPV (LB 93.9%). Thus, all a *priori* threshold performance values were met and this was considered the final algorithm. The prevalence of *H pylori* based on the final algorithm was 11%, which is consistent with estimates among the VHA population (unpublished data). The estimated sensitivity and specificity were 84.7% and 99.4%, respectively.

Algorithm 3: Diagnostic Codes

The presence of one inpatient or outpatient ICD-9 or ICD-10 code alone achieved a 98% PPV (LB 93.8%) for a true diagnosis of *H pylori* infection based on EHR documentation among individuals with at least one ICD-9 or ICD-10 code for *H pylori* (Table). Often this was confirmed based on *H pylori* testing that occurred outside of the VHA.

Algorithm 4: H Pylori *Eradication Treatment*

We validated the performance of the eradication treatment algorithm for correctly identifying that these regimens were indeed being used for *H pylori* treatment and not another indication, even in the absence of other structured data indicators of *H pylori* infection (ie, *H pylori* testing and/or ICD codes).

The first round of validation after the development phase yielded an observed PPV of 98% (LB 93%) for *H pylori* eradication treatment indication for each of the guideline-recommended therapies (see Table A2) with the exception of the combination of amoxicillin/levofloxacin (Table). EHR review of amoxicillin/levofloxacin cases revealed that this regimen was most often prescribed for skin and soft tissue or other non-*H pylori* infections. As such, in order to improve performance, we modified the algorithm for amoxicillin/levofloxacin to remove any patients with ICD codes for non-*H pylori* infections in the ±30 days surrounding the date of amoxicillin and levofloxacin prescription fills, as well as removing individuals who had prescription fills that were not consistent with *H pylori* treatment recommendations (ie, removing any day supply that was not 7 days, 10 days, or 14 days). We also narrowed the overlap period from 7 days to 5 days. Other elements of the prescription including the date prescribed and duration of therapy were 100% accurate; however, *indication* for the prescription combination could not achieve the *a priori* threshold for validity (observed PPV 55%, LB 44.7%). Thus, in the absence of other indicators such as *H pylori* laboratory testing or ICD code, the combination of amoxicillin/levofloxacin alone is not indicative of *H pylori* eradication treatment, and this combination was therefore removed from the algorithm. All other guideline-recommended

Table. Summary Performance of Validation Algorithms

Algorithm	Iterative rounds[a]	Prevalence[b]	PPV (lower bound)[c]	NPV (lower bound)[c]
1 – Laboratory tests	1	0.26	0.99 (0.95)	1.00 (0.96)
2 – Histopathology	1	0.13	0.94 (0.87)	0.97 (0.915)
	2	0.12	0.95 (0.86)	0.97 (0.92)
	3	0.11	0.95 (0.89)	0.98 (0.94)
3 – ICD codes	1	0.10	0.98 (0.94)	n/a
4 – Eradication treatment	1	0.10	0.98 (0.94)	n/a
	2 (amoxicillin/levofloxacin combination only)	0.10	0.55 (0.46)	n/a

n/a, not applicable.
[a]100 algorithm-positive and 100 algorithm-negative controls were manually reviewed for the validation phase in each round with the following exceptions: 110 algorithm-positive and 110 algorithm-negative EHRs were manually reviewed for round 2 of Algorithm 2, and 140 algorithm-positive and 140 algorithm-negative EHRs were manually reviewed for round 3 of Algorithm 2. Independent samples of cases and controls were drawn for each round of review.
[b]Prevalence estimates are derived from applying the respective algorithms to the full data set of individuals ≥18 years old with longitudinal VHA care.
[c]95% one-sided confidence lower bound for PPV and NPV was based on binomial exact test with Bonferroni correction for Algorithms 1 and 2.

regimens achieved very high PPV for an *H pylori* eradication treatment indication independent of any other structured *H pylori* data indicators in the VHA record (eg, laboratory tests, ICD codes) and thus comprised the final algorithm for *H pylori* eradication treatment.

H Pylori *Cohort Details*

Of 15,594,932 individuals who received care through the VHA between 1999 and 2018, we identified 1,199,032 individuals with *H pylori* testing (laboratory and upper endoscopy with gastric biopsies [algorithms 1 and 2]) (n = 913,328; 76.2%), *H pylori* ICD-9 or ICD-10 code only without evidence of laboratory testing [algorithm 3] (n = 60,472; 5.0%), or *H pylori* eradication treatment only without evidence of laboratory testing or ICD-coding [algorithm 4] (n = 225,232; 18.8%). The majority of eradication regimens with clarithromycin-based triple therapy (Table A2), while amoxicillin/levofloxacin comprised less than 7% of all eradication combinations. These 1,199,032 individuals comprised the starting VA-*Helicobacter pylori* (VA-HP) cohort. Basic demographics of the cohort included mean age 58.1 (standard deviation, 14.9) years, 90.4% men, 66.2% non-Hispanic White, 15.3% non-Hispanic Black, 6.1% non-Hispanic other, and 6.4% unknown, and 6.0% Hispanic.

Discussion

We achieved the primary aim of this study and validated a large nationwide cohort of individuals receiving care through the VHA who underwent testing and/or treatment for *H pylori* infection (VA-HP cohort). Indeed, of the nearly 1.2 million people identified, 913,328 (76.2%) had evidence of formal *H pylori* testing. By using both structured and unstructured data to develop the algorithms, we optimized both data capture and performance, as evidenced by most algorithms only requiring one round of manual chart review to achieve *a priori* performance thresholds. All algorithms for *H pylori* diagnosis based on testing or ICD codes achieved PPV 94%–99% (LB 90%–94.6%) and NPV 97.9%–100% (LB 93.9%–96.4%) and, with the exception of the combination of amoxicillin/levofloxacin, all guideline-recommended combinations of *H pylori* eradication treatments achieved PPV 98% (LB 93.8%) for *H pylori* treatment indication.

To our knowledge, this is the largest validated cohort of individuals with confirmed *H pylori* testing and results available. The VA-HP national cohort can be used to describe contemporary *H pylori* epidemiology, risk factors, diagnostic and treatment practices, and outcomes, including *H pylori* eradication treatment failure. In addition to validating the results of testing, we confirmed that we were able to ascertain related *H pylori* testing and procedural dates, testing modality (serology vs non-serology), and prescription details. Notably, the VHA data are readily linked to the VA Central Cancer Registry and National Death Index, which accurately identify cancer diagnoses and causes of death, respectively, as well as the Million Veteran Program, one of the largest actively enrolling genomic biobanks globally.[11] Accordingly, the VA-HP cohort can not only be leveraged for time-to-event analyses related to cancer and cancer mortality outcomes but also to characterize genetic and gene x environment underpinnings of *H pylori* susceptibility and its plethora of complications. Importantly, the VA-HP cohort includes all individuals *tested* for *H pylori* which reduces the potential for misclassification since *H pylori* infection is often asymptomatic and may therefore go undiagnosed unless someone is formally tested.

Our work extends limited prior literature reporting on the PPV and NPV of using administrative data to identify individuals with an *H pylori* diagnosis. Thirumurthi et al[12] evaluated the performance of *H pylori* ICD-9 code or prescription of triple or quadruple eradication therapy for a diagnosis of *H pylori*. They reported that the PPV of the *H*

pylori ICD-9 code ranged between 97.4% and 100%, while their algorithm for triple and quadruple therapy had a PPV of 73.7% and 97.7%, respectively. Notably, this study did not validate laboratory tests or histopathology obtained from EGD with gastric biopsies. This study published in 2008 also predated the most recent 2017 US American College of Gastroenterology guidelines on *H pylori* treatment and predated the introduction of ICD-10 coding.[3] It is also worth noting that, while ICD codes have excellent PPV for *H pylori*, these only capture a very small fraction of the pool of individuals with *H pylori* infection in the VHA and also miss all individuals who were tested but who tested negative.

Our study has several strengths. We leveraged comprehensive algorithm development using both structured and unstructured data and a rigorous validation process. We validated all guideline-recommended *H pylori* eradication treatments and reported on the accuracy of capturing the correct medications, dates of treatment, as well as indication (i.e., treatment of *H pylori* infection). Lastly, we were able to apply our algorithm across the VHA nationwide database, which includes over 160 distinct VHA stations, to create the VA-HP cohort spanning 1999–2018. Our study also has limitations. We are not able to reliably capture care that occurs outside of the VHA, and thus there is still the small possibility of misclassification. Our algorithms were validated against the VHA database and may perform differently in other electronic health record systems; future work can assess the performance of these algorithms on non-VA data.

In conclusion, we created a large, nationwide validated cohort of individuals with established VHA care tested or treated for *H pylori* infection. *H pylori* infection is common, particularly in high-risk groups such as non-White racial and ethnic groups, early-generation immigrants from countries where *H pylori* infection is endemic, and older individuals. Chronic *H pylori* infection is associated with potentially serious complications, including gastric cancer. Accordingly, this powerful cohort comprising nearly 1.2 million individuals can be used to evaluate *H pylori* epidemiology, treatment patterns, and outcomes, as well as risk factors for disease complications related to chronic *H pylori* infection.

Supplementary Materials

Material associated with this article can be found in the online version at https://doi.org/10.1016/j.gastha.2023.09.005.

References

1. Plummer M, Franceschi S, Ome Vignat J, et al. Global burden of gastric cancer attributable to *Helicobacter pylori*. Int J Cancer 2015;136:487–490.
2. Shah S, Hubscher E, Pelletier C, et al. *Helicobacter pylori* infection treatment in the United States: clinical consequences and costs of eradication treatment failure. Expert Rev Gastroenterol Hepatol 2022;16:341–357.
3. Chey WD, Leontiadis GI, Howden CW, et al. ACG clinical guideline: treatment of *Helicobacter pylori* infection. Am J Gastroenterol 2017;112:212–239.
4. Shah SC, Iyer PG, Moss SF. AGA clinical practice update on the management of refractory *Helicobacter pylori* infection: expert review. Gastroenterology 2021;160:1831–1841.
5. Grad YH, Lipsitch M, Aiello AE. Secular trends in *Helicobacter pylori* seroprevalence in adults in the United States: evidence for sustained race/ethnic disparities. Am J Epidemiol 2012;175:54–59.
6. Everhart JE, Kruszon-Moran D, Perez-Perez GI, et al. Seroprevalence and ethnic differences in *Helicobacter pylori* infection among adults in the United States. J Infect Dis 2000;181:1359–1363.
7. Howden CW, Hunt RH. Guidelines for the management of Helicobacter pylori infection. Am J Gastroenterol 1998; 93:2330–2338.
8. Shah S, Cappell K, Sedgley R, et al. Diagnosis and treatment patterns among patients with newly diagnosed *Helicobacter pylori* infection in the United States 2016-2019. Sci Rep 2023;13:1375.
9. Liu L, Bustamante R, Earles A, et al. A strategy for validation of variables derived from large-scale electronic health record data. J Biomed Inform 2021;121:103879.
10. Shah SC, Halvorson AE, Lee D, et al. *Helicobacter pylori* burden in the United States according to individual demographics and geography: a nationwide analysis of the Veterans Healthcare System. Clin Gastroenterol Hepatol 2023. http://doi.org/10.1016/j.cgh.2023.05.016.
11. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: a mega-biobank to study genetic influences on health and disease. J Clin Epidemiol 2016;70:214–223.
12. Thirumurthi S, Desilva R, Castillo DL, et al. Identification of *Helicobacter pylori* infected patients, using administrative data. Aliment Pharmacol Ther 2008;28:1309–1316.

Received May 31, 2023. Accepted September 7, 2023.

Correspondence:
Address correspondence to: Shailja C. Shah, MD, MPH, VA San Diego Healthcare System, 3350 La Jolla Village Drive, 3rd Floor South, GI Suite. Mail Code 111D, San Diego, California 92161. e-mail: s6shah@health.ucsd.edu.

Authors' Contributions:
Shailja C. Shah conceived the study. Shailja C. Shah, Mark Lamm, Hanin Yassin conducted manual chart review. Shailja C. Shah, Ranier Bustamante, Lin Liu conducted the analysis. Shailja C. Shah, Ranier Bustamante, Lin Liu, Samir Gupta, Christianne L. Roumie interpreted the results. Shailja C. Shah drafted the manuscript.

Conflicts of Interest:
The author discloses the following: Shailja C. Shah serves as an *ad hoc* consultant and advisor for Phathom Pharmaceuticals and *ad hoc* consult for RedHill Biopharma. The remaining authors disclose no conflicts.

Funding:
The results reported herein correspond to specific aims of grant ICX002027A01 awarded to investigator Shailja C. Shah from the United States Department of Veterans Affairs. This work was also supported by a 2019 Research Scholar Award from the American Gastroenterological Association and NIH P30 DK120515 awarded to the San Diego Digestive Diseases Research Center from the National Institutes of Health.

Ethical Statement:
The corresponding author, on behalf of all authors, jointly and severally, certifies that their institution has approved the protocol for any investigation involving humans or animals and that all experimentation was conducted in conformity with ethical and humane principles of research.

Data Transparency Statement:
These data are not publicly available. Data may be accessed by approved Veterans Affairs investigators with appropriate research privileges and distinct institutional review board project approval.

Reporting Guidelines:
Declaration of Helsinki.