Supplemental Methods Section

This retrospective cohort study was conducted within the VHA Corporate Data Warehouse, which includes data accumulated since October 1999. To identify patients with potential HP infection, we identified if patients had any one of the following criteria: 1) HP-associated International Classification of Diseases (ICD) 9/10 codes (ICD-9: 041.86; ICD-10: B96.81), 2) prescription for one of 11 accepted eradication regimens for HP as recommended by the American College of Gastroenterology [1], 3) positive urea breath test, 4) positive stool antigen test, or 5) pathology report indicating active HP infection.

Natural language processing was performed to ensure true infection on pathology. Natural language processing included identifying all pathology reports that indicated HP testing was performed, and sentiment analysis to ensure that testing was positive (by identification of organisms or specific stains). Steps included spell correction, data reduction, data transformation including disambiguation of abbreviations, regular expression matching, negative expression matching, and finally, validation. Upon validation, we found 100% positive predictive value for HP if classified as positive by our natural language processing algorithm.

We then identified patients with AGAC via ICD 9/10 codes (ICD-9: 151.9; ICD-10: C16.1-C16.9) and/or the Veterans Affairs Central Cancer Registry.[2] We filtered to include intestinal type non-cardia gastric cancers, to avoid capturing non-

adenocarcinomas and cardiac/gastroesophageal junction tumors, which are less clearly associated with HP.[1, 6] Of the patients identified by ICD code, we had pathology reports for 2,071, and sample validation showed >90% were truly distal adenocarcinomas.

Patients with AGAC were subcategorized according to the method by which their HP status was determined*:* 1) negative serum antibody (Ab) with documented laboratory test with a negative result (whether HP infection was confirmed by another method or not); 2) a positive serum Ab (documented laboratory test with a positive result) with or without additional evidence for HP; 3) a documented history of active infection without serum Ab testing (a prescription for an accepted eradication regimen, administrative code, or positive diagnostic test); 4) no evidence of serum Ab testing and no history of an active infection (via any criteria.)

REFERENCES

1. Thirumurthi S, Desilva R, Castillo DL, et al. Identification of Helicobacter pylori infected patients, using administrative data. Aliment Pharmacol Ther 2008;28:1309-16.
2. Zullig LL, Sims KJ, McNeil R, et al. Cancer Incidence Among Patients of the U.S. Veterans Affairs Health Care System: 2010 Update. Mil Med 2017;182:e1883-e1891.

Supplemental Table 1: Comparison of *H. pylori* seronegative and *H. pylori* seropositive gastric cancer patients

	H. pylori seronegative (n=1,071)	*H pylori* seropositive (n=1,672)	History of active *H pylori* (n=2,006)	Unknown *H pylori* history (n=11,049)	P-value
Age at cancer diagnosis, years median (IQR)	66.4 (59.4, 75.0)	68.2 (60.4, 76.6)	68.3 (60.9, 76.5)	69.1 (61.8, 77.2)	0.001
Age at serum test, years median (IQR)	65.8 (58.7, 74.3)	62.3 (53.5, 72.5)			0.001
Time from serum to cancer diagnosis, years median (IQR)	0.1 (-0.1, 1.5)	3.8 (0.1, 10.1)			0.001
Race					0.001
White	687 (69.3%)	843 (57.2%)	965 (54.8%)	6264 (65.9%)	
Black or African American	221 (22.3%)	506 (34.3%)	649 (36.8%)	2482 (26.1%)	
American Indian or Alaska Native	4 (0.4%)	12 (0.8%)	15 (0.9%)	67 (0.7%)	
Asian	8 (0.8%)	14 (0.9%)	17 (1.0%)	56 (0.6%)	
Native Hawaiian/ Pacific Islander	10 (1.0%)	18 (1.2%)	22 (1.2%)	87 (0.9%)	
Unknown	61 (6.2%)	81 (5.5%)	94 (5.3%)	551 (5.8%)	
Ethnicity					
Hispanic or Latino	88 (8.8%)	186 (12.5%)	198 (11.1%)	747 (7.8%)	0.001
Not Hispanic/Latino	870 (86.8%)	1235 (83.1%)	1520 (85.0%)	8402 (87.8%)	
Unknown	44 (4.4%)	66 (4.4%)	71 (4.0%)	425 (4.4%)	
Male	1033 (96.5%)	1627 (97.3%)	1944 (97.0%)	10,769 (97.5%)	0.17