Helicobacter pylori infection in Arctic Indigenous communities: assessing the validity of infection status measures and estimating the effect of bacterial load on gastritis severity

by

Amrit Singh Passi

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Clinical Epidemiology

School of Public Health
University of Alberta

Abstract

Helicobacter pylori (*Hp*) causes chronic inflammation of the stomach lining, also known as gastritis, in nearly all people with a persistent infection. Chronic inflammation of the stomach lining can lead to atrophy of stomach glands, a condition known as atrophic gastritis, believed to be the start of a sequence of pathological conditions associated with an increased risk of stomach cancer. The Canadian North *Helicobacter pylori* (CAN*Help*) Working Group conducted community projects to respond to concerns of Arctic Indigenous communities about health risks from *Helicobacter pylori* (*Hp*) infection. CAN*Help* community projects used multiple methods for classifying *Hp* infection, including the urea breath test (UBT), pathological evaluation of stomach tissue, and tissue culture. My thesis assessed the validity of *Hp* infection detection methods and estimated the effect of the *Hp* bacterial load on the severity of gastritis among CAN*Help* project participants.

Estimating the accuracy of alternative staining methods provides information needed to evaluate the cost-effectiveness of special stains for detecting *Hp*. In this thesis, I investigated whether using special stains (Giemsa, Warthin-Starry and/or immunohistochemical (IHC) stains) on tissue slides for pathological assessment of stomach tissue biopsies improves the accuracy of *Hp* detection. To do this, I analyzed data from participants with stomach tissue biopsies collected from 2008 through 2013 in 4 CAN*Help* community projects. I estimated accuracy measures (agreement, sensitivity, specificity, and predictive values) for each stain's classification of tissue samples as *Hp*-positive or *Hp*-negative compared against different definitions of the 'gold standard' classification. I also compared the *Hp* density gradings of 2 pathologists and assessed the relationship between *Hp* density and UBT values. The special stains evaluated for this study had similar accuracy based on estimated measures. Because IHC has a much higher cost than the

other stains assessed here, either Warthin-Starry or Giemsa would be more cost-effective than IHC for improving *Hp* detection during the pathological assessment of stomach tissue biopsies relative to not using a special stain. The *Hp* density grades of the 2 pathologists showed good consistency of agreement. The *Hp* density grades were positively associated with UBT values.

Considering the high global mortality from stomach cancer, a 2014 International Agency for Research on Cancer *Hp* Working Group recommended that countries explore the possibility of introducing *Hp* screening and treatment programs; this panel cautioned, however, that decisions about whether and how to implement such programs must be based on local considerations of disease burden and other factors of relevance to health policy. Identifying local indicators of high risk for targeted interventions is key for effective cancer prevention programs. Thus, understanding the relationship between *Hp* density and gastritis severity in specific populations at elevated risk of stomach cancer has great public health significance. In this thesis, I also investigated whether the frequency of severe *Hp*-associated gastritis increases as *Hp* density increases. To do this, I estimated the effect of *Hp* density on gastritis severity among *Hp*-positive participants in CAN*Help* community projects. This analysis used data from participants with stomach tissue biopsies collected from 2008 through 2017 in 7 CAN*Help* community projects. The results showed that gastritis severity increases sharply with increasing *Hp* density among *Hp*-positive participants.

Preface

This thesis is an original work by Amrit Passi (AP). The research conducted by the CAN*Help* Working Group, of which this thesis is a part, received ethics approval from the University of Alberta Research Ethics Board under the project name "Addressing Community Concerns about Risks from *H. pylori* Infection in the Circumpolar North" (No. Pro00007868) on January 20, 2017. The research conducted by the CAN*Help* Working Group, of which this thesis is a part, also received approval from the Northwest Territories (Licence No. 15785) and Yukon (Licence No. 16-13S&E) research licensing authorities.

Research conducted as a part of this thesis was a result of several collaborations. Collaborators include Dr. Karen J Goodman (KJG), Dr. Dean Eurich (DE), Dr. Gregory Charrois (GC), Dr. Safwat Girgis (SG), Dr. Sander Veldhuyzen van Zanten, Taylor Cromarty (TC), Tyler Dang (TD), CAN*Help* Working Group project staff and CAN*Help* Working Group community planning committees. AP wrote each chapter of this thesis under the supervision of KJG, with input from DE and GC. AP was responsible for the literature review in Chapter 2 and was assisted by TD in developing a search strategy and reviewing article titles and abstracts in Chapter 2.1. AP was responsible for data analysis in Chapters 3 and 4 with assistance from TC for data management and cleaning. The analyses in Chapters 3 and 4 used data previously collected by project staff in CAN*Help* community projects, including histopathology results evaluated by both GC and SG in Chapter 3 and histopathology results evaluated by SG in Chapter 4.

No part of this thesis has been previously published.

Acknowledgments

First and foremost, I would like to express my sincerest gratitude to my supervisor, Dr. Karen Goodman. I am privileged to have called you “coach” through my graduate program. Without your persistent help, guidance, and knowledge, I would not have been able to keep up the momentum needed to complete this thesis. Thank you so much for your unwavering support of my academic and professional goals.

I thank my thesis committee members, Dr. Dean Eurich and Dr. Gregory Charrois, for their expertise and prompt feedback. Thank you for always being available to answer questions. I would especially like to thank Dr. Charrois for the significant time he spent evaluating over 700 gastric biopsy slides for the central chapter of my thesis.

I want to thank the CAN*Help* community project planning committee members and community members for their guidance and gracious hospitality during my time spent participating in fieldwork in the communities. Thank you to the members of the CAN*Help* Working Group, including Ali Assi, Hsiu-Ju Chang, Taylor Cromarty, Tyler Dang, Katharine Fagan-Garcia, Dr. Nancy Price, and Dr. Douglas Quilty, who helped me throughout my thesis project's various stages. I would also like to thank everyone I had the opportunity to work with and be mentored by at the University of Alberta and Alberta Health Services.

I am grateful for all the support I received from my friends and family. To my parents, Mom Passi and Dad Passi, thank you for keeping my spirits high. To my sister, Gagan, thank you for being there to listen. To our puppy, Som, thank you for sitting beside me as I wrote my thesis and knowing when it was the right time to get up for a walk. To my wife, Yashi, I would not have been able to do this without your loving encouragement every step of the way. Thank you for always being there for me and keeping me held together until the finish line.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CAN*Help*	Canadian North *Helicobacter pylori*
^{13}C	Carbon-13
^{14}C	Carbon-14
CI	Confidence interval
CO_2	Carbon dioxide
DOB	Delta-over-baseline
ELISA	Enzyme-linked immunosorbent assay
Hp	*Helicobacter pylori*
ICC	Intraclass correlation coefficient
IHC	Immunohistochemistry
LR	Likelihood ratio
MNC	Mononuclear cells
NPV	Negative predictive value
NSAID	Non-steroidal anti-inflammatory drug
OR	Odds ratio
PCR	Polymerase chain reaction
PNC	Polymorphonuclear cell
PPI	Proton pump inhibitor
PPV	Positive predictive value
RUT	Rapid urease test
SAT	Stool antigen test
UBT	Urea breath test

Chapter 1: Introduction

1.1 Background

Helicobacter pylori (Hp), a fastidious Gram-negative bacterium, was first identified by Barry Marshall and Robin Warren in 1982[1] as an organism colonizing the gastric mucosa, which often persists indefinitely. Since its discovery, *Hp* infection has been identified as a cause of gastritis, gastroduodenal ulcers, and gastric cancer, including gastric carcinoma and some types of gastric lymphoma. *Hp* infection is classified as a Class I carcinogen by the International Agency for Research on Cancer (IARC)[2] and has been recognized as a public health concern for high-prevalence populations, including Northern Canadian communities. Individuals diagnosed with *Hp* infection can undergo treatment to eliminate the infection and thereby reduce the risk of diseases caused by chronic *Hp* infection[3].

The Canadian North *Helicobacter pylori* (CAN*Help*) Working Group was formed to respond to the concerns of Arctic Indigenous communities about health risks from *Hp* infection. The CAN*Help* Working Group links communities with northern healthcare officials and the University of Alberta scientists to conduct research focused on community priorities. This team launched community projects in the Northwest Territories and Yukon to address community-driven research questions. A planning committee made up of community representatives guided each project. These projects aimed to describe the local burden of disease from *Hp* infection, conduct policy analysis to develop clinical management strategies that are cost-effective and culturally appropriate for northern communities and conduct knowledge exchange activities to help participants understand available solutions and unsolved challenges for reducing health risks from this infection.

Project staff offered CAN*Help* community projects participants screening for *Hp* infection using the non-invasive urea breath test (UBT)[4]. Some participants also consented to undergo endoscopy, during which endoscopists collected gastric biopsies. The endoscopy sampling protocol used in CAN*Help* endoscopy activities outlined the collection from prespecified locations in the stomach of 2 biopsies processed for microbiological culture and 4-6 biopsies processed for histopathological assessment by a single pathologist using the updated Sydney system to grade *Hp* density, inflammation, atrophy, and neoplasia[5]. The initial standard histopathology protocol called for duplicate slides (one stained with hematoxylin and eosin (H&E) and one with Giemsa). Giemsa-stained gastric tissue slides were used to detect and quantify *Hp* in gastric tissue samples.

The pathologist's classification of *Hp* presence or absence in the Giemsa-stained gastric tissue slides was not fully concordant with the results of microbiological culture and/or UBT. For discordant results, the CAN*Help* Working Group uses an algorithm based on the probability that *Hp* status is negative or positive[6]. While perfect concordance would not be expected, the pattern of discrepancy is not random. A non-negligible number of participants who were *Hp*-negative by histopathology had *Hp* confirmed by 16s rRNA PCR amplification cultured from gastric tissue. This discrepancy raises the question of whether there is a need to use a more accurate staining method for detecting *Hp* infection in community projects.

1.2 Significance of Research

The Gastrointestinal Pathology Society and other expert bodies recommend using ancillary staining methods only when biopsies show chronic or active gastritis without detectable *Hp* in H&E stains[7–10]. Other experts, however, recommend that ancillary stains be ordered

routinely along with H&E or used alone in all gastric biopsies to be evaluated for *Hp* to identify infected individuals more accurately so they can be treated to eliminate *Hp* infection[11–13]. Assessing the accuracy of alternative staining methods may identify a cost-effective alternative to Giemsa staining for detecting difficult-to-visualize bacteria in gastric tissue slides, especially in cases with a patchy distribution of *Hp*. It will also reveal the extent to which the bacterial density classification varies by staining method. Assessing the validity of CAN*Help Hp* density classifications will facilitate an investigation of the hypothesis that the frequency of severe gastritis increases as *Hp* density increases.

Considering the high global mortality from stomach cancer, a 2014 IARC *Hp* Working Group recommended that countries explore the possibility of introducing population-based *Hp* screening and treatment programs[14]; this panel cautioned, however, that decisions about whether and how to implement such programs must be based on local considerations of disease burden and other factors of relevance to health policy[14]. Identifying local indicators of high risk for targeted interventions is key for effective cancer prevention programs. Thus, understanding the relationship between *Hp* density and gastritis severity in specific populations at elevated risk of stomach cancer has great public health significance. For example, given that *Hp* infection can be particularly resistant to treatment in high-prevalence populations, if *Hp* density substantially impacts gastritis severity, cost-effective gastric cancer prevention could likely be achieved with therapies aimed at reducing the bacterial load in cases that are difficult to cure.

1.3 Specific Aims

The specific aims of this thesis are to:

1. Critically review the current literature on (1) the accuracy of UBT, microbiological culture and histopathological evaluation to detect *Hp* infection; (2) the correlation between UBT delta-over-baseline values and *Hp* density; and (3) the relationship of *Hp* density to gastritis severity.
2. Investigate whether incorporating ancillary stains (Giemsa, Warthin-Starry and/or immunohistochemical (IHC) staining methods) into the histopathological assessment of gastric biopsies improves the accuracy of *Hp* detection compared to H&E staining alone using gastric tissue samples from CAN*Help* community project participants.
3. Investigate the hypothesis that the frequency of severe gastritis increases as *Hp* density increases by estimating the association between *Hp* density and the severity of either active or chronic gastritis in CAN*Help* community project participants.

1.4 Thesis Structure

The remainder of my thesis contains a literature review, two papers and a conclusion. Chapter 2 contains a review that critically assesses the literature on current methods used to detect *Hp* infection, reported estimates of the specificity and sensitivity of different staining methods for detecting *Hp* in gastric biopsies, the accuracy of histopathological and UBT-based assessments of *Hp* density, and the relationship of *Hp* density to gastritis severity. Chapter 3 contains a paper entitled "Comparing the accuracy of staining methods for the detection of *Helicobacter pylori* in gastric biopsies," which assesses the agreement and validity of *Hp* status *classifications* used in CAN*Help* community projects, incorporating information from a validation study that included multiple ancillary stains. Chapter 4 contains a paper entitled "Investigating the effect of *Helicobacter pylori* density on the severity of gastritis in Arctic

Indigenous communities," which reports the results of an analysis that estimates the association between *Hp* density and the severity of each of two types of gastritis: active and chronic. Chapter 5 contains the conclusion, in which I summarize my thesis research and provide directions for future work.

Chapter 2: Literature Review

2.1 Comparing the accuracy of tests used to diagnose Hp infection

Accurately diagnosing *Hp* infection is essential for identifying individuals who need treatment. There is no optimal gold standard for detecting *Hp* infection[15]. The choice of test method depends on several factors, including the clinical question of interest, the accuracy of the method, and its availability in the target population[15]. *Hp* can be detected using different diagnostic tests ranging in invasiveness. Less invasive detection methods include but are not limited to, urea-breath tests (UBT), stool antigen tests (SAT), and serological testing for *Hp*-specific immunoglobulin G (IgG) antibodies. UBT and SAT can identify active infection, while serology does not differentiate active from past infection[16].

More invasive techniques require a gastric tissue biopsy using endoscopy procedures. *Hp* infection can be diagnosed by microbiological culture of gastric biopsy tissue can use non-selective (e.g., blood agar, etc.) or selective (e.g., brain heart infusion (BHI) agar, etc.) media incubated under microaerophilic conditions[16]. Catalase, oxidase, and urease testing can confirm *Hp* bacteria identification during microbiological culture. *Hp* bacteria in gastric biopsies can be detected using polymerase chain reaction (PCR) methods or fluorescence in-situ hybridization (FISH)[16]. *Hp* bacteria in gastric biopsies can also be visualized on histological sections.

Histological sections must be stained to permit clear visualization of cells. Stains can be nonspecific or specific, either staining most cells the same way or selectively staining certain types of cells differently. H&E is a common non-specific stain that reveals the overall cellular structure in gastric biopsy sections and facilitates pathological evaluation. Special stains can be used along with H&E to selectively stain *Hp* in gastric biopsies to facilitate the identification of the bacteria, especially in cases where there is a patchy *Hp* distribution, which is often the case

for this organism[10]. Special stains used along with H&E are often called ancillary stains. Pathologists commonly evaluate stained slides prepared from gastric biopsy tissue using the updated Sydney system[5] as a grading system to classify gastritis severity. Pathology reports based on the updated Sydney system[5] characterize the tissue in terms of the stomach subsite and cell types, as well as the presence and amount of *Hp*, chronic inflammation, active inflammation, glandular atrophy, intestinal metaplasia, dysplasia, and carcinoma. Amounts are classified on an ordinal scale (none, mild, moderate, marked). An ordinal evaluation of *Hp* density (mild, moderate, marked) is reported in cases where *Hp* is present.

I conducted a literature review to summarize validation studies that have estimated agreement and diagnostic accuracy of the UBT, microbiological culture, and/or histopathology for classifying *Hp* infection status. I also extracted information on factors (such as characteristics of hosts, bacterial colonization, or test protocols) that may modify the estimated agreement and/or accuracy of these detection methods in published reports. I searched the PubMed/Medline database to find articles indexed through 2021. The search terms used for the literature review were:

1. *Helicobacter pylori*:

   "*Helicobacter pylori*"[Mesh] OR "Helicobacter Infections"[Mesh]

2. UBT:

   UBT OR urea breath test

3. Histology:

   "Histology"[Mesh] OR "Gastritis, Atrophic"[Mesh] OR "Staining and Labeling/anatomy and histology"[Mesh] OR “Staining and

Labeling/classification"[Mesh] OR “Staining and Labeling/epidemiology"[Mesh] OR “Staining and Labeling/statistics and numerical data"[Mesh] OR “Staining and Labeling/therapeutic use"[Mesh] OR “Staining and Labeling/trends"[Mesh]

4. Culture:

   "Culture Techniques/epidemiology"[Mesh] OR “Culture Techniques/immunology"[Mesh] OR “Culture Techniques/microbiology"[Mesh] OR “Culture Techniques/statistics and numerical data"[Mesh] OR “Culture Techniques/trends"[Mesh]

5. #2 OR #3 OR #4
6. Validation study:

   Validat* OR "Validation Studies as Topic"[Mesh] OR "Validation Study" [Publication Type]

7. #1 AND #5 AND #6

I restricted my review to both original research and review articles that were available in English. With the aid of a research assistant, I reviewed the title and abstract of all publications included in the search results. I included validation studies that evaluated at least one of the index tests (UBT using either ^{13}C or ^{14}C isotopes, microbiological culture, and histopathological evaluation using diverse staining techniques) against a different one of these tests as a reference test. I reviewed all available full-text articles and either extracted information or excluded the article if it did not compare at least one of the index tests against a different one of these tests.

Following this initial search, I conducted a sensitivity analysis to search for any reports of studies that estimated the specificity and sensitivity of histopathological evaluation for detecting

Hp in gastric biopsies missed by the original search strategy. I repeated the literature review search strategy using "Sensitivity AND Specificity" for search term #6. I reviewed the title and abstract of all search results. I included validation studies that evaluated histopathological evaluation against either culture or UBT as a reference test. I reviewed all available full-text articles and either extracted information or excluded the article if it did not compare histology against either culture or UBT as a reference test.

The geographic regions of Asia and Europe are overrepresented in the 24 articles selected for data extraction. Eight of 24 articles (33%) report studies conducted in Asia, of which the largest representation was from Pakistan and Korea, with 2 studies each. Meanwhile, 9 of 24 articles (38%) report European studies from different countries, including England, Sweden, Germany, Slovakia, Latvia, Italy, Greece, and the Netherlands. A study from Turkey could be considered to represent either Europe or Asia. The geographic region of North America is underrepresented, with 1 article from Canada, 1 from the USA, and 2 from Mexico. A collaborative study was done in Canada and Korea, representing North America and Asia. The geographic region of Africa is underrepresented, with 1 article from Egypt. Studies from South America and Oceania were absent from the selected articles.

All articles report a validation study which enrolled patients through one or more healthcare facilities in the study location. All validation studies used cross-sectional analysis; some were part of a larger cohort or case-control study, intended to validate the testing method against a selected gold standard. Sampling strategies for most studies selected patients from referrals for gastroenterology consults and endoscopy follow-ups. Four of 24 studies (17%) selected patients from pediatric populations (<18 years of age), while the remainder selected only

adult patients. The sex distribution of patients was not reported in 5 of 24 articles (21%); of the studies that reported this, most had roughly equivalent numbers of male and female patients.

Exclusion criteria for most studies included the recent use of proton pump inhibitors (PPI), nonsteroidal anti-inflammatory drugs (NSAID), and/or antibiotics within a specified period before enrollment. Herold et al. (2002)[17] only included recent antibiotic use in their exclusion criteria. Reilly et al. (1997)[18] imposed no exclusion criteria; therefore, there were patients enrolled in the study who had previous *Hp* elimination therapy or were currently taking PPIs or NSAIDs. It was the only study that did a sub-analysis, including and excluding subjects with recent PPI, NSAID and antibiotic treatment. The results of this sub-analysis showed that these exclusions led to a reduction in the number of indeterminate UBT results, maintaining the same sensitivity but improving the specificity of test methods[18].

Among studies described in the selected articles, Roma-Giannikou et al. (2010)[19] conducted the only case-control study that aimed to validate a rapid urease test (RUT). This case-control study selected 530 *Hp* positive cases and 1,060 *Hp* negative controls from a pediatric population. A positive *Hp* infection was defined by either a positive culture or two other positive tests (histology and RUT, or UBT when histology and RUT were discordant). Only 182 patients were tested using culture, and 154 (84.6%) were culture positive. The sensitivity and specificity of histology were estimated using culture results from these 182 patients and reported to be 90.3% and 100%, respectively. Information on recent antibiotic or PPI use was not collected for either cases or controls. The use of antibiotics or PPIs could inhibit *Hp* growth and reduce the sensitivity of the gold standard by increasing false negative culture results.

Hp testing protocols varied across studies. Each study used distinct protocols for UBT, culture and histology. Four of 24 studies (17%) used ^{14}C UBT, while the remainder used ^{13}C

UBT. Culture was performed in 16 of 24 studies (67%), and protocols differed on the type of growth media used. Histological evaluation was performed in all studies included, of which 9 studies used a combination of H&E and Giemsa stains, 7 used Giemsa stain alone, 2 used a combination of H&E and Warthin-Starry stains, 2 used Warthin-Starry stain alone, 1 used a combination of Giemsa and IHC, and 1 did not specify the stain used. Three of 24 studies (13%) compared different histological stains.

A few studies used a single test method as the gold standard. UBT alone was used as a gold standard in 2 studies. Histology alone was used as a gold standard in 4 studies. Culture alone was used as a gold standard in 2 studies. The remaining studies created gold standards from multiple test methods. The resulting sensitivity and specificity of combining multiple tests depends on how the results are combined. Some studies defined a gold standard positive result as positive results on 2 or 3 test methods. Requiring multiple positive test results for a positive gold standard classification generally reduces the frequency of false positives arising from the gold standard classification. Meanwhile, other studies used either a positive culture or 2 other positive tests with a negative culture to define a gold standard positive classification. Defining a gold standard positive as a positive on either one test (or set of tests) or another test (or set of tests) generally reduces the frequency of false negatives arising from the gold standard classification.

Validation and/or optimization of UBT protocols was the most common study objective, characterizing 12 of 24 studies (50%). These studies aimed to identify the optimal UBT value cut point, with optimal defined in varied manners as the point with the highest sensitivity, specificity, and/or diagnostic accuracy, to identify patients with and without *Hp* infection. UBT cut points were selected using cluster analysis, as in Mauro et al. (2006)[20], or receiver operating characteristic (ROC) curves, as in Herold et al. (2002)[17] and Ortiz-Olvera et al. (2007)[21], which

both compared the sensitivity and specificity of different cut points against the study's gold standard.

All UBT validation studies concluded that UBT was as or more sensitive than the gold standard, except for Kwon et al. (2015)[22]. Kwon et al. (2015)[22] conducted a retrospective analysis to evaluate the diagnostic accuracy of their local UBT assay cut point on patients following *Hp* infection treatment. A cohort of 14,972 patients underwent UBT screening in the study period; the study selected 1,891 patients who had UBT testing and endoscopy after treatment to eliminate *Hp* infection. The gold standard positive was defined as a positive result on two of the following tests: RUT, culture, or histology. Compared to the study's gold standard, the estimated sensitivity of UBT was 99.3% (95% CI 96.4-100%), while the estimated specificity was only 47.1% (95% CI 38.6-55.8%) (95% confidence intervals calculated from information in report). A higher frequency of apparently false positive UBT results occurred in patients with multiple prior therapies to eliminate *Hp* infection and those with moderate to marked gastric intestinal metaplasia[22]. Kwon et al. (2015)[22] also attributed the low specificity of ^{13}C-UBT results to their UBT protocol, which did not administer the labelled urea with citric acid, which is commonly included in UBT protocols. Using citric acid to administer ^{13}C-urea is hypothesized to increase the amount of labelled CO_2 in the breath after the administration of ^{13}C-urea through several mechanisms[22], including increasing *Hp* urease activity and reducing non-*Hp* urease by reducing pH and delaying gastric emptying. Evidence from studies that have investigated this hypothesis, however, is inconsistent[4,23].

Mauro et al. (2006)[20] conducted a cohort study, which analyzed the results of UBT performed on 2,232 outpatients. *Hp* infection was diagnosed in 1,209 patients, and 1,023 patients of these patients were tested to confirm treatment success following *Hp* treatment. Cluster

analysis was performed on the data from this cohort to select the optimal UBT cut point. These cut points were then validated using a group of 176 patients who underwent endoscopic biopsy to allow comparison of the UBT result against histology and culture. The optimized UBT demonstrated 100% sensitivity and 98.5% specificity compared with histology and microbiology as a gold standard.

Cohen's kappa has known performance limitations, and its values lack a straightforward interpretation[24]. Only two studies calculated and reported kappa as a measure of agreement. Li et al. (2017)[25] conducted a cohort study that analyzed the UBT results of 21,857 subjects, of whom 300 were randomly selected for endoscopy for a validation sub-study. The sensitivity and specificity of two different UBT cut-off values were compared to histology results. A kappa value of 0.866 was calculated by comparing UBT to histology results. Meanwhile, in a cohort study conducted by Storskrubb et al. (2005)[26], a serology validation sub-study was conducted to compare serology results to histology and culture results; the sub-study enrolled 1,000 of 3,000 subjects who had been selected by inviting every 7th adult by birth date in the enumerated census data for two Swedish towns. These 3,000 people were invited to Questionnaire respondents were contacted by telephone at random and invited to undergo endoscopy until 1,000 participants were enrolled in the sub-study. The validation study analysis included a sub-analysis that compared the concordance of culture and histology results and yielded a kappa value of 0.96 (95% CI 0.94-0.98). Given the high kappa value, a gold standard positive was defined as a positive result on either culture or histology. Both studies that used kappa report the proportion of subjects who tested positive and negative by UBT or culture compared to histology, which aids in the interpretation of their measure of agreement.

Shin et al. (2009)[27] selected 651 subjects who had undergone gastroscopy and were enrolled in a hospital-based case-control study on gastric cancer. In addition to comparing culture and histology to a gold standard, each method's sensitivity, specificity, and predictive values were estimated within strata of histologically classified grades of atrophic gastritis and intestinal metaplasia (none, mild, or moderate/marked). The overall accuracy of both culture and histology decreased in increasing strata of atrophic gastritis and intestinal metaplasia[27]. The authors concluded that at least two tests are necessary to diagnose *Hp* infection, especially in patients with atrophic gastritis and intestinal metaplasia.

The sensitivity analysis identified studies that reported estimates of the specificity and sensitivity of different staining methods for detecting *Hp* in gastric biopsies. Laine et al. (1997)[28] and Jonkers et al. (1997)[29] both reported studies comparing staining methods to culture alone as a gold standard and acknowledged the limitation of using it as a gold standard due to the vulnerability of culture to false negative results. Meanwhile, van der Wouden et al. (1999)[30] reported comparing histology stains against a combination of tests. Both van der Wouden et al. (1999)[30] and Laine et al. (1997)[28] reported that Giemsa was more specific than H&E staining alone. Jonkers et al. (1997)[29] reported that IHC was more sensitive and specific than both Warthin-Starry and Giemsa stains. It is important to note that these earlier studies did not report the use of automated staining methods. This could have resulted in slides of poorer and less consistent quality compared to modern staining platforms.

In summary, the geographic areas of Asia and Europe were overrepresented in the articles identified by the search strategy. Reports of UBT validation studies generally concluded that UBT was as sensitive as histology or culture. Reports of other validation studies concluded that culture and histology were not as sensitive as UBT. There was heterogeneity in specificity

estimates for UBT when the gold standard was composed of biopsy-based methods. Observations of low specificity of UBT are consistent with low sensitivity of biopsy-based methods due to the limitations of biopsy sampling, including its vulnerability to false negative results arising from patchy colonization[10]. Different studies used different histology staining methods, but very few of the identified articles estimated the sensitivity or specificity of specific stains or assessed the agreement of different stains. Other than changes in UBT cut points, the identified study reports did not mention factors that might modify the estimated agreement and/or accuracy of detection methods, such as characteristics of hosts or bacterial colonization, or variations in test protocols.

Figure 1: Total number of articles identified by the search strategy and articles excluded for Chapter 2.1

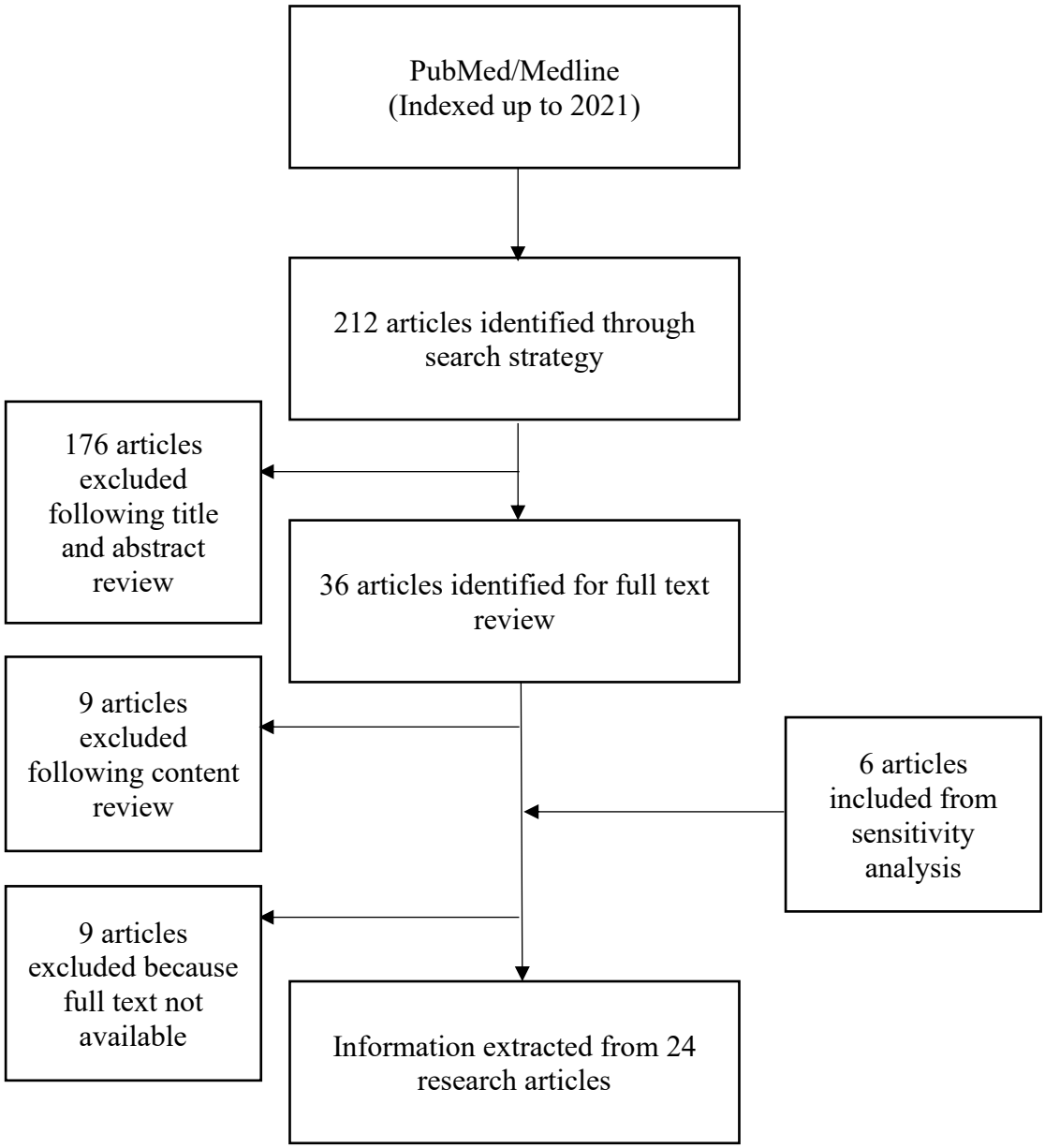


Table 1: Summary of studies that compared urea breath test, microbiological culture, and histopathology, used alone or in combination, for classifying *Hp* infection status

Authors (Year)	**Country**	**Population**	**Gold Standard Used and *Hp* Prevalence**	**Tests**			
				UBT	Culture	Histology	Other(s)
Šeligová et al. (2020)[31]	Slovakia	81 patients (36 male, 45 female)	UBT - 39.5% (32/81)	^{13}C		Giemsa	RUT, SAT, PCR, nested PCR
Results: Compared to UBT, histology demonstrated 53.1% SE, 100% SP, 100% PPV, and 74.6% NPV **Conclusion:** Histology is less sensitive for classifying *Hp* infection status than UBT							
Li et al. (2017)[24]	China	300 patients (133 male, 167 female)	UBT - 67% (201/300)	^{13}C		Giemsa	IgG ELISA
Results: Compared to histology, optimized UBT demonstrated 94.53% SE and 89.90% SP **Conclusion:** UBT is as sensitive for classifying *Hp* infection status as histology							
Kwon et al. (2015)[21]	Korea	289 patients (177 male, 112 female)	Histology & culture - 100% (289/289)	^{13}C	BHI agar	Giemsa	RUT
Results: Compared to histology and culture, UBT demonstrated 99.3% SE, 47.1% SP, 67.3% PPV, and 98.5% NPV **Conclusion:** UBT is as sensitive, but not as specific, for classifying *Hp* infection status post-eradication therapy than histology and culture							

Authors (Year)	Country	Population	Gold Standard Used and *Hp* Prevalence	Tests			
				UBT	Culture	Histology	Other(s)
Sudraba et al. (2011)[32]	Latvia	119 patients (28 male, 91 female)	Combo** - 87.1% (104/119)	^{13}C	Not specified	H&E and Giemsa	RUT, IgG ELISA
Results: Compared to combo*, histology demonstrated 100% SE and 100% SP, UBT demonstrated 96% SE and 100% SP **Conclusion:** UBT is as sensitive for classifying *Hp* infection status as histology							
Roma-Giannikou et al. (2010)[18]	Greece	530 pediatric cases (254 male, 276 female) and 1060 controls (526 male, 534 female)	Combo* - *Hp* prevalence 50% (530/1060)	^{13}C	Thioglycollate medium	H&E and Giemsa	RUT
Results: Compared to combo*, histology demonstrated 90.3% SE and 100% SP **Conclusion:** Histology is as sensitive for classifying *Hp* infection status as culture in combination with other tests							
Shin et al. (2009)[27]	Korea	651 patients (397 male, 254 female)	Combo* - 64.5% (420/651)		BHI agar	Giemsa and IHC	IgG ELISA
Results: Compared to combo*, culture demonstrated 56.2% SE (95% CI 51.1-60.6%), 100% SP (95% CI 96.9-100%), 100% PPV and 40.8% NPV while histology demonstrated 93.0% SE (95% CI 90.3-95.0%), 94.0% SP (95% CI 88.6-97.1%), 98.8% PPV and 59.8% NPV **Conclusion:** Histology is more sensitive for classifying *Hp* infection status than culture alone							

Authors (Year)	Country	Population	Gold Standard Used and *Hp* Prevalence	Tests			
				UBT	Culture	Histology	Other(s)
Ortiz-Olvera et al. (2007)[21]	Mexico	88 patients (39 male, 49 female)	Combo* - 57.95% (51/88)	^{13}C	Trypticasein-soy agar	H&E and Giemsa	IgG ELISA
Results: Compared to combo*, UBT demonstrated 90.2% SE, 93.3% SP, 95.83% PPV and 84.8% NPV **Conclusion:** UBT is as sensitive classifying *Hp* infection status than culture and other tests used in combination							
Rasool et al. (2007)[33]	Pakistan	94 patients (60 male, 34 female)	Histology – 70% (66/94)	^{14}C		H&E and Giemsa	RUT
Results: Compared to histology, UBT demonstrated 92% SE (95% CI 87-95%), 93% SP (95% CI 79-99%), 97% PPV (95% CI 91-99%) and 84% NPV (95% CI 72-89%) **Conclusion:** UBT is as sensitive for classifying *Hp* infection status as histology							
Bilal et al. (2007)[34]	Pakistan	90 patients (79 male, 11 female)	Combo** - prevalence not reported	^{13}C	Columbia agar	H&E and Giemsa	RUT
Results: Compared to combo**, UBT demonstrated 100% SE, 82% SP, 90% PPV and 93.3% accuracy, culture demonstrated 70% SE, 100% SP, 100% PPV and 81.1% accuracy, and histology demonstrated 98% SE, 74% SP, 86% PPV and 88.9% accuracy **Conclusion:** UBT is as sensitive for classifying *Hp* infection status as histology							

Authors (Year)	Country	Population	Gold Standard Used and *Hp* Prevalence	Tests			
				UBT	Culture	Histology	Other(s)
Mauro et al. (2006)[20]	Canada	176 patients	Histology & culture – prevalence not reported	^{13}C	Brucella agar, Mueller-Hinton agar, and egg yolk emulsion	Warthin-Starry	RUT
Results: Compared to histology and culture, UBT demonstrated 100% SE and, 98.5% SP, 94.5% PPV and 100% NPV **Conclusion:** UBT is as sensitive for classifying *Hp* infection status as histology and culture							
Frenck et al. (2006)[35]	Egypt	100 pediatric patients	Combo* - 46% (46/100)	^{13}C	Columbia agar	H&E and Giemsa	RUT, SAT, IgG ELISA
Results: Compared to combo*, UBT demonstrated 98% SE, 89% SP, 88% PPV and 98% NPV **Conclusion:** UBT is as sensitive for classifying *Hp* infection status as histology and culture							
Pellicano et al. (2005)[36]	Italy	46 patients (18 male, 28 female)	Combo*** - 65.2% (30/46)	^{13}C	Not specified	Not specified	RUT, SAT, IgG ELISA
Results: Compared to combo***, culture demonstrated 96.7% SE, 100% SP, 100% PPV, 94.1% NPV and 97.8% accuracy **Conclusion:** Culture is as sensitive for classifying *Hp* infection status as other tests used in combination							

Authors (Year)	Country	Population	Gold Standard Used and *Hp* Prevalence	Tests			
				UBT	Culture	Histology	Other(s)
Storskrubb et al. (2005)[26]	Sweden	1,000 patients (49% male, 51% female)	Histology or culture - 33.9% (339/1000)		Columbia agar	H&E and Warthin-Starry	IgG ELISA
Results: Compared to culture, histology demonstrated 95.8% SE and 99.6% SP **Conclusion:** Histology is as sensitive for classifying *Hp* infection status as culture							
Oztürk et al. (2003)[37]	Turkey	75 patients (19 male, 56 female)	Histology - 65% (48/75)	^{14}C†		Giemsa	RUT
Results: Compared to histology, UBT† demonstrated 100% SE, 80% SP, 90% PPV, 100% NPV and 93% accuracy **Conclusion:** UBT† is as accurate for classifying *Hp* infection status as histology							
Herold et al. (2002)[17]	Germany	251 pediatric patients (135 male, 116 female)	Histology - 53% (133/251)	^{13}C		Giemsa	
Results: Compared to histology, validated UBT demonstrated 95% SE, 97% SP, 97% PPV, and 95% NPV **Conclusion:** UBT is as sensitive for classifying *Hp* infection status as histology							

Authors (Year)	Country	Population	Gold Standard Used and *Hp* Prevalence	Tests			
				UBT	Culture	Histology	Other(s)
Shirin et al. (2000)[38]	Israel	97 patients (41 male, 56 female)	Histology & RUT - 47.4% (46/97)	^{13}C		Giemsa	RUT
Results: Compared to histology & RUT, UBT demonstrated 97.8% SE, 96.1% SP, 95.7% PPV, and 98.0% NPV **Conclusion:** UBT is as sensitive for classifying *Hp* infection status as histology							
Yanez et al. (2000)[39]	Mexico	59 pediatric patients (20 male, 39 female)	Combo*** - 37.3% (22/59)	^{13}C	Skirrow agar	H&E and Giemsa	RUT
Results: Compared to combo***, UBT demonstrated 90.9% SE and 91.9% SP, culture demonstrated 77.3% SE and 100% SP, and histology demonstrated 81.8% SE and 94.6% SP **Conclusion:** UBT is as sensitive for classifying *Hp* infection status as histology and culture							
Mock et al. (1999)[40]	Korea & Canada	205 total patients (97 male, 108 female)	Histology & RUT – 44.9% (92/205)	^{13}C		Giemsa	RUT
Results: Compared to histology and RUT, UBT demonstrated 93.5% SE (95% CI 88.5-98.5%), 97.3% SP (95% CI 94.3-100%), and 95.6% accuracy (95% CI 92.8-98.4%) **Conclusion:** UBT is as sensitive for classifying *Hp* infection status as histology							

Authors (Year)	Country	Population	Gold Standard Used and *Hp* Prevalence	Tests			
				UBT	Culture	Histology	Other(s)
Yu et al. (1999)[41]	Singapore	79 patients	Histology - 53.2% (42/79)	^{14}C		H&E and Warthin-Starry	RUT
Results: Compared to histology, UBT demonstrated 100% SE (95% CI 92-100%), 97.2% SP (95% CI 86-100%), 100% PPV (95% CI 90-100%) and 97.7% NPV (95% CI 88-100%) **Conclusion:** UBT is as sensitive for classifying *Hp* infection status as histology							
van der Wouden et al. (1999)[30]	Netherlands	197 patients	Combo** - 79.2% (156/197)	^{13}C	Belo-horizonte medium	H&E alone and Giemsa alone	RUT, PCR
Results: Compared to combo**, H&E alone demonstrated 94% SE and 95% SP, Giemsa alone demonstrated 81% SE and 99% SP, and culture had 100% SE and 100% SP **Conclusion:** Culture is more sensitive for classifying *Hp* infection status than histology, and Giemsa alone is more specific than H&E staining alone							
Maconi et al. (1999)[42]	Italy	115 patients (59 male, 56 female)	Combo** - 70.4% (81/115)	^{13}C	Not specified	H&E and Giemsa	RUT, IgG ELISA
Results: Compared to combo**, UBT demonstrated 100% SE and 100% SP, culture demonstrated 97.5% SE and 100% SP, routine histology demonstrated 92.6% SE and 93.3% SP, and expert histology demonstrated 98.8% SE and 100% SP **Conclusion:** UBT is as sensitive for classifying *Hp* infection status as histology and culture; pathologist experience impacts accuracy of histology							

Authors (Year)	Country	Population	Gold Standard Used and *Hp* Prevalence	Tests			
				UBT	Culture	Histology	Other(s)
Reilly et al. (1997)[18]	England	300 patients	Combo*** - 45.7% (137/300)	^{13}C		H&E and Giemsa	RUT, IgG ELISA
Results: Compared to combo***, UBT demonstrated 100% SE (95% CI 91-100%) and 100% SP (95% CI 91-100%) **Conclusion:** UBT is as sensitive for classifying *Hp* infection status as histology							
Laine et al. (1997)[28]	USA	101 patients (42 male, 59 female)	Culture - prevalence not reported		Skirrow and BHI agar	H&E alone and Giemsa alone	RUT
Results: Compared to culture, H&E demonstrated 92% SE and 89% SP, Giemsa demonstrated 88% SE and 98% SP **Conclusion:** Giemsa staining is more specific for classifying *Hp* infection status than H&E alone							
Jonkers et al. (1997)[29]	Netherlands	40 patients	Culture - 50% (20/40)		Blood and Skirrow agar	Giemsa alone, Warthin-Starry alone and IHC alone	
Results: Compared to culture, Giemsa demonstrated 90% SE and 53.8% SP, Warthin-Starry demonstrated 70% SE and 82.5% SP, and IHC demonstrated 83.8% SE and 90% SP **Conclusion:** IHC is more specific for classifying *Hp* infection status than Giemsa and Warthin-Starry stains							

UBT, urea breath test; H&E, hematoxylin and eosin; IHC, immunohistochemical; RUT, rapid urease test; SAT, Stool antigen test; IgG ELISA, immunoglobulin enzyme-linked immunoassay; PCR, polymerase chain reaction; SE, sensitivity; SP, specificity; PPV, positive predictive value; NPV, negative predictive value.

*positive *Hp* infection was defined by positive culture or, if culture negative, positive histology and serology
**positive *Hp* infection was defined by any three tests positive
***positive *Hp* infection was defined by any two tests positive
†breath samples were collected using two different methods (Heliprobe vs standard method); standard UBT results reported

2.2 Using UBT as an estimate for Hp density

The ^{13}C-UBT uses urea labelled non-radioactively with ^{13}C to detect *Hp*'s enzymatic activity[4,43]. If *Hp* bacteria are present, the urease they secrete will catabolize the ^{13}C-urea to form ammonia and $^{13}CO_2$. Two breath samples, one taken before ^{13}C-urea consumption and one after, can be compared to measure the change from baseline, or delta-over-baseline (DOB), in the level of $^{13}CO_2$. ^{13}C-UBT DOB values are a function of the $^{13}C/^{12}C$ ratio and range from values around 0 to around 100[4,44]. The ^{14}C-UBT follows the same concept, except it uses ^{14}C instead of ^{13}C to detect *Hp*'s enzymatic activity. The measurement of ^{14}C differs from ^{13}C and can be expressed as a rate of counts per unit of time, like counts per minute (CPM), as registered by a radiation monitoring instrument.

As occurs with bacterial infections in general, the number of bacteria that colonize the stomach of an infected person, referred to as the bacterial load or density, varies greatly, with implications for disease and treatment outcomes, as well as the immune response. Methods for bacterial density assessment in the literature include histology, UBT, PCR and semi-quantitative culture. The UBT has been theorized to provide a quantitative measure of *Hp* density, assuming its numeric DOB is a function of the amount of gastric urease, that this amount increases with the bacterial load, and that *Hp* bacteria are the only source of urease in the stomach.

In the ideal UBT reaction, all *Hp* urease present would be saturated by labelled urea, and steady hydrolysis would be reached. The DOB value would therefore reflect total gastric urease activity[45]. If total gastric urease activity accurately reflects *Hp* density, it could be a valuable and non-invasive screening technique compared to methods requiring endoscopy. However, accurate assessment of *Hp* density by the UBT alone may not be possible for multiple reasons: CO_2 production varies widely across individuals in a manner related to age- and sex-dependent

metabolic variation; the type of test meal may further vary CO_2 production[46]; there may be other urease-secreting bacteria in the stomach, as occurs more frequently during disease processes involving hypochlorhydria; and there may be non-pylori helicobacters, as occurs in a small fraction of cases, with the precise frequency varying across populations. Also, a wide variation in urease activity has been observed across *Hp* strains in vitro[47]. When UBT is repeated for an individual, a large difference in the delta-over-baseline values likely reflects changes in bacterial load[45].

I conducted a literature review to summarize studies that estimated the association of *Hp* density classified by histopathological evaluation with UBT DOB values. I also assessed the quality of studies examining the relationship between *Hp* density and UBT DOB values. I searched the PubMed/Medline database to find articles indexed through 2021. The search terms I used for the literature review were:

1. *Helicobacter pylori*:

   "*Helicobacter pylori*"[Mesh] OR "Helicobacter Infections"[Mesh]

2. UBT:

   UBT OR urea breath test

3. *Hp* density:

   Density OR Infiltration OR Load

4. #1 AND #2 AND #3

I restricted my review to both original research and review articles that were available in English. I reviewed the title and abstract of all search results. I included studies that evaluated *Hp*

infection using either ^{13}C or ^{14}C UBT and compared DOB values to density scores classified by histopathological evaluation. I reviewed all available full-text articles and either extracted information or excluded the article if it did not compare DOB values to *Hp* density scores. I did not include full-text articles that were not accessible either online or through the University of Alberta library service. In tables and text, I translated diverse labels for *Hp* density grades reported in articles to terms used in the updated Sydney system[5] (i.e., none, mild, moderate, marked) for consistent nomenclature.

The geographic regions of Asia and Europe are overrepresented in the 17 articles selected for data extraction. Eight of 17 articles (47%) report on studies conducted in Asia, of which 3 are from Taiwan and 2 are from Japan. Meanwhile, 6 of 17 articles (35%) report on studies conducted in Europe, of which 3 are from Germany and 2 are from Italy. The geographic regions of North and South America are underrepresented, with only 1 article from Canada and 1 from Brazil. Studies from Africa and Oceania were absent from the selected articles.

All articles report cross-sectional studies that enrolled patients through one or more healthcare facilities in the study location. Patients in these studies were referred for gastroenterology consults and endoscopy follow-ups, often presenting with GI symptoms (e.g., dyspepsia). Eligibility criteria for 5 of 17 studies (29%) included positive *Hp* infection, while other studies included *Hp*-negative patients. Only 2 of 17 studies (12%) included pediatric populations (<18 years of age), while the remainder enrolled only adult patients. The distribution of sex in the study population was not reported in 5 of 17 articles (29%); of the studies that reported them, 29% had nearly twice the number of male patients compared to female patients.

Hp testing protocols varied across studies. Each study used distinct protocols for UBT and histology except for the two studies by Perri et al. (1998)[48,49]. Only 2 of 17 studies (12%)

used ^{14}C UBT, while the remainder used ^{13}C UBT. Of the 16 studies that reported the wait time between labelled urea consumption and the second breath sample collection, 12 of 16 studies (75%) waited 30 minutes (T_{30}), while the remainder (25%) waited 15 minutes (T_{15}). Regarding staining methods used to assess *Hp* presence and density, 6 studies used a combination of H&E and Giemsa stains, 5 used a combination of H&E and Warthin-Starry stains, 2 used Giemsa stain alone, 2 used a combination of H&E and cresyl-violet, 1 used H&E alone, and 1 used a combination of hematoxylin-phloxin-safran and argentaffin stains.

The Sydney System[5] was used to grade *Hp* density in 10 of 17 studies, while 5 used an arbitrary grading system resembling the Sydney System. The remaining 2 studies (Vincent et al. (1999)[50] and Desroches et al. (1997)[51]) classified the presence or absence of *Hp* by histology and used semi-quantitative microbiological culture methods to estimate *Hp* density. Vincent et al. (1999)[50] report collecting biopsies from the antrum for histology and culture, including quantitative culture by a validated method (Intertechniques, Saint-Nom, France) which provided log bacterial enumeration. Desroches et al. (1997)[51] also report collecting biopsies from the antrum for histology and culture but used the number of colonies on the culture plate to measure density. Using samples collected at the same time and anatomical location for microbiology and histology enhances the relevance of culture for estimating *Hp* density; additionally, increasing the number of anatomical sites from which samples are collected reduces the probability of misclassification in cases with a patchy distribution of *Hp* colonization[10].

Fifteen of 17 studies (88%) estimated a positive correlation between UBT values and *Hp* density; 12 of these 15 studies used Spearman correlation coefficients as the measure of correlation. Vincent et al. (1999)[50] used the Pearson correlation coefficient. It should be noted that the Pearson correlation coefficient assumes a linear relationship, while the Spearman

correlation coefficient is appropriate for monotonic trends where the rate of increase/decrease is not constant, as occurs with ordinal categories. The Pearson correlation coefficient may be justified for the study by Vincent et al. (1999)[50], which correlated the natural log of the UBT DOB values with the log of the bacterial colonization count. Other studies that used the Spearman correlation coefficient used non-logarithmic UBT DOB values and ordinal *Hp* density scores assigned by the Sydney System[5] or an arbitrary ordinal scoring system assigned by the evaluating pathologist.

Zagari et al. (2005)[52] interpreted a comparison of means with a Student's t-test for statistical significance as indicating a positive correlation between UBT values and *Hp* density. Mean ± standard deviation (SD) UBT DOB values were 17.4 ± 7.9, 29.9 ± 14.4, and 48.7 ± 20.4 for mild, moderate, and marked density, respectively, showing a gradient of increasing mean UBT values with increasing density; p-values for mean comparisons were <0.005 for mild versus moderate, <0.05 for mild versus marked, and <0.05 for moderate versus marked. It should be noted, however, that neither the comparison of means nor the p-values estimate the strength of association; furthermore, these authors did not comment on the large SD for each estimated mean UBT DOB value.

Sheu et al. (1999)[53] used linear regression and estimated a positive correlation between UBT values and *Hp* density. R^2 was 0.7574 for T_{15} and 0.7432 for T_{30} ($p<0.0001$). Unique to this study was the collection of gastric biopsies from the antrum, body, and cardia, which were individually assessed for *Hp* density (range 0-5) and then those scores were added together to quantify the total bacterial density (range 0-15). Assessing and calculating *Hp* presence this way may increase sensitivity compared to a single biopsy from one anatomical location. However, no

analysis was done to examine variations in density across antrum, body, or cardia biopsies or if giving equal weight to each of the 3 biopsied sites makes sense.

Kobayashi et al. (2002)[54] and Epple et al. (1997)[55] reported a correlation between ^{13}C UBT DOB values and *Hp* density, including and not including *Hp* negative patients. The other 15 studies included only *Hp*-positive patients. The interpretation of the UBT DOB values is not the same for *Hp*-negative and *Hp*-positive patients. Theoretically, DOB values are random variations around 0 for *Hp* negative results, while *Hp* positive results measure the amount of bacterial urease activity. Including negative patients, therefore, has the effect of mixing the assessment of whether urease activity increases with bacterial density with whether UBT and histology agree on whether *Hp* bacteria are present. This effect may be less significant if the comparison is based on categories rather than a correlation of continuous values. Not including *Hp* negative results before computation of trend estimates in this instance is justified[56].

Kumar et al. (2001)[57] reported no correlation between UBT values and *Hp* density. Their Kruskal-Wallis test had a p-value >0.05, which they interpreted as not statistically significant. However, the reported median ^{14}C UBT counts for patients with none, mild, moderate, and marked *Hp* density assessed on samples from the corpus were 1351.6 counts per minute (CPM), 2117.5 CPM, 2460.4 CPM, and 2314.2 CPM, respectively[57], a trend that is not consistent with an interpretation of "no correlation". The authors provided no additional analysis or discussion of the UBT counts and their distribution within density categories. Thus, it is not possible to draw informative conclusions from their report.

Machado et al. (2006)[58] also reported no correlation between UBT values and histological grades for *Hp* density. Like Kumar et al. (2001)[57] they did not estimate a correlation coefficient; instead, they drew their conclusion from an analysis that used ANOVA to test the

statistical significance of a comparison of means. Of interest, they compared mean log DOB UBT values for mild, moderate, and marked histological grades assigned to *Hp*-positive patients in their analysis. The authors' rationale for treating UBT DOB values logarithmically in the analysis was that DOB values do not present in a normal distribution. Theoretically, log transformation would reduce or remove the skewness and can be justified on this basis. In this study, the ANOVA p-value for the comparison of mean log DOB UBT values across density categories was 0.42. The correlation between *Hp* density and UBT DOB values was not assessed or discussed in any other way. Therefore, the claims that there was no correlation are not substantiated.

In summary, the geographic areas of Asia and Europe were overrepresented in the articles identified by the search strategy. Eligible studies generally revealed a positive correlation between ^{13}C UBT values and *Hp* density. Investigators that reported no correlation between ^{13}C UBT values and *Hp* density did not substantiate their conclusions with informative analysis. Almost all studies included only *Hp*-positive patients; excluding *Hp*-negative results before estimating associations between ^{13}C UBT values and measures of *Hp* density is justified[56].

Figure 2: Total number of articles identified by the search strategy and articles excluded for Chapter 2.2

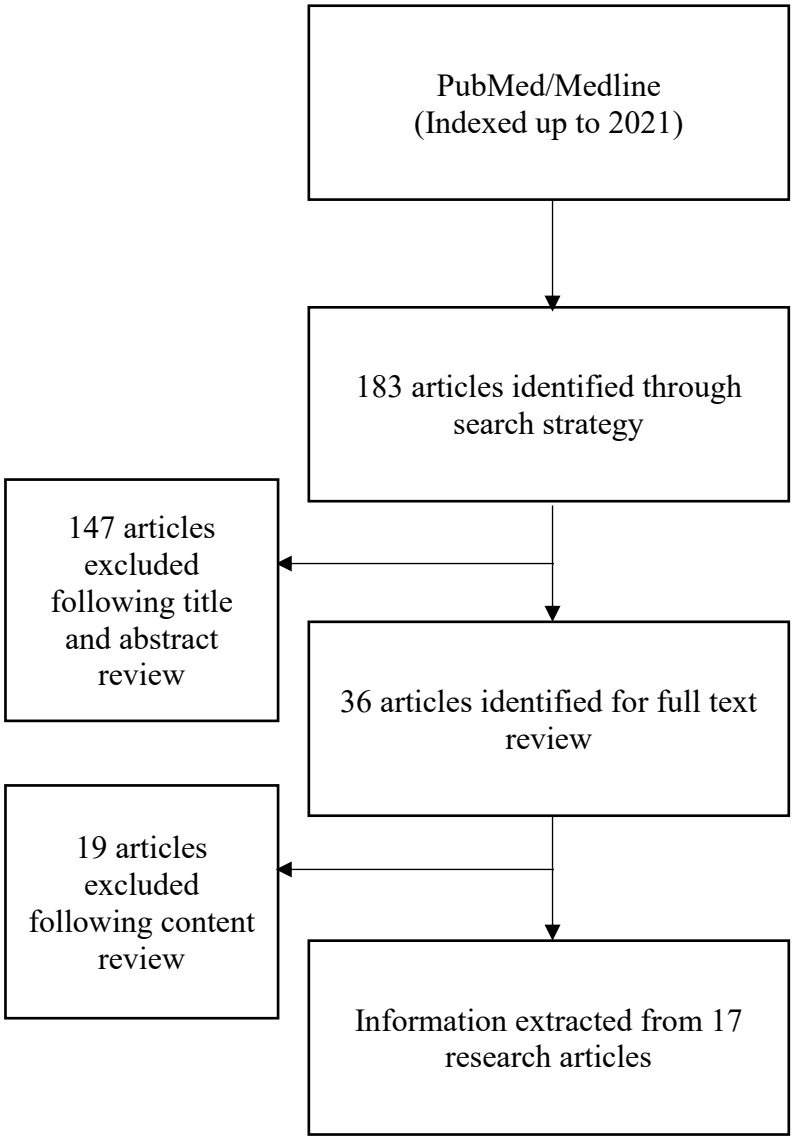


Table 2: Summary of studies comparing *Hp* density and urea breath test values

Authors (Year)	Country	Population	UBT		Histopathology		Statistical Analysis
			Isotope	Time*	Stain	Density Grades	
Rasheed et al. (2014)[59]	Pakistan	93 patients (71 male, 22 female)	^{13}C	30†	H&E, Giemsa	Updated Sydney: None Mild Moderate-Marked (combined)**	Spearman's rank correlation
Results: Mean ± SD DOB values were 10.52 ± 8.14 and 18.40 ± 15.31 for mild and moderate-marked density, respectively (r_s = 0.277, p=0.037) **Conclusion:** Positive correlation between ^{13}C UBT values and *Hp* density							
Machado et al. (2006)[58]	Brazil	44 *Hp*-positive pediatric patients (19 male, 25 female)	^{13}C	30	H&E, Giemsa	Updated Sydney: Mild Moderate Marked	ANOVA
Results: Means ± SD of log DOB were 1.545 ± 0.259, 1.407 ± 0.269, and 1.46 ± 0.175 for mild, moderate, and intense density, respectively (p=0.42) **Conclusion:** No apparent correlation between ^{13}C UBT values and *Hp* density							
Zagari et al. (2005)[52]	Italy	192 *Hp*-positive patients (98 male, 94 female)	^{13}C	30‡	H&E, Giemsa	Updated Sydney: Mild Moderate Marked	Student's t-test
Results: Means ± SD DOB were 17.4 ± 7.9, 29.9 ± 14.4, and 48.7 ± 20.4 for mild, moderate, and marked density; mild versus moderate p<0.005; mild versus marked p<0.05; moderate versus marked p<0.05 **Conclusion:** Positive correlation between ^{13}C UBT values and *Hp* density							

Authors (Year)	Country	Population	UBT		Histopathology		Statistical Analysis
			Isotope	Time*	Stain	Density Grades	
Lai et al. (2004)[60]	Taiwan	113 *Hp*-positive outpatients (73 male, 40 female)	^{13}C	15	H&E, Giemsa	Updated Sydney: Mild Moderate Marked	Spearman's rank correlation
Results: Median DOB (P_{75}-P_{25}) values were 9.6 (13.2), 17.9 (26.9) and 33.2 (25.4), for mild, moderate, and marked, respectively; r_s for the DOB and *Hp* density was 0.47 **Conclusion:** Positive correlation between ^{13}C UBT values and *Hp* density							
Chang et al. (2003)[61]	South Korea	100 *Hp*-positive patients (50 male, 50 female)	^{13}C	30	H&E, Giemsa	Updated Sydney: Mild Moderate Marked	Spearman's rank correlation
Results: Spearman's correlation coefficients were 0. 45 ($p<0.0001$) for *Hp* density in the antrum, 0.42 ($p<0.0001$) for *Hp* density in the body, and 0.16 ($p=0.11$) for *Hp* density in the cardia, and 0.53 ($p<0.0001$) for the stomach overall **Conclusion:** Positive correlation between ^{13}C UBT values and *Hp* density							
Chang et al. (2002)[62]	Taiwan	62 patients (36 male, 26 female)	^{13}C	30	H&E, Warthin-Starry	Updated Sydney: None Mild Moderate Marked	Spearman's rank correlation
Results: Spearman's correlation coefficients were 0.799 for *Hp* density in the antrum and 0.816 for *Hp* density in the corpus **Conclusion:** Positive correlation between ^{13}C UBT values and *Hp* density							

Authors (Year)	Country	Population	UBT		Histopathology		Statistical Analysis
			Isotope	Time*	Stain	Density Grades	
Kobayashi et al. (2002)[54]	Japan	88 patients (60 male, 28 female)	^{13}C	15	Giemsa updated	Updated Sydney: None Mild Moderate Marked	Spearman's rank correlation
Results: The Spearman correlation coefficient were 0.81 (p=0.001) for all patients and 0.67 (p=0.001) and when the 16 patients without *Hp* infection were excluded **Conclusion:** Positive correlation between ^{13}C UBT values and *Hp* density							
Kumar et al. (2001)[57]	India	60 patients (57 male, 3 female)	^{14}C	15	H&E, Warthin-Starry	Updated Sydney: None Mild Moderate Marked	Kruskal-Wallis test
Results: Results: Median ± SD DOB values ranged from 331.3 to 8770.5 cpm and did not correlate with *Hp* density in either the corpus or antrum (p > 0.1729) **Conclusion:** No correlation between ^{14}C UBT values and *Hp* density							
Suto et al. (1999)[63]	Japan	137 patients	^{13}C	15	Giemsa	Updated Sydney: Normal Mild Moderate Marked	Spearman's rank correlation
Results: Spearman's correlation coefficients were 0.67 (p<0.05) for *Hp* density in the antrum and 0.62 (p<0.05) for *Hp* density in the corpus **Conclusion:** Positive correlation between ^{13}C UBT values and *Hp* density							

Authors (Year)	Country	Population	UBT		Histopathology		Statistical Analysis
			Isotope	Time*	Stain	Density Grades	
Vincent et al. (1999)[50]	France	39 pediatric patients (19 male, 20 female)	^{13}C	30	H&E, Giemsa	Presence Absence	Pearson's correlation coefficient using ln DOB and log *Hp* density determined by quantitative culture
Results: Pearson's correlation coefficient was 0.63 (p=0.001) for log *Hp* density in the antrum **Conclusion:** Positive correlation between ^{13}C UBT values and *Hp* density							
Sheu et al. (1999)[53]	Taiwan	196 patients	^{13}C	15, 30	H&E	Biopsies from 3 topographical locations scored using a semi-quantitative scale (0-5); categories used based on sum of scores (0-4)	Linear regression
Results: R^2 was 0.7574 for T_{15} and 0.7432 for T_{30} ($p<0.0001$) **Conclusion:** Positive correlation between ^{13}C UBT values and *Hp* density							
Perri et al. (1998)[64]	Italy	108 *Hp* positive patients	^{13}C	30	H&E, cresyl-violet	Arbitrarily scored: 1+ (very few bacteria) 2+ (between 1+ and 3+) 3+ (a lot of bacteria)	Spearman's rank correlation
Results: Spearman's correlation coefficient was 0.76 for *Hp* density in antral biopsies **Conclusion:** Positive correlation between ^{13}C UBT values and *Hp* density							

Authors (Year)	Country	Population	UBT		Histopathology		Statistical Analysis
			Isotope	Time*	Stain	Density Grades	
Perri et al. (1998)[49]	Italy	172 outpatients	^{13}C	30	H&E, cresyl-violet	Arbitrarily scored: 0 (no bacteria) 1+ (very few bacteria) 2+ (between 1+ and 3+) 3+ (a lot of bacteria)	Spearman's rank correlation
Results: Spearman's correlation coefficient was 0.32 for *Hp* density in antral biopsies **Conclusion:** Positive correlation between ^{13}C UBT values and *Hp* density							
Ellenrieder et al. (1997)[65]	Germany	145 adult patients (71 male, 74 female)	^{13}C	30	H&E and Warthin-Starry,	Arbitrarily scored: 0 (no bacteria) 1 (a few bacteria) 2 (more than a few)	Spearman's rank correlation
Results: Pearson's correlation coefficients were 0.56 for *Hp* density in the antrum and 0. 58 for *Hp* density in the corpus **Conclusion:** Positive correlation between ^{13}C UBT values and *Hp* density							
Desroches et al. (1997)[51]	Canada	56 dyspeptic patients	^{14}C		Hematoxylin-Phloxin-Safran and argentaffin (Dieterle)	Presence Absence	Spearman's regression analysis using DOB and *Hp* density determined by semi-quantitative culture
Results: Spearman's correlation coefficient for *Hp* density in the antrum was 0.84 (95% CI, 0.72-0.91) **Conclusion:** Positive correlation between ^{14}C UBT values and *Hp* density							

Authors (Year)	Country	Population	UBT		Histopathology		Statistical Analysis
			Isotope	Time*	Stain	Density Grades	
Epple et al. (1997)[55]	Germany	126 patients (56 male, 70 female)	^{13}C	30	H&E, Warthin-Starry	Sydney system: None Mild Moderate Marked	Spearman's rank correlation
Results: Spearman's correlation coefficient was 0.78 ($p < 0.001$) including *Hp*-negative patients and 0.33 ($p < 0.01$) excluding *Hp*-negative patients **Conclusion:** Positive correlation between ^{13}C UBT values and *Hp* density							
Labenz et al. (1996)[66]	Germany	70 patients (37 male, 33 female)	^{13}C	30	H&E, Warthin-Starry	Arbitrarily scored: None Mild Moderate Marked	Spearman's rank correlation
Results: Spearman's correlation coefficient was 0.52 for *Hp* density in antral and body biopsies **Conclusion:** Positive correlation between ^{13}C UBT values and *Hp* density							

UBT, urea breath test; H&E, hematoxylin and eosin; Sydney, Sydney grading method; SD, standard deviation; CI, confidence interval

*Sample collection time in minutes following labelled urea intake
**Combined in a single category due to a smaller number of marked cases
†DOB values reported for 10-60-minute collection times, 30-minute collection time used for analysis
‡DOB values reported for 10-60-minute collection times for 52 patients, 30-minute collection time used for analysis

2.3 The relationship between Hp density and gastritis severity

Gastritis is generally described as an inflammation of the stomach lining. Gastritis can be classified based on severity, anatomical location, and histological evaluation as active or chronic. Active gastritis is characterized by polymorphonuclear cell (PNC) infiltration of the gastric mucosa[5]. Chronic gastritis includes increased mononuclear cells (MNC) (e.g., lymphocytes, plasma cells, monocytes, mast cells, and/or eosinophils) in the lamina propria and possible PNC infiltration in cases of chronic active gastritis[5]. Pathologists commonly evaluate stained slides prepared from gastric biopsy tissue and can use the updated Sydney system[5] as a grading system to classify gastritis. There are many causes of gastritis, including *Hp* infection.

I conducted a literature review to summarize studies investigating the hypothesis that gastritis severity increases as *Hp* density increases. I also assessed the quality of studies examining the relationship between *Hp* density and gastritis severity. I searched the PubMed/Medline database to find articles indexed through 2021. The search terms I used for the literature review were:

1. *Helicobacter pylori*:

   "*Helicobacter pylori*"[Mesh] OR "Helicobacter Infections"[Mesh]

2. *Hp* density:

   Density OR Infiltration OR Load

3. Gastritis Severity:

   Gastritis severity OR gastric mucosal inflammation severity

4. #1 AND #2 AND #3

I restricted my review to both original research and review articles that were available in English. I reviewed the title and abstract of all search results. I included studies that evaluated and compared *Hp* density and gastritis severity by histopathological evaluation. I reviewed all available full-text articles and either extracted information or excluded the article if it did not compare *Hp* density and gastritis severity. In tables and text, I translated diverse labels for active and chronic gastritis reported in articles to terms used in the updated Sydney system[5] (i.e., neutrophil infiltrate, mononuclear cell infiltrate) for consistent nomenclature.

The geographic regions of Asia and Europe are overrepresented in the 24 articles selected for data extraction. Nine of 24 articles (38%) are from a study location in Asia, of which 3 are from India and 2 are from Japan. Two studies from Turkey could be considered to represent either Europe or Asia. Six of the 24 articles (25%) are from a study location in Europe, of which 3 are from Italy and 2 are from England. Four of 24 studies (17%) are from a study location in North America, 3 of which are from the United States of America (USA). A collaborative study was done in the USA and Colombia, representing North and South American geographic areas. The geographic region of South America is underrepresented, with only 2 articles from Brazil. Studies from Africa and Oceania were absent from the selected articles.

All articles report cross-sectional studies which enrolled patients through one or more healthcare facilities in the study location. Patients in these studies were referrals for gastroenterology consults and endoscopy follow-ups, often presenting with GI symptoms (e.g., dyspepsia). Exclusion criteria for 17 of 24 studies (71%) included, but were not limited to, age, history of gastric surgery, and a recent history of PPI or antibiotic use. Tiwari et al. (2020)[67] reported excluding neoplastic cases or patients who received *Hp* treatment within a month of endoscopy, while the remaining 6 studies[61,68–72] did not report any exclusion criteria. Five of 24

studies included children (<18 years of age), while the remaining 19 studies were restricted to adults (≥18 years of age).

All investigators reported using the H&E stain for evaluating gastritis severity except for Taşçı et al. (2020)[73], who did not explicitly report their gastritis evaluation methods. Methods for evaluating *Hp* density varied across studies. Four studies used UBT DOB values to estimate *Hp* density. Four studies used semi-quantitative culture methods to estimate *Hp* density; all 4 compared *Hp* density estimated using culture methods to *Hp* density evaluated using histology and confirmed that results from the 2 methods were correlated. For studies that used histological staining methods to assess *Hp* presence and density, 10 used the Giemsa stain, 3 used the Warthin-Starry stain, 1 used either Giemsa and/or Warthin Starry stain(s), and 1 used Loeffler's methylene blue stain alone.

Most studies provided evidence of a positive correlation or association of *Hp* density with gastritis severity, except for 4 studies[58,68,69,71]. Nai et al. (2007)[68] analyzed gastric biopsies from 200 patients with chronic gastritis and aimed to estimate the association between *Hp* density and the combined frequency of histopathological findings, including chronic gastritis, lymphoid follicles, intestinal metaplasia, and regenerative cell atypia. Marked chronic gastritis was detected in 9.8% of mild *Hp* density cases, 3.7% of moderate *Hp* density cases, 42% of marked *Hp* density cases, and 14.4% of cases without *Hp* bacteria (p=0.03). Overall, chronic gastritis was more prevalent in *Hp*-negative cases (72.3%) than in *Hp*-positive cases (53%). No exclusion criteria were reported, nor was information collected on antibiotic use or previous *Hp* treatment; these variables could result in misclassification of *Hp* density during the relevant etiologic window of exposure; failure to account for these variables could result in underestimating the strength of the association.

Machado et al. (2006)[58] also reported no correlation between UBT values and histological grades for neutrophilic or mononuclear infiltrate. They drew their conclusion from an analysis that used ANOVA to test the statistical significance of a comparison of means. Their analysis compared mean log DOB UBT values for mild, moderate, and marked histological grades of gastritis severity. In this study, the ANOVA p-values for the comparison of mean log DOB UBT values across neutrophilic and mononuclear infiltrate categories were 0.41 and 0.29, respectively. The correlation between UBT DOB values and gastritis severity was not assessed or discussed in any other way. Therefore, the claims that there was no correlation are not substantiated. Molaei et al. (2010)[74] conducted a study to investigate correlations between different *Hp* genotypes and histological findings; they reported a correlation between *Hp* density and gastritis severity but did not actually investigate the effect of different *Hp* genotypes on this relationship.

In the selected articles, analysis of confounding or effect-modifying variables is absent. Multiple investigators reported an association or correlation of *Hp* density with gastritis severity but do not provide any additional analysis or discussion of potential confounding variables. Conclusions from these studies only suggest that pathologists evaluating gastric biopsies by histology remain vigilant for *Hp* bacteria once gastritis is detected. From an epidemiological perspective, the overall quality of these studies is poor.

Figure 3: Total number of articles identified by the search strategy and articles excluded for Chapter 2.3

Table 3: Summary of studies investigating the association of *Hp* density with gastritis severity

Authors (Year)	**Country**	**Sample Size & Prevalence**	**Test Method**		**Association of *Hp* density and Inflammation**			
			Hp Density	Gastritis Severity	Antrum		Body/Corpus	
					PNC	MNC	PNC	MNC
Tiwari et al. (2020)[67]	Nepal	250 patients - 60% (150/250)	Giemsa	H&E	n/a	n/a	-	-
Statistical testing: No statistical testing performed **Conclusion:** Positive correlation between *Hp* density and both neutrophil infiltration and mononuclear cell infiltrate								
Taşçı et al. (2020)[73]	Turkey	181 patients - 100% (181/181)	Giemsa	Not specified	p=0.001	p=0.002	"	"
Statistical testing: Pearson's $\chi 2$ **Conclusion:** Positive correlation between *Hp* density and both neutrophil infiltration and mononuclear cell infiltrate; negative correlation between *Hp* density and IM								
Garg et al. (2012)[75]	India	300 patients - 43.7% (131/300)	Giemsa	H&E	p=0.00	p=0.00	-	-
Statistical testing: $\chi 2$ **Conclusion:** Positive association of *Hp* density with both neutrophil infiltration and mononuclear cell infiltrate but not IM								

Authors (Year)	Country	Sample Size & Prevalence	Test Method		Association of *Hp* density and Inflammation			
			Hp Density	Gastritis Severity	Antrum		Body/Corpus	
					PNC	MNC	PNC	MNC
Molaei et al. (2010)[74]	Iran	166 patients - 51.8% (86/166)	Giemsa	H&E	p<0.001	p=0.003	-	-
Statistical testing: $\chi 2$ **Conclusion:** Positive correlation between *Hp* density and both neutrophil infiltration and mononuclear cell infiltrate but not IM								
Jaramillo-Rodriguez et al. (2008)[76]	Mexico	107 patients - 26.3% (40/152)	Giemsa	H&E	r_s=0.977, p<0.0001	r_s=0.387, p<0.0001	"	"
Statistical testing: Spearman's correlation coefficient **Conclusion:** Positive correlation between *Hp* density and both neutrophil infiltration and mononuclear cell infiltrate								
Tutar et al. (2008)[69]	Turkey	152 patients - 26.3% (40/152)	Giemsa	H&E	p>0.05	p>0.05	"	"
Statistical testing: $\chi 2$ **Conclusion:** No association of *Hp* density with either neutrophil infiltration and mononuclear cell infiltrate								

Authors (Year)	Country	Sample Size & Prevalence	Test Method		Association of *Hp* density and Inflammation			
			Hp Density	Gastritis Severity	Antrum		Body/Corpus	
					PNC	MNC	PNC	MNC
Nai et al. (2007)[68]	Brazil	200 patients - 58.5% (117/200)	Warthin-Starry	H&E	-	p>0.05	-	-
Statistical testing: ANOVA **Conclusion:** No association of *Hp* density with mononuclear cell infiltrate								
Tummala et al. (2007)[77]	USA	323 patients - 26.6% (86/323)	H&E, Semi-quantitative Culture	H&E	R^2=0.22, p=0.045	Not reported	Not reported	Not reported
Statistical testing: Linear regression **Conclusion:** Positive correlation between *Hp* density and neutrophil infiltration in antrum only, not in the corpus or mononuclear cell infiltrate in either antrum or corpus								
Machado et al. (2006)[58]	Brazil	44 patients	UBT	H&E	p=0.41	p=0.288	-	-
Statistical testing: ANOVA **Conclusion:** No correlation between *Hp* density and neutrophil infiltration or mononuclear cell infiltrate								

Authors (Year)	Country	Sample Size & Prevalence	Test Method		Association of *Hp* density and Inflammation			
			Hp Density	Gastritis Severity	Antrum		Body/Corpus	
					PNC	MNC	PNC	MNC
Zsikla et al. (2006)[72]	Switzerland	126 patients - 48.9% (54/126)	Giemsa & PCR	H&E	n/a	n/a	-	-
Statistical testing: No statistical testing performed **Conclusion:** Positive but weak correlation between *Hp* density and mononuclear cell infiltrate by histology but not PCR								
Zagari et al. (2005)[52]	Italy	192 patients - 100% (192/192)	Giemsa	H&E	-	-	-	G=0.59
Statistical testing: Goodman and Kruskal Gamma **Conclusion:** Positive correlation between *Hp* density and mononuclear cell infiltrate								
Chang et al. (2003)[61]	Korea	112 patients - 89.2% (100/112)	UBT	H&E	r_s=0.34, p=0.0005	r_s=0.69, p<0.0001	"	"
Statistical testing: Spearman's correlation coefficient **Conclusion:** Positive correlation between *Hp* density and neutrophil infiltration								

Authors (Year)	Country	Sample Size & Prevalence	Test Method		Association of *Hp* density and Inflammation			
					Antrum		Body/Corpus	
			Hp Density	Gastritis Severity	PNC	MNC	PNC	MNC
Gallo et al. (2003)[70]	Italy	911 patients - 38.2% (348/911)	Giemsa and/or Warthin Starry	H&E	p<0.001	-	"	"
Statistical testing: $\chi 2$ **Conclusion:** Positive correlation between *Hp* density and neutrophil infiltration								
Dzierzanowska-Fangrat et al. (2002)[71]	Italy	41 patients - 100% (41/41)	Giemsa, Semi-quantitative Culture	H&E	r=0.27, p=0.09	r=0.22, p=0.17	-	-
Statistical testing: Pearson's correlation coefficient **Conclusion:** No correlation between *Hp* density and neutrophil infiltration or mononuclear cell infiltrate								
Kumar et al. (2001)[57]	India	60 patients - 100% (192/192)	Warthin-Starry	H&E	-	p=0.02	-	p=0.002
Statistical testing: Kruskal-Wallis ANOVA **Conclusion:** Positive correlation between *Hp* density and mononuclear cell infiltrate								
Misra et al. (2000)[78]	India	50 patients - 98.0% (49/50)	Loeffler's methylene blue stain	H&E	r_s=0.7096, p<0.0001	"	"	"
Statistical testing: Spearman's correlation coefficient **Conclusion:** Positive correlation between *Hp* density and both neutrophil infiltration and mononuclear cell infiltrate								

Authors (Year)	Country	Sample Size & Prevalence	Test Method		Association of *Hp* density and Inflammation			
					Antrum		Body/Corpus	
			Hp Density	Gastritis Severity	PNC	MNC	PNC	MNC
Fareed et al. (2000)[79]	Pakistan	150 patients - 82.7% (124/150)	Giemsa	H&E	r_s=0.542, p=0.00	r_s=0.173, p=0.035	r_s=0.644, p=0.00	r_s=0.245, p=0.003
Statistical testing: Spearman's correlation coefficient **Conclusion:** Positive correlation between *Hp* density and active gastritis severity								
Chen et al. (2000)[80]	Japan	162 patients - 83.3% (135/162)	UBT	H&E	p<0.05	p=0.176	p=0.262	p=0.537
Statistical testing: Multivariate logistic regression **Conclusion:** Positive correlation between *Hp* density and neutrophil infiltration in antrum but not in the corpus or mononuclear cell infiltrate in either antrum or corpus								
Yamaoka et al. (1999)[81]	USA & Columbia	41 patients - 100% (41/41)	Giemsa & Semi-quantitative Culture	H&E	r_s=0.83	r_s=0.80	Rs<0.2	Rs<0.2
Statistical testing: Spearman's correlation coefficient **Conclusion:** Positive correlation between *Hp* density and neutrophil infiltration and mononuclear cell infiltrate in antrum but not in the corpus								

Authors (Year)	Country	Sample Size & Prevalence	Test Method		Association of *Hp* density and Inflammation			
					Antrum		Body/Corpus	
			Hp Density	Gastritis Severity	PNC	MNC	PNC	MNC
Yamamura et al. (1999)[82]	Japan	81 patients - 65.4% (53/81)	Giemsa	H&E	r_s=0.6050, p<0.00001)	r_s=0.5754, p<0.00001	"	"
Statistical testing: Spearman's correlation coefficient **Conclusion:** Positive correlation between *Hp* density and both neutrophil infiltration and mononuclear cell infiltrate								
Atherton et al. (1996)[83]	USA	29 patients - 100% (29/29)	Giemsa & Semi-quantitative Culture	H&E	p=0.006 (Giemsa), p=0.009 (Culture)	p=0.02 (Giemsa), p=0.007) (Culture)	p=0.11 (Giemsa)	p=0.23 (Giemsa)
Statistical testing: Student's t test **Conclusion:** Positive correlation between *Hp* density and both neutrophil infiltration and mononuclear cell infiltrate in antrum but not in the corpus								
Khulusi et al. (1995)[84]	England	81 patients	Semi-quantitative Culture	H&E	r_s=0.53, p<0.001	r_s=0.51, p<0.001	r_s=0.50, p<0.001	r_s=0.53, p<0.001
Statistical testing: Spearman's correlation coefficient **Conclusion:** Positive correlation between *Hp* density and both neutrophil infiltration and mononuclear cell infiltrate								
Alam et al. (1992)[85]	USA	903 patients	Warthin-Starry	H&E	p<0.0001	p<0.0001	-	-
Statistical testing: Cochran-Mantel-Haenszel correlation **Conclusion:** Positive correlation between *Hp* density and both neutrophil infiltration and mononuclear cell infiltrate								

Authors (Year)	Country	Sample Size & Prevalence	Test Method		Association of *Hp* density and Inflammation			
			Hp Density	Gastritis Severity	Antrum		Body/Corpus	
					PNC	MNC	PNC	MNC
Logan et al. (1991)[86]	England	50 patients - 68% (34/50)	UBT	H&E	p<0.0001	p<0.07	-	-

Statistical testing: Not specified

Conclusion: Positive correlation between *Hp* density and both neutrophil infiltration and mononuclear cell infiltrate

PNC, polymorphonuclear cell; MNC, mononuclear cell; IM, intestinal metaplasia; H&E, hematoxylin and eosin; r_s, Spearman's correlation coefficient; r, Pearson's correlation coefficient
"gastric biopsy samples from antrum and body/corpus pooled

Chapter 3: Comparing the accuracy of staining methods for the detection of *Helicobacter pylori* in histopathologically evaluated gastric biopsies

3.1 Introduction

The accurate diagnosis of *Hp* infection is essential to ensure that people with this infection receive appropriate treatment and follow-up care and that people without this infection do not receive unnecessary treatment. There are multiple methods for detecting *Hp* infection, but there is no optimal gold standard *Hp* detection method. All CAN*Help* community project participants were invited to undergo screening for *Hp* infection using the ^{13}C UBT and upper gastrointestinal endoscopy. A smaller number of participants consented to endoscopy, during which endoscopists collected gastric biopsies for histopathological evaluation and tissue culture. CAN*Help* community project participants were tested for *Hp* infection by up to 3 detection methods: UBT; histopathological evaluation of gastric tissue slides prepared with H&E and Giemsa stains; and tissue culture.

Various sources of error may impact the results of methods used to classify *Hp* infection status. Fundamental among them is that all *Hp* detection methods may fail to detect low bacterial loads; because of this, gold standard classifications of mild infections are especially prone to false-negative classifications. This is particularly the case for biopsy-based *Hp* detection methods, including histology and tissue culture, because gastric tissue samples collected by endoscopy can miss *Hp* in the stomach due to its patchy distribution. Inadequate fixation or physical agitation during tissue biopsy transport can cause the mucosa's delicate components, including gastric mucus, to separate; if the bacteria are in the mucus rather than the submucosal tissue, they could be missed on histological examination. Poor staining of slides can cause either

false-negative or false-positive results because too little stain may fail to highlight bacteria adequately, and inconsistent staining may create visual artifacts. False-negative culture results can arise from the fastidious culture requirements for *Hp* that make the recovery of viable bacteria difficult. False-positive culture results can occur from contamination across samples in the laboratory.

While the non-invasive UBT is not vulnerable to missing *Hp* due to its patchy colonization, false-positive UBT results can occur because of gastric colonization by non-*pylori* urease-positive bacteria, which can arise under hypochlorhydric conditions in the stomach[87]. False-positive UBT results can also occur from inter-individual variation in CO_2 production, particularly in young children[88]. False-negative UBT results can occur in individuals taking medications that suppress *Hp* activity, such as antisecretory drugs, bismuth, or antibiotics; they can also occur in individuals with upper gastrointestinal bleeding and those with reduced lung capacity who cannot produce a breath sample with adequate CO_2 concentration[4].

The multiple *Hp* detection methods used in CAN*Help* projects created the need for a protocol for classifying discrepant results. This analysis aims to investigate whether the use of Giemsa, Warthin-Starry and/or IHC ancillary staining improves the accuracy of histopathological classification of *Hp* presence relative to H&E staining alone. Given the significant cost of ancillary stains, an assessment of the accuracy of these stains relative to alternatives will contribute information needed to evaluate their cost-effectiveness for histological assessment of *Hp* infection. The results of this validation study will provide information on *Hp* infection status classification accuracy in CAN*Help* community projects. To further assess the accuracy of ancillary stains, this study will also investigate the correlation of *Hp* density graded using diverse histopathological stains with ^{13}C-UBT values.

3.2 Methods

Study Design

I used a validation study design to estimate agreement and diagnostic accuracy of histopathology, tissue culture, and the ^{13}C-UBT for classifying *Hp* infection status, focusing on the relative accuracy of distinct histological stains. This validation study used cross-sectional data pertaining to samples collected from 2008 through 2013 in four CAN*Help* community projects focused on the disease burden from *Hp* infection in Arctic Indigenous Canadian communities: the Aklavik, Old Crow, Tuktoyaktuk and Fort McPherson *Hp* Projects. This sample collection served as the sampling frame for a histopathology validation sub-study to assess the accuracy of diverse histopathology stains for detecting *Hp* in gastric tissue slides. This analysis does not include data from CAN*Help* projects launched after 2013.

Research Setting

Recruitment for the Aklavik *Hp* Project started in November 2007[89]. The community of Aklavik, NT, had a population of 594 people in 2006, according to local census data[90]. Initial endoscopy activities took place in Aklavik in February 2008; endoscopy follow-up took place in Inuvik in March 2017. Recruitment for the Old Crow *Hp* Project started in December 2010. The community of Old Crow, YT, had a population of 245 people in 2011, according to local census data[91]. Initial endoscopy activities took place in Old Crow in January 2012; endoscopy follow-up took place in Inuvik in June 2017. Recruitment for the Tuktoyaktuk *Hp* Project started in February 2011[92]. The community of Tuktoyaktuk, NT, had a population of 854 people in 2011, according to local census data[93]. Initial endoscopy activities for Tuktoyaktuk participants took place in Inuvik in March 2013; a few Tuktoyaktuk residents participated in endoscopy in 2013,

and none had endoscopy follow-up. Recruitment for the Fort McPherson *Hp* Project started in June 2012[94]. The community of Fort McPherson, NT, had 792 people in 2011, according to local census data[93]. Initial endoscopy activities took place in Fort McPherson in March 2013; endoscopy follow-up took place in Inuvik in June 2017.

Endoscopy Methods

Upper gastrointestinal endoscopy for consenting participants was performed by study endoscopists in temporary endoscopy clinics organized in community health centers, except for Tuktoyaktuk residents, whose endoscopies took place at the Inuvik Regional Hospital. Information on endoscopic abnormalities, such as gastric and duodenal inflammation, erosions, and ulcers was collected. Gastric biopsies were also collected during endoscopy: 4-6 biopsies (2 from the antrum, 2 from the corpus, and 1-2 from the incisura) were collected for histopathological evaluation and 2-4 biopsies (1-2 from the antrum, 1-2 from the corpus) were collected for tissue culture. The initial endoscopy activities in Aklavik in February 2008 used Olympus 4.9 mm ultrathin trans-nasal gastroscopes. The remaining endoscopy activities in community projects used Olympus 5.4 mm ultra-slim gastroscopes. All gastric biopsies collected during endoscopy activities were transported to the University of Alberta for histopathological evaluation and tissue culture.

Histopathological Evaluation

Initially, a single pathologist (Pathologist 1, SG) assessed pairs of tissue slides, including one stained with H&E and one stained with Giemsa, examined side-by-side, to grade *Hp* density, inflammation severity, and gastric neoplasia using the updated Sydney system[5].

Histopathological variables relevant to this analysis include *Hp* density grade, active gastritis grade and chronic gastritis grade. For the validation sub-study, new tissue cuts were taken from the original formalin-fixed paraffin-embedded gastric tissue blocks (including the 4-6 biopsies collected from each participant at endoscopy) and slides were prepared with Warthin-Starry and IHC ancillary stains at the Red Deer Regional Hospital Centre. Warthin-Starry staining was performed on a Dako Artisan platform using a Dako Artisan Warthin Starry staining kit. IHC was performed on a Dako Omnis platform using a Dako *Hp* antibody clone and a Dako detection kit (EnVision FLEX, High pH - DAB Chromogen). Due to Pathologist 1's reduced availability when this validation study was conducted, a second pathologist (Pathologist 2, GC) used the updated Sydney system to grade *Hp* density in 4 sets of slides, each with a distinct stain (H&E, Giemsa, Warthin-Starry, or IHC) applied and examined individually. In this analysis, I used Pathologist 2's density grading of each of the 4 sets of slides, and Pathologist 1's *Hp* density, chronic gastritis, and active gastritis gradings from the original CAN*Help* assessment of side-by-side H&E-stained and Giemsa-stained paired slides. Given the assessment of multiple biopsies per participant, each participant's classification of *Hp* density grade, chronic gastritis grade and active gastritis grade was based on the highest grade assigned across the biopsies evaluated from that participant.

For the validation sub-study, I took steps to conceal the identity of participants from the study pathologist to reduce observer bias in the assessment of slides from the same participant. To do this, I masked the participant identity on slides by covering the participant ID label, replacing it with a label containing a unique, randomized study ID number, and assigning a different series of ID numbers to each set of stained slides (H&E, Giemsa, Warthin-Starry and IHC), so slides from the same participant could not be identified. I kept the code list that linked

the series of validation study ID numbers to the participant ID number and delivered slides in consecutive batches to the pathologist for histopathological grading. Pathologists 1 and 2 both assessed the slides using the updated Sydney System. Pathologist 2 was blinded to all previous histopathology results for the assessed samples. With each batch, Pathologist 2 received a worksheet to record the grading values; this worksheet included only the validation sub-study ID numbers unique to that batch.

Tissue Culture

Collected biopsies for tissue culture were pooled and mechanically homogenized. The homogenized suspensions were plated on both selective (with antibiotic) and non-selective (with no antibiotic) brain heart infusion/yeast extract/5% horse serum (BHI/YE/HS) agar plates. Two selective BHI/YE/HS plates were used for each homogenized suspension; one plate with 2 antibiotics (vancomycin and amphotericin B) and one plate with 6 antibiotics (vancomycin, amphotericin B, cefsulodin, polymyxin B, trimethoprim, and β-cyclodextrin). Inoculated plates were incubated at 37°C in microaerophilic conditions. Cultures were assessed for growth every 48 hours for up to a month. *Hp* identification was confirmed by urease, catalase, oxidase and 16S ribosomal RNA PCR testing. *Hp* was sub-cultured to generate glycerol stocks.

UBT Screening

The CAN*Help* Working Group ^{13}C-UBT protocol adapted manufacturer instructions for collecting breath samples with modifications based on a review article by Gisbert et al. (2004)[4]. CAN*Help* project staff collected a baseline breath sample from participants. Participants were then given a 50mg or 75mg ^{13}C-labelled urea solution (over time available test kits used a

reduced dose) dissolved in 100 mL of water. Another breath sample (T_{30}) was collected 30 minutes after drinking the ^{13}C-labelled urea solution. The collected breath samples were measured using an IRIS® breath test analyzer (Kibion/Wagner Analysen Technik, Bremen, Germany). The test result is based on the delta-over-baseline (DOB) value, calculated by subtracting the baseline $^{13}C/^{12}C$ ratio from the T_{30} $^{13}C/^{12}C$ ratio, with the $^{13}C/^{12}C$ ratio expressed as parts per thousand (‰) relative to the international Pee Dee Belemnite (PDB) standard[4,44]. For the purpose of *Hp* infection classification, DOB values from -1.99 to 2.49 were classified as *Hp* negative; 2.50 to 3.99 were classified as borderline; and ≥4.00 were classified as *Hp* positive. Participants were asked to repeat the UBT if the result was borderline, the DOB was ≤-2.00, or the CO_2 concentration was too low in baseline or T_{30} breath samples for accurate measurement.

Participant Selection

Participants eligible to be selected for the validation sub-study included all participants in CAN*Help* community projects with gastric tissue available from biopsies previously collected by endoscopy, who also had tissue culture and urea breath test (UBT) results as of 2013. Participants with *Hp* treatment between UBT screening and endoscopy were not eligible for inclusion. Project staff offered participation in endoscopy activities to all residents of participating communities aged 15 years or older at the time of endoscopy, regardless of *Hp* infection status; children younger than 15 years were included at a parents' request. Of 919 participants in the 4 community projects launched before 2013, 324 had gastric biopsies taken during endoscopy. Of the 324 participants with gastric biopsies, there were 309 eligible participants, who did not have *Hp* treatment between UBT screening and endoscopy, who also had culture and UBT results. Due to funding limitations, the validation sub-study used stratified

sampling to select 180 of the 309 eligible participants while ensuring an optimal distribution of *Hp* density in the validation sub-study population. Because *Hp* is more easily detected when density is higher, the validity and usefulness of ancillary stains are better assessed in cases that test discrimination between *Hp*-negative tissue samples and tissue samples with low bacterial density or otherwise patchy bacterial distribution[9,10]. The stratified sampling aimed at achieving an adequate number of *Hp*-negative tissue samples and samples with low *Hp* density, based on Pathologist 1's original assessment of paired Giemsa and H&E slides. Selection for this validation sub-study included:

- All 53 participants (29%) who tested negative by histopathology, culture and UBT
- All 41 participants (23%) with discrepant results on the 3 tests
- 86 participants (48%) who tested positive on all 3 tests, distributed by *Hp* density (as graded by Pathologist 1) as follows:
    - All 8 participants (4%) with 1 or more biopsies classified as "rare" *Hp* density, Pathologist 1's label for very few bacteria observed
    - All 31 participants (17%) with "mild" as the highest *Hp* density grade in assessed biopsies
    - 47 participants (26%) selected at random from those with mild as the lowest *Hp* density in assessed biopsies

Demographic Information

Demographic information was collected through structured interviews by trained CAN*Help* Working Group interviewers using multiple questionnaires around the time of ^{13}C-

UBT screening in community projects. Additionally, standardized chart reviews extracted data from participant health records. To characterize the study population so that generalizability can be assessed, I used basic demographic variables (age, sex, ethnicity) collected through the CAN*Help* Participant Questionnaire and Health Survey. I used the community of residence at the time of project enrolment, as indicated by the participant ID, to account for differences in biopsy collection methods used by different clinical teams at different times.

To exclude participants who had *Hp* treatment provided through a CAN*Help* community project between UBT testing and endoscopy, which occurred with a small fraction of participants, I consulted CAN*Help* study information. To account for antibiotic use, I used self-reported information on previous *Hp* treatment collected in CAN*Help* community project participants by interviewer-administered questionnaire. I used information on proton pump inhibitor (PPI) and non-steroidal anti-inflammatory drug (NSAID) use collected by interviewer-administered questionnaire to assess whether they influence detection method accuracy.

Validation Measures and Statistical Methods

To assess inter-rater reliability, I compared the 2 pathologist's Giemsa classifications of *Hp* presence/absence and density. I also qualitatively assessed the trend in the relationship between the two pathologists' Giemsa-based classifications of *Hp* density, an ordinal measure, and I calculated percent agreement of the ordinal classifications. I did not use Cohen's kappa for observer comparisons because it has known performance limitations for assessing ordinal data, and its values lack a straightforward interpretation[24]. Instead, I used the two-way random effects intraclass correlation coefficient (ICC) to measure absolute agreement and consistency of agreement. Absolute agreement assesses whether different raters assigned the same score to the

same subject, while consistency assesses if raters' scores are correlated in an additive manner[95]. I used a two-way random effects model rather than a one-way random effects model or a two-way mixed effects model because it is preferred for generalizing reliability results to any raters with the same characteristics as the raters compared in a reliability assessment[95]. A two-way random effects model ICC is considered appropriate for evaluating rater-based clinical assessment methods (e.g., histopathological evaluation) designed for routine clinical use by clinicians with specific characteristics (e.g., specialization in pathology)[95]. ICC values <0.5, from 0.5 to <0.75, from 0.75 to <0.9 and ≥0.90 are generally interpreted as poor, moderate, good, and excellent reliability between raters, respectively[95].

Since there is no optimal gold standard method for detecting *Hp* infection, I used diverse available reference classifications in lieu of a single gold standard to estimate multiple measures of accuracy for the classification of tissue samples as *Hp*-negative or *Hp*-positive by each of the staining methods evaluated for the validation sub-study (H&E, Giemsa, Warthin-Starry, and IHC). I treated the following reference classifications successively as 'gold standards' to estimate accuracy measures: the original CAN*Help* histopathology evaluation using paired H&E and Giemsa-stained slides examined side-by-side by Pathologist 1; tissue culture; UBT; and the algorithm developed by the CAN*Help* Working Group for classifying *Hp* positivity or negativity (based on the probability that the participant was negative or positive given the combined results).

As one measure of accuracy, I estimated percent agreement with each 'gold standard' for each histopathology stain evaluated for the validation sub-study (H&E, Giemsa, Warthin-Starry, and IHC) as the percent of participants whose classification is the same on the pair of methods compared. I also estimated sensitivity, specificity, predictive values, and likelihood ratios for the

classification resulting from each staining method compared against each 'gold standard'. I used the 95% confidence interval (CI) for all estimated proportions as a measure of statistical precision. The use of likelihood ratios in validation studies is limited[96], but I used them as a marker of the potential utility of each diagnostic test. Likelihood ratios for diagnostic tests compare the probability that a test result is correct to the probability that it is incorrect; they are calculated for both positive ('LR+' = sensitivity / 1- specificity) and negative test results ('LR-' = 1 – sensitivity/specificity)[97].

I examined whether the following histopathological factors impact accuracy: *Hp* density; chronic gastritis severity; and active gastritis severity. I dichotomized *Hp* density, active gastritis, and chronic gastritis as none-mild or moderate-marked. Using stratified analysis to examine the influence of these factors on accuracy, I estimated accuracy measures within categories of histopathological factors, previous *Hp* treatment, NSAID use, and PPI use to the extent possible for each stain evaluated for the validation sub-study, given that the precision of accuracy estimates was reduced in these categories.

I used linear regression to model the relationship between *Hp* density, as classified by Pathologist 1's original histopathological evaluation, and ^{13}C-UBT DOB values, as a measure of bacterial load. I also calculated Spearman's correlation coefficient for *Hp* density and ^{13}C-UBT delta-over-baseline values, separately for Pathologist 1's density classifications and each of the ancillary stains assessed by Pathologist 2. For the linear regression and Spearman's correlation coefficient analysis, I excluded *Hp*-negative participants to avoid mixing the assessment of whether urease activity increases with bacterial density with the assessment of whether UBT and histology agree on the presence of *Hp*[56].

Lastly, I conducted a bias analysis using the estimated accuracy parameters to estimate the amount of bias that would result from each staining method when estimating the prevalence of *Hp* infection in participants. I calculated the adjusted prevalence (P_{adj}) using specificity and sensitivity estimates and compared them to the observed prevalence (P_{obs}). Per Rogan et al. (1978)[98], if sensitivity and specificity are known, an unbiased estimate of P_{adj} is calculated using the following equation: $P_{adj} = (P_{Obs} + \text{specificity} - 1) / (\text{specificity} + \text{sensitivity} - 1)$[98,99]. I estimated 95% confidence intervals as recommended by Rogan et al. (1978)[98]. For this bias analysis, I used sensitivity and specificity estimated using the 'gold standard' based on the CAN*Help* Working Group algorithm due to its applicability to published research by the CAN*Help* Working Group.

I conducted all statistical analyses using STATA Data Analysis and Statistical Software (Version 14, StataCorp LLC, Texas, USA).

Ethics Approval

The research conducted by the CAN*Help* Working Group, of which this study is a part, received ethics approval from the University of Alberta Research Ethics Board under the project name "Addressing Community Concerns about Risks from *H. pylori* Infection in the Circumpolar North" (No. Pro00007868). This research was also approved by the Northwest Territories (Licence No. 15785) and Yukon (Licence No. 16-13S&E) research licensing authorities. CAN*Help* Working Group staff members provided verbal and written study information, approved by the University of Alberta Research Ethics Board, to all participants. All participants, or their guardians if they were under 17 years old at the time of enrollment, signed consent forms to indicate they understood the project information before enrolling as

project participants. Younger children whose parents deemed them old enough to understand gave informed assent. Through the Northern Alberta Clinical Trials and Research Centre (NACTRC), Alberta Health Services provided administrative and operational approvals for processing and pathological assessment of gastric tissue samples collected in community health centres and the Inuvik Regional Hospital.

3.3 Results

Of the 919 participants enrolled in the four community projects, 309 participated in endoscopy and had eligible histopathology results. From these 309 participants, 180 were selected for the histopathology validation sub-study. Table 4 describes the demographic characteristics of these 3 study populations. The Aklavik *Hp* Project had both the largest number of total participants (n=383) and participants with histopathology results (n=194). The Old Crow and Fort McPherson *Hp* Projects had comparable numbers of total participants and participants with histopathology results. Only 5 Tuktoyaktuk residents had eligible histopathology results because of logistic barriers to participation in endoscopy.

The distribution of demographic characteristics in these three study populations was similar except for age. A large proportion of total CAN*Help* project participants (20%) were <20 years of age at enrollment. The proportion of participants who were <20 years of age at enrollment was lower for those with histopathology results (10%) and those selected for the validation sub-study (12%). A larger proportion of participants were female than male among total participants (54%), participants with histopathology results (55%) and those selected for the validation sub-study (60%). Similarly, most participants were Indigenous among total

participants (84%), participants with histopathology results (88%) and those selected for the validation sub-study (87%).

Table 5 shows the distribution of grades of *Hp* density, chronic gastritis, and active gastritis in the study population of participants with histopathology results and the validation sub-study population. Among participants with histopathology results, 27% were *Hp*-negative, 17% had *Hp* density classified as mild, 32% as moderate, and 24% as marked. Among validation sub-study participants, 41% were *Hp*-negative, 21% had *Hp* density classified as mild, 26% as moderate, and 13% as marked. The higher frequency of *Hp* negativity and the lower frequency of moderate or marked *Hp* density in the validation sub-study population reflects the stratified sampling aimed at including as many samples as possible with no *Hp* or low *Hp* density. A similar difference was apparent for chronic gastritis, and active gastritis; the proportion of participants with lower grades of gastritis was larger among validation sub-study participants compared to all participants with histopathology results. Table 5 also compares frequencies of receiving *Hp* treatment before project enrollment and use of PPIs or NSAIDs. In the validation sub-study population, 23% had previous *Hp* treatment, 19% reported PPI use, and 57% reported NSAID use; frequencies of these variables were similar among the 309 participants with histopathology results.

A comparison of the 2 pathologists' classifications of participants' *Hp* status as positive or negative showed absolute agreement for 95% (171/180) of participants. The remaining 9 (5%) participants were classified as *Hp* positive by Pathologist 1 and *Hp* negative by Pathologist 2. A comparison of the 2 pathologists' classifications of the *Hp* density grade showed absolute agreement on 63% (114/180) of participants. Of the 66 (37%) participants with discordant *Hp* density grades, 74% (49/66) are explained by Pathologist 2 grading one grade lower than

Pathologist 1 (e.g., none instead of mild, mild instead of moderate, or moderate instead of marked). A two-way random effects model, measuring the absolute agreement, estimated an ICC of 0.66 (95% CI 0.35, 0.80). Meanwhile, a two-way random effects model, measuring the consistency of agreement, estimated an ICC of 0.73 (95% CI 0.66, 0.80).

Table 6 shows the distribution of *Hp* density classified using each ancillary stain (Warthin-Starry, IHC, Giemsa, H&E) examined one-by-one by Pathologist 2 and the original CAN*Help* histopathology grade (based on side-by-side examination of H&E and Giemsa-stained slides by Pathologist 1), among 180 (100%) validation sub-study participants. The distribution of *Hp* density in Pathologist 2's validation sub-study assessments had proportionally more participants classified at lower grades across all stains relative to the original classifications by Pathologist 1. For example, the proportion of participants with *Hp* density classified as marked was 13% according to the original classification, 9% when assessed by IHC, 3% by Warthin-Starry, 1% by Giemsa, and 0 by H&E. Pathologist 2 noted some quality issues during histopathological evaluation: poor quality of staining or faded staining for 18 participants in at least one of two H&E slides evaluated per participant; precipitate on one H&E slide for 1 participant; focal non-specific staining for 14 (8%) participants in at least one of two IHC slides evaluated per participant; all Giemsa-stained slides for all participants were faded; no quality issues were noted for Warthin-Starry stained slides. Pathologist 2 indicated the need for special staining when evaluating H&E slides for 35 (19%) participants.

Table 7 shows concordance of detection methods on classification of *Hp* positivity or negativity; 65% (117/180) of participants' classifications were concordant across all staining methods evaluated for the validation study, as well as the original histopathology classification, culture and ^{13}C UBT. The most frequent discordant results pertained to participants whose

classifications were positive only on culture, representing 9% (16/180) of the validation sub-study population, while 5% (9/180) of participants were *Hp*-negative on culture but *Hp*-positive on all other tests. Eleven participants (6%) were negative on H&E staining but positive on all other tests. Two participants (1%) were negative on Giemsa staining but positive on all other tests. Conversely, a few participants were positive only on one ancillary stain (3 (2%) on IHC, 3 (2%) on H&E, and 2 (1%) on Warthin-Starry). Two participants (1%) were positive on both IHC staining and culture only. Only 2 participants (1%) were positive by all endoscopy-based methods and negative on UBT. The remaining variation of discordance did not follow any discernable pattern.

Table 8 shows the estimated sensitivity, specificity, and predictive values of ancillary stains evaluated for the validation sub-study against different 'gold standards.' All evaluated stains performed similarly, regardless of the 'gold standard', except for H&E staining. H&E staining had the lowest sensitivity and agreement of the four stains evaluated. Warthin-Starry and IHC staining showed very similar sensitivity and agreement. Warthin-Starry staining had higher specificity relative to IHC against all 'gold standards' except for tissue culture, for which the estimated specificity (77%, 95% CI 66, 86) of these 2 stains was identical. Relative to IHC and Warthin-Starry staining, Giemsa staining had lower sensitivity but higher specificity across all 'gold standards'. All stains had lower sensitivity, specificity, and agreement against tissue culture as the 'gold standard'. Against the CAN*Help* classification algorithm as the 'gold standard', a positive Warthin-Starry result was nearly 24 times more likely among participants with *Hp* infection than among patients without *Hp* infection, but only 3.5 times more likely when tissue culture was the 'gold standard'.

Table 9 shows the agreement of *Hp* positivity classifications based on each of the 4 stains with diverse 'gold standards' across categories of chronic gastritis severity, acute gastritis severity, *Hp* density, previous *Hp* treatment, PPI use and NSAID use. For all stains, agreement was more frequent when *Hp* density was moderate-marked than when it was none-mild. The difference in agreement was largest when tissue culture was the 'gold standard', which also showed higher agreement within higher grades of chronic gastritis and active gastritis than within lower grades. When tissue culture was the 'gold standard', estimated agreement was higher for all four staining methods among participants with no previous *Hp* treatment. Overall, the precision of accuracy estimates within categories of stratification variables was high.

Among all CAN*Help* participants, ^{13}C-UBT DOB values ranged from -10.74 to 103.85 (median 9.08); this range was -10.39 to 72.33 (median 12.61) among participants with histopathology results and -10.39 to 60.01 (median 6.80) among validation sub-study participants (Figure 1). (Theoretically, UBT DOB values should not be negative except for random variation and for this reason, participants with values below -2.00 were asked to repeat the UBT, but not all instances of lower negative values were resolved.) A general trend of increasing UBT DOB values with increasing *Hp* density grades was apparent across stains evaluated for the validation sub-study (Figure 2), except for Giemsa staining, for which participants with marked *Hp* density had lower UBT DOB values (median 14.02) on average than those with moderate *Hp* density (median 23.70).

After excluding participants with *Hp* density classifications of none (n=102, 33%), a weak positive correlation between *Hp* density and UBT DOB values was apparent. Among the 207 (66%) participants with histopathology results and *Hp* density classifications of mild or higher, the Spearman correlation coefficient was 0.34 (p=0.00) for the correlation of *Hp* density

(based on the original classification by Pathologist 1) with UBT DOB values. When limited to the validation sub-study population, excluding those with *Hp* density classifications of none, Spearman correlation coefficients were 0.23 (p=0.02) for the original CAN*Help* histopathology classification (n=106), 0.15 (p=0.16) for H&E staining (n=89), 0.17 (p=0.09) for Giemsa staining (n=97), 0.21 (p=0.03) for IHC staining (n=107), and 0.17 (p=0.09) for Warthin-Starry staining (n=104).

A linear regression model with the UBT DOB value as the outcome variable and *Hp* density as classified by Pathologist 1 and modelled as a continuous variable, with data from all participants with histopathology results excluding those with *Hp* density classifications of none, suggested that UBT DOB values increase in a linear manner with density grade (F-statistic p-value = 0.00); *Hp* density accounted for 10.2% of the explained variability in the UBT DOB, and the regression coefficient was 6.7 (95% CI 4.0, 9.3), indicating an average increase of 6.7 UBT DOB units as the density grade increases from mild to moderate and from moderate to marked. A linear regression model including *Hp* density as a nominal ordinal variable (F-statistic p-value = 0.00) showed similar results: *Hp* density accounted for 9.8% of the explained variability in the UBT DOB; regression coefficients were 6.4 (95% CI 1.3, 11.51) for moderate *Hp* density compared to mild, indicating an average increase of roughly 6 UBT DOB units with an increase from mild to moderate density, and 13.3 (95% CI 7.9, 18.67) for marked *Hp* density compared to mild, indicating an average increase of roughly 13 UBT DOB units for an increase from mild to marked density.

I estimated corrected prevalence based on the classifications arising from each stain evaluated for the validation sub-study, with sensitivity and specificity estimated using the CAN*Help* classification algorithm as the ‘gold standard’. Starting with the prevalence observed

in the 180 (100%) validation sub-study participants for *Hp* positivity classifications yielded by each histopathology stain, I adjusted these prevalence estimates for the proportion of false negatives and false positives corresponding to the estimated sensitivity and specificity. According to the 'gold standard' CAN*Help* classification, *Hp* prevalence in the validation sub-study population was 58.9%. For H&E staining, the observed prevalence was 50.0%, and the corrected prevalence was 58.9% (95% CI 47.3, 70.5), indicating that the prevalence estimated from H&E-stained tissue slides was 85% of the 'gold standard' prevalence (or 8.9 percentage points lower). For Giemsa staining, the observed prevalence was 54.4%, and the corrected prevalence was 58.9% (95% CI 50.4, 67.4), indicating that the prevalence estimated from Giemsa-stained tissue slides was 92% of the 'gold standard' prevalence (or 4.5 percentage points lower). For IHC staining, the observed prevalence was 60.0%, and the corrected prevalence was 58.9% (95% CI 49.9, 67.9), indicating that prevalence estimated from IHC-stained tissue slides was roughly 1 percentage point higher than the 'gold standard' prevalence. For Warthin-Starry staining alone, observed prevalence was 58.3%, and corrected prevalence was 58.9% (95% CI 49.9, 67.9), indicating that prevalence estimated from Warthin-Starry-stained tissue slides was within 1 percentage point of the 'gold standard' prevalence. This exercise demonstrates how sensitivity and specificity can be used to correct prevalence estimated using a particular classification method so that the prevalence estimate matches that of the gold standard.

3.4 Discussion

This study showed a high degree of concordance on Giemsa, Warthin-Starry and IHC staining for the detection of *Hp* infection in gastric tissue slides, with similar estimates of sensitivity, specificity, and agreement against diverse 'gold standards.' Given that the evaluation

of histopathology stains was based on slides cut from the same tissue block, the bacterial distribution should be similar, providing a fair comparison of the value of each stain for enhancing bacterial detection. In contrast, the Warthin-Starry and IHC stains were applied more recently than the H&E and Giemsa stains; thus, fading of stains over time may have contributed to an increased number of false negative classifications of slides stained by H&E and Giemsa. The observed low specificity of staining results when tissue culture is used as a 'gold standard' is consistent with the relatively low sensitivity of tissue culture. The use of multiple gold standards helped guard against the misleading interpretation that random error alone impacts estimated accuracy.

Higher accuracy of histopathological classification in cases with higher grades of *Hp* density makes sense because the bacteria are more numerous and easier for a pathologist to find. This is especially the case when comparing staining results to tissue culture because more bacteria in a gastric biopsy makes a positive culture result more likely. Evidence suggests that the severity of chronic and active gastritis increases with increased *Hp* density[57,67,73,75,100]. Since the bacteria and the inflammation co-occur in the tissue, the pathologist evaluating *Hp* density may be influenced by background pathology, including any gastritis present, and may explain why accuracy is greater in cases with more severe gastritis.

This analysis showed a weak positive correlation between *Hp* density and UBT DOB values. The Spearman correlation coefficients observed in this study are lower than other reports of the correlation between *Hp* density and UBT DOB values[48,50,51,54,55,61–66]. Restricting this analysis to the validation sub-study participants weakened the correlation, which would be expected because the stratified sampling of participants for the validation sub-study aimed to maximize statistical efficiency for evaluating tissue samples with sparse *Hp* colonization, given

their higher probability of misclassification. Including all participants with histopathology results increased the proportion with higher *Hp* density and strengthened the observed correlation somewhat. Linear regression showed that average UBT DOB values increased with increasing *Hp* density grades of mild, moderate, and marked.

While this validation sub-study was not a study of the reliability of pathologist evaluations, it assessed the reliability of two pathologists' *Hp* density classifications. A qualitative assessment of inter-rater reliability in this sub-study showed higher consistency of agreement than absolute agreement. This observation was supported by ICC models, indicating moderate reliability when measuring absolute agreement, 0.66 (95% CI 0.35, 0.80) and good reliability when measuring the consistency of agreement, 0.73 (95% CI 0.66, 0.80). The correlation of measurements made on the same participant when measuring absolute agreement was lower and less precise with wider confidence intervals than when measuring the consistency of agreement. In addition to reflecting the degree of rater agreement, a low ICC can also reflect a small number of subjects or raters[95].

It should be noted that the Giemsa slides were not evaluated at the same time by both pathologists. Any Giemsa stain fading over the years since Pathologist 1's evaluation could have impacted Pathologist 2's evaluation. Unlike Pathologist 1, Pathologist 2 was blinded to all participant information, including endoscopy reports, and did not evaluate Giemsa slides concurrently with H&E slides. To accurately assess the reliability of different pathologists' evaluations, they should be done simultaneously with the same information. Given that both pathologists' assessments were used for this validation study, however, an assessment of interobserver agreement was warranted.

Correcting *Hp* prevalence estimates for the estimated classification errors of each stain, revealed that H&E staining had the largest difference between observed and corrected prevalence, reflecting relatively low sensitivity resulting in underestimated prevalence. These results confirm that adding an ancillary stain to the histopathological evaluation of gastric biopsies for the presence or absence of *Hp* improves sensitivity and specificity. No one ancillary stain appeared better than others overall in this validation sub-study. From a cost perspective, IHC staining cost nearly twice the cost of Warthin-Starry or Giemsa staining. While a cost-effectiveness analysis is beyond the scope of this validation study, given comparable accuracy and lower cost, either the Warthin-Starry or Giemsa stain would be more cost-effective than IHC relative to H&E alone for improved detection of *Hp* in gastric biopsy tissue.

Table 4. CAN*Help* study population characteristics by data availability

	All CANHelp participants* n=919		**Participants with histopathology data n=309**		**Participants in validation sub-study n=180**	
Variables	n	%*	n	%*	n	%*
Sex						
Male	422	46	138	45	72	40
Female	497	54	171	55	108	60
Age at enrollment						
0-14	117	13	7	2	5	3
15-24	133	14	46	15	28	16
25-34	137	15	46	15	23	13
35-44	113	12	42	14	27	15
45-54	168	18	75	24	41	23
55-64	129	14	54	17	30	17
65-96	122	13	39	13	26	14
Ethnicity						
Indigenous	768	84	272	88	156	87
Inuvialuit (Inuit)	305	33	121	39	79	44
Gwich'in First Nation	393	43	132	43	67	37
Other/mixed**	70	8	19	6	10	6
Non-Indigenous	82	9	22	7	19	11
Missing	69	8	15	5	5	3
Community of Residence						
Aklavik	383	42	194	63	131	73
Old Crow	200	22	60	19	19	11
Fort McPherson	229	25	50	16	26	14
Tuktoyaktuk	107	12	5	2	4	2

*Of projects in Aklavik NT, Old Crow YT, Tuktoyaktuk NT, Fort McPherson NT
**% of column total; totals may not sum to 100% due to rounding
**Includes: Métis; mixed Indigenous ethnicities; unspecified Indigenous ethnicity; Indigenous mixed with non-Indigenous ethnicities

Table 5. Distribution of the highest grade of chronic gastritis, active gastritis, and *Hp* density among gastric biopsies graded by CAN*Help* study pathologist, by data availability

	Participants with histopathology data n=309		**Participants in validation sub-study n=180**	
Variables	n	%*	n	%*
Chronic Gastritis				
None	75	24	67	37
Mild	27	9	17	9
Moderate	100	32	60	33
Marked	107	35	36	20
Missing	*0*	*0*	*0*	*0*
Active Gastritis				
None	89	29	75	42
Rare	1	0	1	1
Mild	103	33	66	37
Moderate	81	26	34	19
Marked	33	11	4	2
*missing***	*2*	*1*	*0*	*0*
Hp Density				
None	84	27	73	41
Mild	51	17	38	21
Moderate	99	32	46	26
Marked	75	24	23	13
*missing***	*2*	*1*	*0*	*0*
Previous *Hp* treatment				
Yes	61	20	41	23
No	243	79	139	77
Missing	*5*	*2*	*0*	*0*

	Participants with histopathology data n=309		Participants in validation sub-study n=180	
Variables	n	%*	n	%*
PPI Use				
Yes	58	19	34	19
No	221	72	146	81
Missing	*30*	*10*	*0*	*0*
NSAID Use				
Yes	165	53	102	57
No	144	47	78	43
Missing	*0*	*0*	*0*	*0*

Hp, Helicobacter pylori; PPI, proton pump inhibitor; NSAID, non-steroidal anti-inflammatory drug
*% of column total; totals may not sum to 100% due to rounding
**2 participants lack classification for active gastritis and *Hp* density

Table 6. Distribution of *Hp* density classified histopathologically, by ancillary stain (Warthin-Starry, IHC, Giemsa, H&E examined one-by-one by Pathologist 2) and the original CAN*Help* histopathology grade (based on side-by-side examination of H&E and Giemsa-stained slides by Pathologist 1), among 180 validation sub-study participants

	H&E		**Giemsa**		**IHC**		**Warthin-Starry**		**Original CAN*Help* Histopathology**	
	n	%	n	%	n	%	n	%	n	%
Hp Density										
None	90	50	82	46	72	40	75	42	73	41
Mild	80	44	79	44	64	36	68	38	38	21
Moderate	10	6	17	9	28	16	31	17	46	26
Marked	0	0	2	1	16	9	6	3	23	13

Hp, Helicobacter pylori; IHC, immunohistochemical; H&E, hematoxylin and eosin
*% of column total; totals may not sum to 100% due to rounding

Table 7. Concordance of *Hp* infection status (present/absent across classification methods validation sub-study histopathology stain (Warthin-Starry, IHC, Giemsa, H&E), original CAN*Help* histopathology grade, urea breath test (UBT) result, and culture result, among 180 validation sub-study participants (*dark grey box=positive test, white box=negative test)

Number of Participants (n)	**Warthin-Starry**	**IHC**	**Giemsa**	**H&E**	**Original CAN*Help* Histopathology**	**UBT**	**Culture**
72	■	■	■	■	■	■	■
45							
16							■
11	■	■	■		■	■	■
9	■	■	■	■	■	■	
3		■					
3				■			
2	■						
2	■	■	■	■	■		■
2		■					■
2	■	■		■	■	■	■

Number of Participants (n)	Warthin-Starry	IHC	Giemsa	H&E	Original CAN*Help* Histopathology	UBT	Culture
1	■	■		■	■		
1					■		
1	■	■	■		■	■	
1	■			■	■	■	
1	■	■			■	■	
1		■			■	■	
1	■		■		■	■	
1						■	
1	■	■	■		■		■
1					■		■
1	■	■			■	■	■
1		■	■		■	■	■
1						■	■

IHC, immunohistochemical; H&E, hematoxylin and eosin; UBT, urea breath test

Table 8. Sensitivity, specificity, and predictive values of ancillary histopathology stains by 'gold standard' and stain (Warthin-Starry, IHC, Giemsa, H&E) among 180 validation sub-study participants

	Sensitivity		**Specificity**		**PPV**		**NPV**		**Accuracy**		**LR+**	**LR-**
	%*	95%CI†	%*	95%CI†	%*	95%CI†	%*	95%CI†	%*	95%CI†		
Gold Standard: Original Histopathology												
H&E	81	73, 88	96	88, 99	97	91, 99	78	68, 86	87	81, 92	19.78	0.19
Giemsa	92	85, 96	100	95, 100	100	96, 100	89	80, 95	95	91, 98	-	0.08
IHC	96	91, 99	93	85, 98	95	90, 98	94	86, 98	95	91, 98	14.05	0.04
Warthin-Starry	96	91, 99	97	90, 100	98	93, 100	95	87, 99	97	93, 99	35.12	0.04
Gold Standard: UBT												
H&E	82	73, 86	92	84, 97	93	86, 98	79	69, 87	86	80, 91	10.47	0.20
Giemsa	92	85, 96	96	89, 99	97	91, 99	90	82, 96	94	89, 97	23.65	0.08
IHC	96	90, 99	88	79, 95	92	85, 96	94	86, 98	93	88, 96	8.22	0.04
Warthin-Starry	96	90, 99	92	84, 97	94	87, 98	95	87, 99	94	90, 97	12.33	0.04
Gold Standard: Culture												
H&E	69	60, 78	80	69, 89	84	75, 91	62	51, 72	73	66, 80	3.45	0.39
Giemsa	79	70, 86	84	74, 92	89	81, 94	72	61, 81	81	75, 87	5.03	0.25
IHC	84	75, 90	77	66, 86	85	77, 91	75	63, 84	81	75, 87	3.66	0.21
Warthin-Starry	81	72, 88	77	66, 86	85	76, 91	72	60, 82	79	73, 85	3.53	0.25

	Sensitivity		Specificity		PPV		NPV		Accuracy		LR+	LR-
	%*	95%CI†	%*	95%CI†	%*	95%CI†	%*	95%CI†	%*	95%CI†		
Gold Standard: CANHelp Algorithm												
H&E	80	72, 87	95	87, 99	96	89, 99	77	67, 85	86	80, 91	14.67	0.21
Giemsa	92	85, 96	100	95, 100	100	96, 100	89	80, 95	95	91, 98	-	0.08
IHC	95	89, 98	92	83, 97	94	88, 98	93	85, 98	94	89, 97	11.60	0.05
Warthin-Starry	95	89, 98	96	88, 99	97	92, 99	93	85, 98	96	91, 98	23.19	0.05

IHC, immunohistochemical; H&E, hematoxylin and eosin; UBT, urea breath test; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; LR+, positive likelihood ratio; LR-, negative likelihood ratio
*% of column total; totals may not sum to 100% due to rounding
†binomial Wald CI for numerators ≥10 and <100, binomial exact CI for numerators <10, one-sided binomial exact 97.5% CI for numerators=100)

Table 9. Agreement of ancillary histopathology stains (Warthin-Starry, IHC, Giemsa, H&E) with diverse 'gold standards', by gastritis severity, previous *Hp* treatment, PPI use, and NSAID use, among 180 validation sub-study participants.

		H&E		Giemsa		IHC		Warthin-Starry	
		%*	95%CI†	%*	95%CI†	%*	95%CI†	%*	95%CI†
Gold Standard: Original Histopathology									
	All	87	81, 92	95	91, 98	95	91, 98	97	93, 99
Chronic gastritis	None-Mild	90	82, 96	87	78, 93	94	87, 98	96	90, 99
	Moderate-Marked	84	76, 91	92	84, 96	97	91, 99	97	91, 99
Active Gastritis	None-Mild	86	79, 91	94	89, 98	94	88, 97	96	92, 99
	Moderate-Marked	92	78, 98	97	86, 100	100	91, 100	97	92, 99
Hp Density	None-Mild	84	76, 90	95	89, 98	93	86, 97	95	90, 99
	Moderate-Marked	93	84, 98	96	88, 99	99	92, 100	99	92, 100
Previous *Hp* treatment	No	88	81, 93	94	88, 97	95	90, 98	96	91, 98
	Yes	85	71, 94	100	91, 100	95	83, 99	100	91, 100
PPI Use	No	88	81, 93	95	90, 98	97	92, 99	96	91, 98
	Yes	85	69, 95	94	80, 99	88	72, 97	100	91, 100
NSAID Use	No	88	79, 95	97	91, 100	95	87, 99	97	91, 99
	Yes	86	78, 92	93	86, 97	95	89, 98	96	90, 99
Gold Standard: UBT									
	All	86	80, 91	94	89, 97	93	88, 96	94	90, 97
Chronic gastritis	None-Mild	89	80, 95	98	92, 99	92	84, 97	95	88, 99
	Moderate-Marked	83	74, 90	91	83, 96	94	87, 98	94	87, 98
Active Gastritis	None-Mild	85	78, 91	94	88, 97	91	86, 96	94	89, 98
	Moderate-Marked	89	75, 97	95	82, 99	97	86, 100	95	82, 99
Hp Density	None-Mild	84	76, 90	95	89, 98	91	84, 96	94	87, 97
	Moderate-Marked	90	80, 96	93	84, 98	96	88, 99	96	88, 99

		H&E		Giemsa		IHC		Warthin-Starry	
		%*	95%CI†	%*	95%CI†	%*	95%CI†	%*	95%CI†
Previous *Hp* treatment	No	86	79, 92	92	86, 96	92	86, 96	93	87, 96
	Yes	85	70, 94	100	91, 100	95	83, 99	100	91, 100
PPI Use	No	86	79, 91	95	89, 98	94	89, 98	94	89, 97
	Yes	88	73, 97	91	76, 98	85	69, 95	97	85, 100
NSAID Use	No	87	77, 94	96	89, 99	94	86, 98	96	89, 99
	Yes	85	77, 92	92	85, 97	92	85, 97	93	86, 97
Gold Standard: Culture									
	All	73	66, 80	81	75, 87	81	75, 87	79	73, 85
Chronic gastritis	None-Mild	68	57, 78	73	63, 83	73	62, 82	71	61, 81
	Moderate-Marked	78	68, 86	88	79, 93	89	80, 94	86	78, 93
Active Gastritis	None-Mild	69	60, 76	77	69, 84	77	69, 83	75	67, 82
	Moderate-Marked	89	75, 97	95	82, 99	97	86, 100	95	82, 99
Hp Density	None-Mild	66	56, 75	78	70, 86	77	68, 84	76	67, 83
	Moderate-Marked	86	75, 93	86	77, 93	88	78, 95	86	75, 93
Previous *Hp* treatment	No	77	69, 84	83	75, 89	81	74, 87	81	73, 87
	Yes	61	45, 76	76	60, 88	80	65, 91	76	60, 88
PPI Use	No	73	66, 81	80	73, 86	82	74, 87	79	72, 86
	Yes	71	53, 85	85	69, 95	79	62, 91	79	62, 91
NSAID Use	No	73	62, 82	79	69, 88	82	72, 90	79	69, 88
	Yes	74	64, 82	82	74, 89	80	71, 88	79	70, 87
Gold Standard: CANHelp Algorithm									
	All	86	80, 91	95	91, 98	94	89, 97	96	91, 98
Chronic gastritis	None-Mild	89	81, 95	98	92, 100	92	84, 97	95	88, 99
	Moderate-Marked	83	74, 90	93	86, 97	96	90, 99	96	90, 99

		H&E		Giemsa		IHC		Warthin-Starry	
		%*	95%CI†	%*	95%CI†	%*	95%CI†	%*	95%CI†
Active Gastritis	None-Mild	84	77, 90	94	89, 98	92	86, 96	95	90, 98
	Moderate-Marked	92	79, 98	97	86, 100	100	91, 100	97	86, 99
Hp Density	None-Mild	82	74, 89	95	89, 98	91	84, 96	94	87, 97
	Moderate-Marked	93	84, 98	96	88, 99	99	92, 100	99	92, 100
Previous *Hp* treatment	No	86	79, 92	94	88, 97	94	89, 97	94	89, 97
	Yes	85	71, 94	100	91, 100	95	83, 99	100	91, 100
PPI Use	No	86	80, 91	95	90, 98	95	90, 98	95	89, 98
	Yes	85	69, 95	94	80, 99	88	73, 97	100	90, 100
NSAID Use	No	90	81, 95	99	93, 100	96	89, 99	99	93, 100
	Yes	83	75, 90	92	85, 97	92	85, 96	93	86, 97

IHC, immunohistochemical; H&E, hematoxylin and eosin; UBT, urea breath test; CI, confidence interval; *Hp, Helicobacter pylori*; PPI, proton pump inhibitor; NSAID, non-steroidal anti-inflammatory drug
*% of column total; totals may not sum to 100% due to rounding
†binomial Wald CI for numerators ≥10 and <100, binomial exact CI for numerators <10, one-sided binomial exact 97.5% CI for numerators=100

Figure 4. Box plot comparing ^{13}C-urea breath test (UBT) delta over baseline (DOB) (δ ‰ in $^{13}C/^{12}C$ relative to standard reference value) across all CAN*Help* community project participants screened by urea breath test (n=833) (All), participants with histopathology and UBT results (n=276) (Histopathology) and participants selected for the validation sub-study with UBT results (n=178) (Validation)

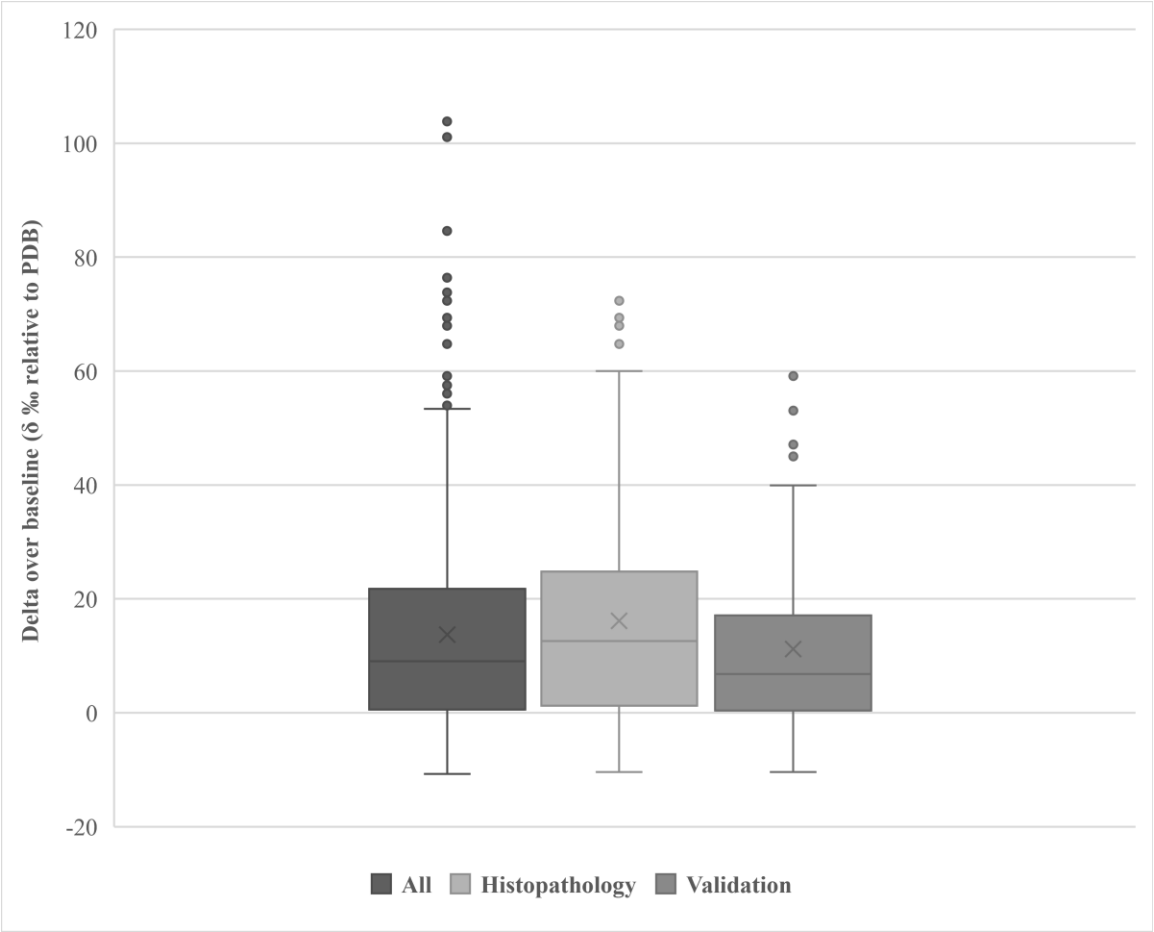
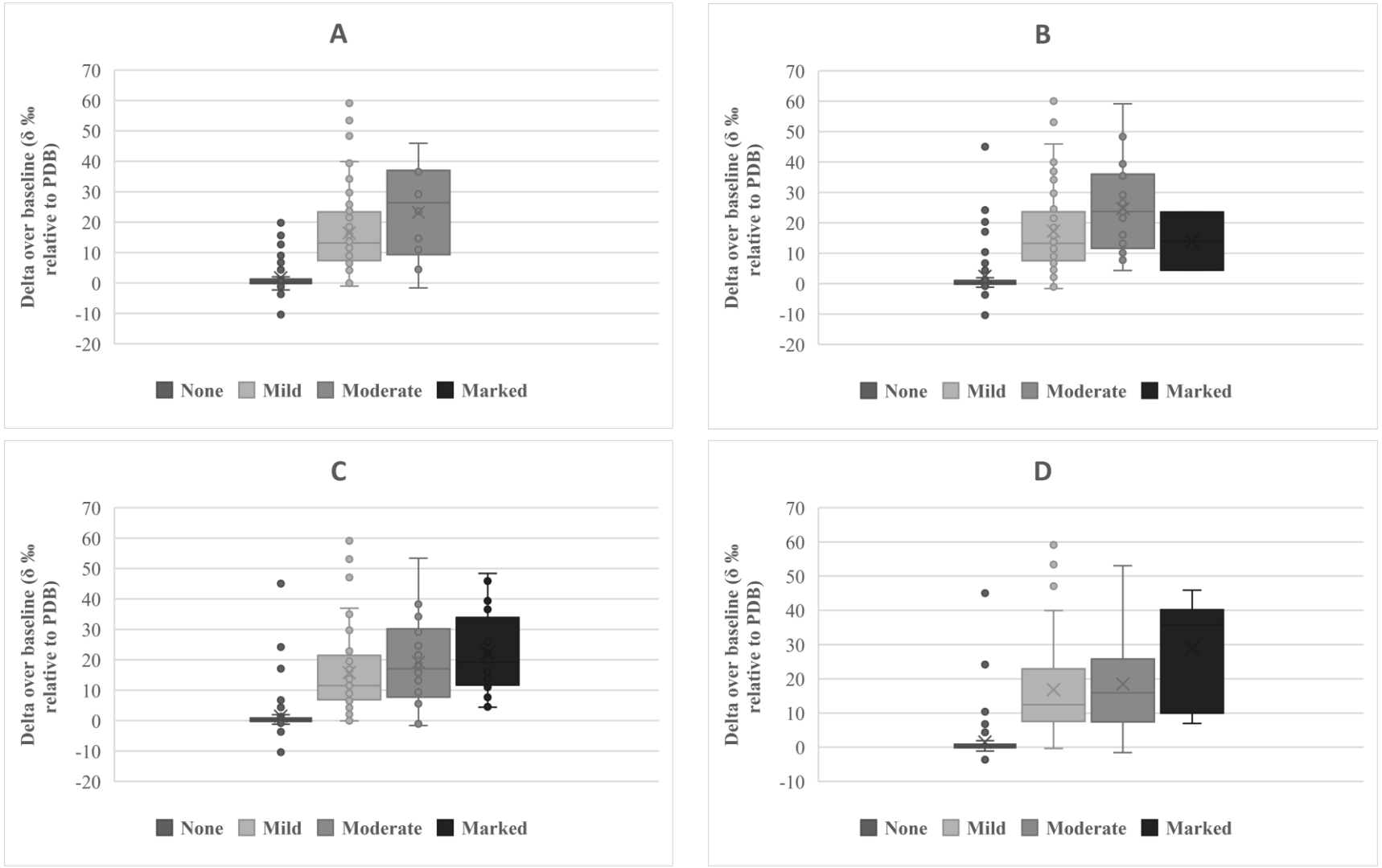


Figure 5. Box plots comparing ^{13}C-UBT delta over baseline (DOB) (δ ‰ in $^{13}C/^{12}C$ relative to standard reference value) by *Hp* density assessed by H&E (A), Giemsa (B), IHC (C), and Warthin-Starry (D) stains among participants selected for the validation sub-study with UBT results (n=178)



Chapter 4: Investigating the effect of *Helicobacter pylori* density on the severity of gastritis in Arctic Indigenous communities

4.1 Introduction

Hp infection causes chronic inflammation of the gastric mucosa in nearly all people with a persistent infection. Chronic gastritis can progress to atrophic gastritis involving the loss of gastric cells and glands, believed to be the start of a sequence of progressively advancing precursors for gastric carcinoma, including intestinal metaplasia and dysplasia[101]. While there are many causes of inflammation, atrophy, and metaplasia of the gastric mucosa, these pathological changes are all strongly associated with *Hp* infection[102]; in particular, there is an increased risk of gastric cancer when one or more of these pathologies are present[102,103].

Pathology reports based on the updated Sydney system characterize the tissue in terms of the stomach subsite and cell types as well as the presence of *Hp*, chronic inflammation, active inflammation, glandular atrophy, intestinal metaplasia, dysplasia, and carcinoma[5]. In cases where *Hp* or gastritis is present, an ordinal evaluation of *Hp* density and gastritis severity (none, mild, moderate, marked) is reported[5]. Multiple investigators have reported an association between *Hp* density and gastritis severity in gastric biopsies[57,67,73–75,83,84,100,104,105]. Some investigators, however, have reported a null association between *Hp* density and gastritis severity[68,106]. In other studies, both severe gastritis and *Hp* density have been identified as factors influencing treatment success[107–109].

This analysis aims to estimate the effect of *Hp* density on the severity of gastritis among *Hp*-positive residents of Arctic Canadian communities who participated in projects conducted by the CAN*Help* Working Group. This analysis will further the CAN*Help* Working Group's aim of

identifying factors that influence the local burden of disease from *Hp* infection. Characterizing the relationship between the *Hp* bacterial load and gastritis severity in Arctic Indigenous communities will have relevance to other populations with a similar prevalence of *Hp* infection and associated disease.

4.2 Methods

Study Design

I used a cross-sectional design to estimate the effect of *Hp* density on active gastritis and chronic gastritis among participants in community projects focused on the disease burden from *Hp* infection in 7 Arctic Indigenous communities: the Aklavik, Old Crow, Tuktoyaktuk, Fort McPherson, Ross River, Teslin, and Inuvik *Hp* Projects. My analysis is limited to the subset of participants with gastric biopsies assessed by histopathology.

Research Setting

Among Northwest Territories (NT) communities, recruitment for the Aklavik *Hp* Project started in November 2007[89]. The community of Aklavik, NT, had a population of 594 people in 2006, according to local census data[90]. Initial endoscopy activities occurred in Aklavik in February 2008; endoscopy follow-up occurred in Inuvik in March 2017. Recruitment for the Tuktoyaktuk *Hp* Project started in February 2011[92]. The community of Tuktoyaktuk, NT, had a population of 854 people in 2011, according to local census data[93]. Initial endoscopy activities for Tuktoyaktuk participants took place in Inuvik in March 2013; a few Tuktoyaktuk residents participated in endoscopy in 2013, and none had endoscopy follow-up. Recruitment for the Fort McPherson *Hp* Project started in June 2012[94]. The community of Fort McPherson, NT, had 792

people in 2011, according to local census data[93]. Initial endoscopy activities occurred in Fort McPherson in March 2013; endoscopy follow-up occurred in Inuvik in June 2017. Recruitment for the Inuvik *Hp* Project started in 2016. The community of Inuvik, NT, had a population of 3,243 people in 2016, according to local census data26. Initial endoscopy activities took place in Inuvik in June 2017.

Among Yukon (YT) communities, recruitment for the Old Crow *Hp* Project started in December 2010. The community of Old Crow, YT, had a population of 245 people in 2011, according to local census data[91]. Initial endoscopy activities occurred in Old Crow in January 2012; endoscopy follow-up occurred in Inuvik in June 2017. Recruitment for the Ross River *Hp* Project started in 2016. The community of Ross River, YT, had a population of 293 people in 2016, according to local census data[111]. Initial endoscopy activities took place in Ross River in June 2017. Recruitment for the Teslin *Hp* Project started in November 2016. The community of Teslin, YT, had a population of 263 people in 2016, according to local census data[111]. Initial endoscopy activities took place in Teslin in June 2017.

Endoscopy Methods

Upper gastrointestinal endoscopy for consenting participants took place in temporary endoscopy clinics organized in community health centers. Information on endoscopic abnormalities (in particular, gastric and duodenal inflammation, erosions, and ulcers) was collected. Gastric biopsies were also collected during endoscopy. In total, 4-6 biopsies (2 from the antrum, 2 from the corpus, and 1-2 from the incisura) were collected for histopathological evaluation, while 2-4 biopsies (1-2 from the antrum, 1-2 from the corpus) were collected for tissue culture. The initial endoscopy activities in Aklavik in February 2008 used Olympus 4.9

mm ultrathin trans-nasal gastroscopes. The remaining endoscopy activities in community projects used Olympus 5.4 mm ultra-slim gastroscopes. All gastric biopsies collected during endoscopy activities were transported to the University of Alberta for histopathology evaluation and tissue culture.

Histopathology Evaluation

The Sydney System defines active gastritis as polymorphonuclear cell (e.g., neutrophils) infiltration of the gastric mucosa and chronic gastritis as increased mononuclear cells (e.g., lymphocytes, plasma cells, monocytes, mast cells, and/or eosinophils) in the lamina propria[5]. A single pathologist assessed H&E and Giemsa-stained slides to grade *Hp* density and inflammation severity (active and chronic gastritis) using the updated Sydney system[5] for all community projects included in this analysis. I classified the main study outcomes, active gastritis and chronic gastritis, and the main study exposure, *Hp* density, using this histopathological evaluation of gastric biopsies collected for community projects. Given the assessment of multiple biopsies per participant, each participant's classification of *Hp* density grade, chronic gastritis grade and active gastritis grade was based on the highest grade assigned across the biopsies from that participant. I re-classified *Hp* density as mild for 6 participants the study pathologist classified as having rare *Hp* density.

Participant Selection

Participants eligible to be selected for this analysis are all participants in CAN*Help* community projects with gastric biopsies assessed by histopathology. Only results from the first endoscopy event were used in this analysis for participants who participated in more than one.

Participation in endoscopy activities was offered to all residents of participating communities aged 15 years or older at the time of endoscopy, regardless of *Hp* infection status. Children under 15 years of age were included at parents' request. Of 1,422 participants in the community projects, 407 had gastric biopsies taken during a first endoscopy event and assessed for *Hp* density and inflammation severity. Two *Hp*-positive participants lacked classification for active gastritis and *Hp* density. Excluding these 2 participants left 405 participants with complete histopathology results.

Demographic Information

I used information collected through structured interviews conducted by trained CAN*Help* Working Group staff members. CAN*Help* Working Group staff used interview tools to collect information from participants at the time of ^{13}C-UBT screening and endoscopy. They also conducted standardized chart reviews to extract data from participant health records. To characterize the study population to assess generalizability, I used basic demographic variables: age at project enrollment, sex, ethnicity, and education level. Common exposures other than *Hp* known to influence gastritis include non-steroidal anti-inflammatory drug (NSAID) use, alcohol intake, and smoking[112,113]. To ascertain other relevant study variables, I used self-reported information on alcohol consumption, NSAID use, and previous *Hp* treatment collected by interviewer-administered questionnaire. Information on smoking was collected inconsistently across community projects and therefore was not available. In my analysis, I assessed whether age, sex, ethnicity, alcohol consumption, NSAID use, previous *Hp* treatment or community of residence confound the effect of *Hp* density on gastritis severity. Because it is not likely that any of these variables are affected by *Hp* density, they would not be intermediate variables in the

causal pathway, and their inclusion in regression models would not introduce bias from controlling for mediating effects. Because previous *Hp* treatment may substantially impact both *Hp* density and gastritis, I examined whether it modified the relationship.

Data collected by questionnaire may be subject to misclassification due to imperfect recall or underreporting behaviour perceived as socially undesirable. For example, participants may not accurately report alcohol consumption, smoking, NSAID use or previous *Hp* treatment. Using interviewers for questionnaire administration trained to use a neutral tone and develop a rapport with participants answering questions helped avoid underreporting or overreporting exposures based on social desirability. Using trained interviewers who could explain the questionnaire to participants and ensure they understood the questions helped ensure accuracy of responses. Having staff members check completed questionnaires for missing data at the time of questionnaire administration and carrying out double entry of responses into a database and comparing databases for discrepancies helped to avoid missing data and data entry errors.

Statistical Methods

I dichotomized gastritis severity based on the previously published severity distribution in CAN*Help* community projects[6]. Among CAN*Help* participants with *Hp* detected by histology, 3% had no active gastritis, 46% had mild active gastritis, 36% had moderate active gastritis, 14% had marked active gastritis, 1% had no chronic gastritis, 8% had mild chronic gastritis, 44% had moderate chronic gastritis, and 47% had marked chronic gastritis[6]. Because of this, I dichotomized active gastritis as none-mild or moderate-marked and chronic gastritis as none-mild-moderate or marked.

I examined the association of *Hp* density with gastritis severity, excluding *Hp*-negative participants to avoid mixing the assessment of whether gastritis severity increases with increasing bacterial density with whether the presence of *Hp* is associated with gastritis; this exclusion also avoided collinearity in regression models because few *Hp*-negative participants had any degree of gastritis and very few *Hp*-positive participants had no gastritis. Though correlation coefficients are not epidemiologic measures of effect, they are commonly reported in the literature on the relationship between *Hp* density and gastritis severity. For comparisons with published reports, I used Spearman's correlation coefficient for ordinal measures to estimate the correlation between *Hp* density and active gastritis severity and between *Hp* density and chronic gastritis severity.

Using logistic regression models for dichotomized gastritis severity outcomes, I estimated prevalence odds ratios (ORs) and 95% CIs as the measure of the effect of *Hp* density on the prevalence of the more severe gastritis category, adjusting for potential confounders. I used ORs instead of prevalence ratios for reasons argued by Pearce (2004)[114]: ORs estimate the incidence rate ratio when the average duration of the disease is the same in the exposed and nonexposed groups (i.e., exposure has no effect on duration) without the rare disease assumption, which is not true for the prevalence ratio;[114] ORs also provide practical, analytical, and theoretical consistency between analyses of a prevalence study and any future case–control analyses based on the same study population[114].

I took multiple steps to build multivariable logistic regression models to estimate the effect of *Hp* density on active gastritis severity and, in separate models, on chronic gastritis severity. First, I estimated crude ORs for the associations of *Hp* density with the gastritis outcome variables. Then, I used purposeful selection of covariates to adjust for confounding[115],

starting with a multivariable model including the set of variables with crude OR p-values ≤0.25. I dropped each variable from the model one at a time. If the coefficients of the exposure of interest (*Hp* density), or their respective CIs, changed by ≥10% upon removal of a variable, then the removed variable was retained as a confounder in the multivariable model. Finally, I used the likelihood ratio (LR) test in applicable instances to determine whether alternate forms of variables improved the model fit.

I also used purposeful selection for alternate categorizations of categorical covariates. In particular, I collapsed smaller categories and used the LR test to assess whether one categorization improved the model fit relative to another categorization. Once I selected the optimal variable form, I subjected it to the change-in-estimates assessment. Because the study population included multiple members of households, I assessed the need to include a random-effects parameter for household membership in case the gastritis outcomes cluster by household. In addition to assessing different categorizations of continuous variables in regression models, I used Lowess plots to visualize the relationship of the variable with gastritis outcomes and then assessed whether using cubic splines improved the fit of the variable in the model. To assess the potential impact of missing data, I compared the unadjusted odds ratio for the association of each study variable with each gastritis outcome in the subset of all participants with data on that variable and the subset of participants with complete data on all variables retained in regression models.

I conducted all statistical analyses using STATA Data Analysis and Statistical Software (Version 14, StataCorp LLC, Texas, USA).

Ethics Approval

The research conducted by the CAN*Help* Working Group, of which this study is a part, received ethics approval from the University of Alberta Research Ethics Board under the project name "Addressing Community Concerns about Risks from *H. pylori* Infection in the Circumpolar North" (No. Pro00007868). The Northwest Territories (Licence No. 15785) and Yukon (Licence No. 16-13S&E) research licensing authorities also approved this research. CAN*Help* Working Group staff members provided verbal and written study information, approved by the University of Alberta Research Ethics Board, to all participants. All participants, or their guardians if they were under 17 years old at enrollment, signed consent forms to indicate they understood the project information before enrolling as project participants. Younger children whose parents deemed them old enough to understand gave informed assent. Through the Northern Alberta Clinical Trials and Research Centre (NACTRC), Alberta Health Services provided administrative and operational approvals for processing and pathological assessment of gastric tissue samples collected in community health centres and the Inuvik Regional Hospital.

4.3 Results

Of the 1,302 participants enrolled in the 7 community projects, 407 participated in endoscopy, and 405 had complete histopathology results. Table 10 describes the demographic characteristics of the study populations by data availability. The Aklavik *Hp* Project had the largest number of total participants (n=392, 30%) as well as participants with histopathology results (n=208, 51%). The Old Crow and Fort McPherson *Hp* Projects had comparable numbers of total participants and participants with histopathology results. The remaining community projects had comparable numbers of total participants and participants with histopathology

results, except for the Tuktoyaktuk *Hp* Project. Only 13 (3%) Tuktoyaktuk residents had histopathology results because of logistic barriers to participation in endoscopy.

The distribution of demographic characteristics in the study population with complete histopathology results was similar to that of the total CAN*Help* study population except for age. Among all 1,302 CAN*Help* project participants, 156 (13%) were 0-14 years of age at project enrollment. Among participants with histopathology results, only 8 participants (2%) were 0-14 years of age at project enrollment. A larger proportion of participants were female than male among total participants (56%) and participants with histopathology results (58%). Similarly, most participants were Indigenous among total participants (79%) and participants with histopathology results (86%). A relatively large proportion of participants had not completed high school among total participants (39%) and participants with histopathology results (43%). Table 10 also compares the frequencies of receiving *Hp* treatment before project enrollment and the use of NSAIDs. Of all CAN*Help* participants, 23% had previous *Hp* treatment, and 57% reported NSAID use; frequencies of these variables were similar among the participants with histopathology results.

Table 11 shows the prevalence of active and chronic gastritis grades within *Hp* density grades among the 405 participants with complete histopathology results: 133 (33%) had no active gastritis; 130 (32%) had active gastritis severity classified as mild, 106 (26%) as moderate, and 36 (9%) as marked. Among the 285 *Hp*-positive participants, 14 (5%) had no active gastritis, 129 (45%) had active gastritis severity classified as mild, 106 (37%) as moderate, and 36 (13%) as marked. Among the 405 participants with histopathology results, 104 (26%) had no chronic gastritis, 47 (12%) had chronic gastritis severity classified as mild, 134 (33%) as moderate, and 120 (30%) as marked. Among the 285 *Hp*-positive participants, none (0%) had no chronic

gastritis, 32 (11%) had active gastritis severity classified as mild, 133 (47%) as moderate, and 120 (42%) as marked. Across increasing density grades, the frequency of moderate and marked severity increases for both active and chronic gastritis.

After excluding participants with *Hp* density classifications of none, positive correlations between *Hp* density and severity of each gastritis outcome were apparent. Among the 285 participants with histopathology results and *Hp* density classifications of mild or higher, the Spearman correlation coefficient was 0.55 (p=0.00) for the correlation of *Hp* density with active gastritis severity and 0.66 (p=0.00) for the correlation of *Hp* density with chronic gastritis severity. In this same subset of *Hp*-positive participants, *Hp* density coefficient p-values, interpretable as p-values for tests of trend, from logistic regression models with dichotomized gastritis outcome variables and *Hp* density modelled as a continuous variable, were 0.00 for both active gastritis severity and chronic gastritis severity.

LR test results indicated that the inclusion of a random effects term for household did not improve the fit of logistic regression models for either active gastritis severity or chronic gastritis severity, suggesting that these outcomes did not cluster in households; therefore, clustering in households as a random effect was not included in the final multivariate model.

The unadjusted ORs for the association of each study variable with each gastritis outcome in Tables 12 and 13 were relatively comparable for the subset of the study population with data on the variable and the subset with complete data on variables included in multivariable models. This indicated that missing data would not greatly impact the estimated effects for either gastritis outcome.

After carefully assessing category boundaries, I recategorized ethnicity, education, and community to avoid categories that were too small for regression models. I recategorized

ethnicity as Inuvialuit (Inuit), Gwich'in First Nations, and other (Métis, mixed Indigenous ethnicities, unspecified Indigenous ethnicity, Indigenous mixed with non-Indigenous ethnicities, non-Indigenous ethnicities). I recoded education into categories of less than high school, high school or equivalent (including less than high school with trades or other certification), and college or university. I recategorized the community of residence variable into Aklavik, Old Crow, other NT communities and other YT communities. After recategorizing, I assessed each newly categorized variable for inclusion in the multivariable model. The recategorized education and ethnicity variables did not produce a change in the coefficients of the exposure of interest (*Hp* density) in either multivariate model and, therefore, were not subsequently included. The recategorized community variable produced a ≥10% change in the coefficients in the multivariable model for active gastritis only and was included in that model.

I assessed multiple categorical and continuous options for including age. I assessed age as a continuous variable using Lowess plots to visualize its relationship with each gastritis outcome. Results of the LR test indicated that age modelled as a cubic spline was the best choice of the options assessed. I included age in the multivariable models as a cubic spline, with knots set at 21, 45 and 67, but it did not produce a ≥10% change in coefficients for either gastritis outcome.

The purposeful selection algorithm did not select variables to adjust for confounding except for community of residence in the multivariable model for active gastritis. Because of this, I decided to present results from multivariable models that included each variable that appeared to be clearly associated with the respective gastritis outcome based on an unadjusted OR p-value <0.20. For the model with chronic gastritis as the outcome, I adjusted for age as a cubic spline, sex, ethnicity, and community. For the model with active gastritis as the outcome, I adjusted for age as a cubic spline, sex, ethnicity, education, NSAID use and community.

However, I subsequently had to drop NSAID use and community from the active gastritis severity model due to statistical instability caused by small categories.

Table 14 presents the results of the multivariable logistic regression models. Among *Hp*-positive participants with histopathology data, *Hp* density had the strongest association with either gastritis outcome among the variables examined. Relative to participants with mild *Hp* density, the odds of moderate-marked relative to none-mild active gastritis were around 5-fold greater in participants with moderate density and 29-fold greater in participants with marked density; the odds of marked relative to none-mild-moderate chronic gastritis were around 6-fold greater in participants with moderate density and 81-fold greater in participants with marked density. Adjusted ORs from the multivariable model also estimated lower prevalence odds of chronic gastritis severity among Aklavik residents relative to residents of other communities; this was also the case for active gastritis severity, but the variable distributions did not permit stable effect estimates. Aside from this, no other variable retained clear associations with chronic gastritis prevalence odds, while ethnicity and education retained moderate associations with active gastritis prevalence odds.

The purposeful selection criteria for variable selection for confounding adjustment resulted in the selection of just 1 adjustment variable (community) for the active gastritis severity model and none for the chronic gastritis severity model. While the analysis assessed several risk factors for gastritis, the only of these variables that would plausibly also be associated with the study exposure, *Hp* density, is previous *Hp* treatment. While previous *Hp* treatment did not meet the purposeful selection criteria for confounder control, I assessed it as a potential modifier, estimating ORs for the effect of *Hp* density on gastritis severity in those with and without previous treatment. For active gastritis, ORs were similar for the moderate-mild *Hp* density

contrast among those with (OR 5.5, 95% CI 0.96, 31) and without (OR 3.9, 95% CI 1.8, 8.1) previous *Hp* treatment and for the marked-mild *Hp* density contrast among those with (OR 22, 95% CI 0.96, 31) and without (OR 19, 95% CI 8.0, 47) previous *Hp* treatment. For chronic gastritis, ORs were lower for both moderate and marked density levels compared to mild in those with previous *Hp* treatment: for the moderate-mild *Hp* density contrast the OR was 4.4 (95% CI 0.46, 42) in those with previous *Hp* treatment and 8.6 (95% CI 2.9, 26) in those without previous *Hp* treatment; marked-mild *Hp* density contrast, the OR was 44 (95% CI 3.4, 573) in those with previous *Hp* treatment and 95 (95% CI 28, 325) in those without previous *Hp* treatment . Estimates for those with previous *Hp* treatment were less precise because of the relatively small number of participants in this category.

4.4 Discussion

This analysis showed a strong increasing trend in gastritis severity with increasing *Hp* density among *Hp*-positive participants in CAN*Help* community projects. Multivariable logistic regression models estimated large increases in the prevalence odds of marked active gastritis and marked chronic gastritis with increasing *Hp* density grades. The Spearman correlation coefficients observed in this study are comparable to other reports of the correlation between *Hp* density and gastritis severity[61,76,78,79,81,82,84].

The low prevalence of mild chronic gastritis in the study population made it impossible to examine increasing levels of chronic gastritis severity using the mild category as the referent. Because the only feasible dichotomization of this outcome was marked vs not marked and the "not marked" category was dominated by moderate cases, the chronic gastritis severity contrast is essentially between marked and moderate chronic gastritis. Conversely, the analysis did not

have sufficient data for a statistically precise distinction between moderate and marked active gastritis.

Despite the gold standard for diagnosing gastritis being the histopathological evaluation of gastric biopsies, the evaluation of both gastritis and *Hp* density are semi-quantitative measurements subject to the discretion of the pathologist grading each. The patchy distribution of *Hp* bacteria in the stomach can adversely impact *Hp* detection by histology. False-negative results may occur in biopsy-based *Hp* detection methods from the localized sampling of biopsies collected through endoscopy. Also, inadequate fixation or physical agitation during specimen transport could cause the mucosa's delicate components, like gastric mucus, to separate; if the bacteria are in the mucus rather than the submucosal tissue, they would be missed on histological examination. Other potential sources of false-negative results in histology samples include poor staining of slides. Insufficient or inconsistently applied staining could lead to either false-negative or false-positive histology results. Results from the validation study reported in Chapter 3 of this thesis temper the concern about a large degree of exposure misclassification in this analysis; in particular, the density grades of 2 pathologists showed good consistency of agreement and the *Hp* density grades were positively associated with UBT DOB values, a distinct measure of bacterial load.

Both the main study outcome, gastritis, and the main study exposure, *Hp* density, were ascertained by the same pathologist at the same time using paired Giemsa and H&E slides. The application of the Sydney System to grade gastritis severity and *Hp* density is subject to variability among pathologists. Using a single experienced pathologist to evaluate all participant results likely enhanced the accuracy of the histopathology results. However, because the bacteria and the inflammation co-occur in the tissue, the pathologist could not be blinded to their grading

of one while grading the other. The pathologist could have been influenced by their assessment of one while grading the other. This means that the misclassification of the exposure cannot be presumed to be independent of the misclassification of the outcome in this analysis. Because *Hp* density is not a dichotomized variable, the direction of bias resulting from the misclassification of the density grade is difficult to predict whether it was non-differential (that is, occurring with equal probability across four levels). Also, misclassification bias may have overestimated the association between *Hp* density and gastritis severity outcomes because it is likely that errors in classifying *Hp* density grades and gastritis grades are not independent.

While previous *Hp* treatment is a suspected confounder based on prior knowledge, this variable may contain substantial misclassification due to imperfect recall resulting in incomplete ascertainment of lifetime history of previous *Hp* treatment. The etiologically relevant time window for an effect of previous treatment is also unclear. Lacking an accurate life history of the timing of previous *Hp* treatment, it was not possible to conduct a sensitivity analysis to determine the relevant time-since-treatment window for classifying this exposure. Given these limitations, adjustment for previous *Hp* treatment may have resulted in residual confounding in this analysis. It should be noted, however, that while this analysis is vulnerable to dependent errors in the classification of exposure and outcome variables, the estimated ORs for the effect of density on gastritis severity are so large that unmeasured confounding is unlikely to be responsible for the observation of strong effects.

Table 10. CAN*Help* study population characteristics by data availability

	All CAN*Help* participants* n=1,302		**Participants with complete histopathology data n=405**	
Variables	n	%**	n	%**
Sex				
Male	568	44	172	42
Female	734	56	233	58
Age				
0-14	156	12	8	2
15-24	168	13	52	13
25-34	200	15	57	14
35-44	189	15	62	15
45-54	237	18	99	24
55-64	197	15	77	19
65-96	153	12	50	12
missing	*2*	*0*	*0*	*0*
Ethnicity				
Indigenous	1,023	79	346	85
Inuvialuit (Inuit)	348	27	143	35
Gwich'in First Nations	425	33	146	36
Tlingit First Nations	51	4	9	2
Kaska Dene First Nations	59	5	17	4
Other/mixed***	140	11	31	8
Non-Indigenous	194	15	41	10
missing	*85*	*7*	*18*	*4*
Education				
< High School	509	39	174	43
High School	160	12	61	15
< HS, plus trade or cert	235	18	70	17
College or uni <Bachelors	108	8	32	8
Bachelors or higher	146	11	44	11
missing	*144*	*11*	*24*	*6*
Previous *Hp* Treatment				
Yes	233	18	83	20
No	1021	78	309	76
missing	*48*	*4*	*13*	*3*

	All CAN*Help* participants* n=1,302		**Participants with complete histopathology data n=405**	
Variables	n	%**	n	%**
NSAID Use				
Yes	721	55	231	57
No	519	40	160	40
missing	*62*	*5*	*14*	*3*
Alcohol Consumption				
Yes	437	34	149	37
No	374	29	135	33
missing	*491*	*38*	*121*	*30*
Community of Residence, Project Launch Year				
Aklavik NT, 2007	392	30	206	51
Old Crow YT, 2010	208	16	63	16
Fort McPherson NT, 2012	237	18	53	13
Tuktoyaktuk NT, 2011	107	8	13	3
Ross River YT, 2016	107	8	24	6
Teslin YT, 2016	124	10	24	6
Inuvik NT, 2016	127	10	22	5

NT, Northwest Territories, YT, Yukon
*Of projects in Aklavik NT, Old Crow YT, Tuktoyaktuk NT, Fort McPherson NT, Ross River YT, Teslin YT and Inuvik NT
**% of column total; totals may not sum to 100% due to rounding
**Includes: Métis; mixed Indigenous ethnicities; unspecified Indigenous ethnicity; Indigenous mixed with non-Indigenous

Table 11. Prevalence of active gastritis grades and chronic gastritis grades within *Hp* density grades among CAN*Help* participants with histopathology results (n=405)

	***Hp* Density**									
	None		Mild		Moderate		Marked		Total	
	n	%*	n	%*	n	%*	n	%*	n	%**
Active Gastritis Severity										
None	119	99	13	17	1	1	0	0	133	33
Mild	1	1	48	64	67	52	14	18	130	32
Moderate	0	0	14	19	55	42	37	46	106	26
Marked	0	0	0	0	7	5	29	36	36	9
Total	120	100	75	100	130	100	80	100	405	100
Chronic Gastritis Severity										
None	104	87	0	0	0	0	0	0	104	26
Mild	15	13	30	40	2	2	0	0	47	12
Moderate	1	1	40	53	82	63	11	14	134	33
Marked	0	0	5	7	46	35	69	86	120	30
Total	120	100	75	100	130	100	80	100	405	100

*% of column total; totals may not sum to 100% due to rounding

Table 12. Unadjusted prevalence odds ratios for the association of selected variables with active gastritis severity among *Hp*-positive CAN*Help* project participants by data availability

	Active Gastritis							
	Participants with histopathology data and *Hp* density >0							
	Total n=285				No missing data*** n=259			
Variables	n	OR	95% CI	p*	n	OR	95% CI	p*
Sex								
Male	126	1.0	-	0.05	115	1.0	-	0.06
Female	159	0.63	0.39, 1.0	-	144	0.62	0.38, 1.0	-
Age								
0-24	46	1.0	-	0.52	43	1.0	-	0.47
25-34	47	0.88	0.39, 2.0	-	42	0.79	0.34, 1.8	-
35-44	45	1.0	0.46, 2.4	-	43	1.0	0.43, 2.3	-
45-54	63	1.0	0.48, 2.2	-	56	0.83	0.37, 1.8	-
55-64	49	1.5	0.64, 3.3	-	45	1.4	0.62, 3.3	-
65-96	35	0.59	0.24, 1.4	-	30	0.55	0.21, 1.4	-
Ethnicity								
Inuvialuit (Inuit)	101	1.0	-	0.03	92	1.0	-	0.06
Gwich'in First Nation	111	2.1	1.2, 3.6	-	108	1.9	1.1, 3.4	-
Other***	59	1.3	0.70, 2.6	-	59	1.2	0.62, 2.3	-
missing	*14*							
Education								
< High School	122	1.0	-	0.24	118	1.0	-	0.32
High School or equivalent	102	1.3	0.74, 2.1	-	98	1.2	0.72, 2.1	-
College or University	43	1.8	0.89, 3.7	-	43	1.7	0.84, 3.5	-
missing	*18*							

	Active Gastritis							
	Participants with histopathology data and *Hp* density >0							
	Total n=285				No missing data*** n=259			
Variables	n	OR	95% CI	p*	n	OR	95% CI	p*
Hp Density								
Mild	75	1.0	-	0.00	71	1.0	-	0.00
Moderate	130	4.0	2.0, 7.8	-	112	3.9	1.9, 7.8	-
Marked	80	21	9.1, 47	-	76	20.0	8.6, 46	-
Previous *Hp* Treatment								
No	232	1.0	-	0.85	218	1.0	-	0.71
Yes	43	0.94	0.79, 1.3	-	41	0.88	0.45, 1.7	-
missing	*10*							
NSAID Use								
No	111	1.0	-	0.00	105	1.0	-	0.00
Yes	162	2.2	1.4, 3.7	-	154	2.3	1.4, 3.8	-
missing	*12*							
Alcohol Consumption								
No	85	1.0	-	0.46	84	1.0	-	0.44
Yes	112	0.81	0.46, 1.4	-	111	0.80	0.45, 1.4	-
missing	*89*							
Community of Residence								
Aklavik	135	1.0	-	0.00	126	1.0	-	0.00
Old Crow	57	21	7.8, 56	-	43	14	5.0, 37	-
Other NT Communities	63	1.6	0.87, 3.0	-	60	1.5	0.79, 2.8	-
Other YT Communities	30	2.6	1.2, 5.9	-	30	2.4	1.0, 5.3	-

*Likelihood ratio chi-square test p-value used for purposeful logistic regression model building

**Includes: Métis; mixed Indigenous ethnicities; unspecified Indigenous ethnicity; Indigenous mixed with non-Indigenous ethnicities; non-Indigenous
***No missing data for sex, ethnicity, education, NSAID use, and community of residence

Table 13. Unadjusted prevalence odds ratios for the association of selected variables with chronic gastritis severity among *Hp*-positive CAN*Help* project participants by data availability

	Chronic Gastritis							
	Participants with histopathology data and *Hp* density >0							
	Total n=285				No missing data*** n=271			
Variables	n	OR	95% CI	p*	n	OR	95% CI	p*
Sex								
Male	126	1.0	-	0.09	120	1.0	-	0.18
Female	159	0.63	0.39, 1.0	-	151	0.72	0.44, 1.2	-
Age								
0-24	46	1.0	-	0.28	45	1.0	-	0.30
25-34	47	0.62	0.27, 1.4	-	44	0.61	0.26, 1.4	-
35-44	45	0.61	0.27, 1.4	-	43	0.63	0.27, 1.5	-
45-54	63	0.45	0.21, 0.98	-	57	0.47	0.21, 1.1	-
55-64	49	0.74	0.33, 1.7	-	48	0.74	0.32, 1.7	-
65-96	35	0.39	0.15, 0.97	-	34	0.36	0.14, 0.94	-
Ethnicity								
Inuvialuit (Inuit)	101	1.0	-	0.00	101	1.0	-	0.00
Gwich'in First Nation	111	1.7	0.96, 2.9	-	111	1.7	0.96, 2.9	-
Other***	59	0.37	0.18, 0.79	-	59	0.37	0.18, 0.79	-
missing	*14*							
Education								
< High School	122	1.0	-	0.93	122	1.0	-	0.93
High School or equivalent	102	0.94	0.55, 1.6	-	102	0.94	0.55, 1.6	-
College or University	43	0.88	0.43, 1.8	-	43	0.88	0.43, 1.8	-
missing	*18*							

	Chronic Gastritis							
	Participants with histopathology data and *Hp* density >0							
	Total n=285				No missing data*** n=271			
Variables	n	OR	95% CI	p*	n	OR	95% CI	p*
Hp Density								
Mild	75	1.0	-	0.00	74	1.0	-	0.00
Moderate	130	7.7	2.9, 20	-	120	6.9	2.6, 18	-
Marked	80	88	29, 266	-	77	92	30, 284	-
Previous *Hp* Treatment								
No	232	1.0	-	0.29	224	1.0	-	0.32
Yes	43	0.69	0.35, 1.4	-	41	0.77	0.39, 1.5	-
missing	*10*							
NSAID Use								
No	111	1.0	-	0.65	107	1.0	-	0.83
Yes	162	0.89	0.55, 1.5	-	156	0.95	0.58, 1.6	-
missing	*12*							
Alcohol Consumption								
No	85	1.0	-	0.67	85	1.0	-	0.67
Yes	112	0.88	0.50, 1.6	-	112	0.88	0.50, 1.6	-
missing	*89*							
Community of Residence								
Aklavik	135	1.0	-	0.00	134	1.0	-	0.00
Old Crow	57	2.6	1.4, 5.0	-	44	2.7	1.3, 5.5	-
Other NT Communities	63	0.81	0.43, 1.5	-	63	0.80	0.43, 1.5	-
Other YT Communities	30	0.22	0.072, 0.66	-	30	0.21	0.071, 0.65	-

*Likelihood ratio chi-square test p-value used for purposeful logistic regression model building

**Includes: Métis; mixed Indigenous ethnicities; unspecified Indigenous ethnicity; Indigenous mixed with non-Indigenous ethnicities; non-Indigenous
***No missing data for age, sex, ethnicity, and community of residence

Table 14. Adjusted prevalence odds ratios for the estimated effect of *Hp* density on active gastritis and chronic gastritis among *Hp*-positive CAN*Help* project participants with histopathology data

	Active Gastritis		**Chronic Gastritis**	
	Adjusted OR*	95% CI	Adjusted OR**	95% CI
Hp Density				
Mild	1.0	-	1.0	-
Moderate	4.7	1.9, 8.7	5.7	2.1, 16
Marked	29	11, 74	81	25, 261
Adjustment Variables				
Sex				
Male	1.0	-	1.0	-
Female	0.75	0.42, 1.3	0.96	0.50, 1.8
Ethnicity				
Inuvialuit (Inuit)	1.0	-	1.0	-
Gwich'in First Nation	1.8	0.88, 3.5	0.98	0.43, 2.2
Other***	2.2	0.95, 5.3	0.53	0.15, 1.9
Community of Residence				
Aklavik			1.0	-
Old Crow			3.6	1.3, 10
Other NT Communities			1.7	0.70, 4.0
Other YT Communities			1.7	0.32, 8.9
Education				
< High School	1.0	-		
High School or equivalent	1.4	0.72, 2.7		
College or University	2.6	1.1, 6.3		

* Adjusted for age as a cubic spline, sex, ethnicity, and education
**Adjusted for age as a cubic spline, sex, ethnicity, and community of residence
***Includes: Métis; mixed Indigenous ethnicities; unspecified Indigenous ethnicity; Indigenous mixed with non-Indigenous ethnicities; non-Indigenous

Chapter 5: Conclusion

A literature review summarizing studies that have estimated agreement and diagnostic accuracy of the UBT, microbiological culture, and histopathology, used alone or in combination, for classifying *Hp* infection status showed that UBT validation studies generally concluded that UBT is as sensitive as histology or culture while other validation studies concluded that culture and histology were not as sensitive as UBT. Summarizing studies that estimated the association of *Hp* density classified by histopathological evaluation with UBT DOB values generally revealed a positive correlation between ^{13}C UBT values and *Hp* density. Lastly, summarizing studies investigating the hypothesis that gastritis severity increases as *Hp* density increases revealed an association or correlation of *Hp* density with gastritis but did not provide any additional analysis or discussion of potential confounding variables.

The validation study presented in Chapter 3 aimed to investigate whether incorporating ancillary stains into the histopathological assessment of gastric biopsies improves the accuracy of *Hp* detection compared to H&E staining alone using gastric tissue samples from CAN*Help* community project participants. This study showed a high degree of concordance on Giemsa, Warthin-Starry and IHC staining for the detection of *Hp* infection in gastric tissue slides, with similar estimates of sensitivity, specificity, and agreement against diverse 'gold standards.' Given comparable accuracy and lower cost, either the Warthin-Starry or Giemsa stain would be more cost-effective than IHC relative to H&E alone for improved detection of *Hp* in gastric biopsy tissue.

This study's results should generalize to other settings with a wide range in *Hp* prevalence because the stratified sampling protocol selected for lower *Hp* densities, thus testing the usefulness of ancillary stains to discriminate between *Hp*-negative tissue samples and tissue samples with low bacterial density. In clinical practice, using ancillary stains may be crucial for

cases with moderate to marked chronic gastritis in which *Hp* is not identified[9,10]. In particular, the results of this study are expected to apply to communities across Canada, including those where the prevalence of *Hp* infection and the associated disease burden is relatively low[116,117].

The future direction for this research includes characterizing the distribution of *Hp* density in distinct anatomical locations in the stomach and comparing tissue samples collected from the antrum, corpus, and incisura. This additional work will help assess the impact of the non-uniform distribution of *Hp* gastric colonization on the accuracy of histopathological classification of *Hp* density. There is also a need to assess the influence of bacterial genotypes and virulence factors on *Hp* density and its impact on histopathology accuracy.

The cross-sectional study presented in Chapter 4 aimed to investigate the hypothesis that the frequency of marked gastritis increases as *Hp* density increases by estimating the association between *Hp* density and the severity of either active or chronic gastritis in CAN*Help* community project participants. This analysis showed a very strong increasing trend in gastritis severity with increasing *Hp* density among *Hp*-positive participants in CAN*Help* community projects.

This analysis characterized the relationship between *Hp* density and gastritis severity in Arctic Indigenous communities. These results are relevant to other populations with high *Hp* prevalence, particularly those with acute and chronic gastritis severity distributions like the study population. The future direction for this research includes assessing whether the anatomical location in the stomach, differentiating tissue samples collected from the antrum, corpus, and incisura, modifies the association between *Hp* density and gastritis severity. This additional work will help assess the impact of the non-uniform distribution of *Hp* bacteria on histopathological outcomes like gastritis severity. To avoid limitations of biopsy sampling and dependence on histopathological classifications of *Hp* density and gastritis severity, UBT DOB values could be

used as an alternate measure of *Hp* density for estimating its effect on gastritis severity. There is also a need to assess the influence of bacterial genotypes on gastritis severity.

References

1. Robin Warren, J. Marshall B. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. The Lancet. 1983;321(8336):1273–5.
2. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Schistosomes, Liver Flukes and *Helicobacter pylori*. In: IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Lyon: International Agency for Research on Cancer; 1994. p. 270.
3. Goodman KJ, Jacobson K, Veldhuyzen Van Zanten S. *Helicobacter pylori* infection in Canadian and related Arctic Aboriginal populations. Can J Gastroenterol. 2008;22(3):289–95.
4. Gisbert JP, Pajares JM. Review article: 13C-urea breath test in the diagnosis of *Helicobacter pylori* infection -- a critical review. Aliment Pharmacol Ther. 2004 Nov 15;20(10):1001–17.
5. Dixon MF, Genta RM, Yardley JH, 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(10):1161–81.
6. Fagan-Garcia K, Geary J, Chang HJ, McAlpine L, Walker E, Colquhoun A, et al. Burden of disease from *Helicobacter pylori* infection in western Canadian Arctic communities. BMC Public Health. 2019 Jun 11;19(1):1–11.
7. Wright CL, Kelly JK. The use of routine special stains for upper gastrointestinal biopsies. Am J Surg Pathol. 2006;30(3):357–61.
8. Benoit A, Hoyeau N, Fléjou JF. Diagnosis of *Helicobacter pylori* infection on gastric biopsies: Standard stain, special stain or immunohistochemistry? Ann Pathol. 2018;38(6):363–9.
9. Smith SB, Snow AN, Perry RL, Qasem SA. *Helicobacter pylori*: To stain or not to stain? Am J Clin Pathol. 2012 May;137(5):733–8.
10. Batts KP, Ketover S, Kakar S, Krasinskas AM, Mitchell KA, Wilcox R, et al. Appropriate use of special stains for identifying *Helicobacter pylori*: Recommendations from the Rodger C. Haggitt Gastrointestinal Pathology Society. Am J Surg Pathol. 2013;37(11):12–22.

11. Kocsmár É, Szirtes I, Kramer Z, Szijártó A, Bene L, Buzás GM, et al. Sensitivity of *Helicobacter pylori* detection by Giemsa staining is poor in comparison with immunohistochemistry and fluorescent in situ hybridization and strongly depends on inflammatory activity. Helicobacter. 2017 Aug 1;22(4):1–9.
12. Ginestet F, Guibourg B, Doucet L, Thréaux J, Robaszkiewicz M, Marcorelles P, et al. Upfront immunohistochemistry improves specificity of *Helicobacter pylori* diagnosis. A French pathology laboratory point of view. Helicobacter. 2017 Oct 1;22(5):1–6.
13. Lash RH, Genta RM. Routine Anti-Helicobacter Immunohistochemical Staining is Significantly Superior to Reflex Staining Protocols for the Detection of Helicobacter in Gastric Biopsy Specimens. Helicobacter. 2016 Dec 1;21(6):581–5.
14. IARC *Helicobacter pylori* Working Group. *Helicobacter pylori* eradication as a strategy for preventing gastric cancer [Internet]. Lyons, France; 2014. Available from: http://www.iarc.fr/en/publications/pdfs- online/wrk/wrk8/index.p*Hp*
15. Sabbagh P, Mohammadnia-Afrouzi M, Javanian M, Babazadeh A, Koppolu V, Vasigala VKR, et al. Diagnostic methods for *Helicobacter pylori* infection: ideals, options, and limitations. Vol. 38, European Journal of Clinical Microbiology and Infectious Diseases. Springer Verlag; 2019. p. 55–66.
16. Kimberlin D, Long S, Brady M, editors. *Helicobacter pylori* Infections. In: Red Book (2021). 30th ed. Chicago: American Academy of Pediatrics; 2021. p. 357–62.
17. Herold R, Becker M. 13C-urea breath test threshold calculation and evaluation for the detection of *Helicobacter pylori* infection in children. BMC Gastroenterol [Internet]. 2002 May 16;2(12). Available from: http://www.biomedcentral.com/1471-230X/2/12
18. Reilly TG, Poxon V, A Sanders DS, J Elliott TS, Walt RP, Reilly Poxon RP Walt T V, et al. Comparison of serum, salivary, and rapid whole blood diagnostic tests for *Helicobacter pylori* and their validation against endoscopy based tests. Gut. 1997;40(4):454–8.
19. Roma-Giannikou E, Roubani A, Sgouras DN, Panayiotou J, Van-Vliet C, Polyzos A, et al. Endoscopic tests for the diagnosis of *Helicobacter pylori* infection in children: Validation of rapid urease test. Helicobacter. 2010;15(3):227–32.
20. Mauro M, Radovic V, Zhou P, Wolfe M, Kamath M, Bercik P, et al. 13C urea breath test for *Helicobacter pylori*: determination of the optimal cut-off point in a Canadian community population. Can J Gastroenterol. 2006;20(12):770–4.

21. Ortiz-Olvera NX, Morán S, Blancas-Valencia JM, Cabrera-Muñoz L. Validation of a simplified 13C-urea breath test method for the diagnosis of *Helicobacter pylori* infection. Rev Esp Enferm Dig. 2007;99(7):392–7.
22. Kwon YH, Kim N, Lee JY, Choi YJ, Yoon K, Hwang JJ, et al. The Diagnostic Validity of Citric Acid-Free, High Dose 13C-Urea Breath Test After *Helicobacter pylori* Eradication in Korea. Helicobacter. 2015 Jun 1;20(3):159–68.
23. Kwon YH, Kim N, Yoon H, Shin CM, Park YS, Lee DH. Effect of citric acid on accuracy of 13C-Urea breath test after *Helicobacter pylori* eradication therapy in a region with a high prevalence of atrophic gastritis. Gut Liver. 2019;13(5):506–14.
24. Delgado R, Tibau XA. Why Cohen's Kappa should be avoided as performance measure in classification. PLoS One. 2019 Sep 1;14(9):1–26.
25. Li ZX, Huang LL, Liu C, Formichella L, Zhang Y, Wang YM, et al. Cut-off optimization for 13C-urea breath test in a community-based trial by mathematic, histology and serology approach. Sci Rep. 2017 Dec 1;7(1):1–8.
26. Storskrubb T, Aro P, Ronkainen J, Vieth M, Stolte M, Wreiber K, et al. A negative *Helicobacter pylori* serology test is more reliable for exclusion of premalignant gastric conditions than a negative test for current *H. pylori* infection: A report on histology and *H. pylori* detection in the general adult population. Scand J Gastroenterol. 2005 Mar;40(3):302–11.
27. Shin CM, Kim N, Lee HS, Lee HE, Lee SH, Park YS, et al. Validation of diagnostic tests for *Helicobacter pylori* with regard to grade of atrophic gastritis and/or intestinal metaplasia. Helicobacter. 2009;14(6):512–9.
28. Laine L, Lewin DN, Naritoku W, Cohen H, Angeles L. Prospective comparison of H&E, Giemsa, and Genta stains for the diagnosis of *Helicobacter pylori*. Gastrointest Endosc. 1997;45(6):463–7.
29. Jonkerst D, Stobberingh E, De Bruine A, Arends JW, Stockbr R. Evaluation of immunohistochemistry for the detection of *Helicobacter pylori* in gastric mucosal biopsies. J Infect. 1997;35(2):149–54.
30. van der Wouden E, Thijs J, van Zwet A, Oey H, Kleibeuker J. Reliability of biopsy-based diagnostic tests for *Helicobacter pylori* after treatment aimed at its eradication. Eur J Gastroenterol Hepatol. 1999;11(11):1255–8.

31. Šeligová B, Lukáč Ľ, Bábelová M, Vávrová S, Sulo P. Diagnostic reliability of nested PCR depends on the primer design and threshold abundance of *Helicobacter pylori* in biopsy, stool, and saliva samples. Helicobacter. 2020 Apr 1;25(2).
32. Sudraba A, Daugule I, Rudzite D, Funka K, Tolmanis I, Engstrand L, et al. Performance of routine *Helicobacter pylori* tests in patients with atrophic gastritis. J Gastrointestin Liver Dis. 2011;20(4):349–54.
33. Rasool S, Abid S, Jafri W. Validity and cost comparison of 14 carbon urea breath test for diagnosis of *H Pylori* in dyspeptic patients. World J Gastroenterol. 2007;13(6):925–9.
34. Bilal R, Khaar B, Qureshi T, Mirza S, Ahmad T, Latif Z, et al. Accuracy of non-invasive 13C-Urea Breath Test compared to invasive tests for *Helicobacter pylori* detection. J Coll Physicians Surg Pak. 2007;17(2):84–8.
35. Frenck RW, Fathy HM, Sherif M, Mohran Z, El Mohammedy H, Francis W, et al. Sensitivity and specificity of various tests for the diagnosis of *Helicobacter pylori* in Egyptian children. Pediatrics. 2006 Oct;118(4):1195–202.
36. Pellicano R, Smedile A, Ponzetto A, Berrutti M, Astegiano M, Saracco G, et al. How accurate is the culture of *Helicobacter pylori* in a clinical setting? An appraisal. Panminerva Med. 2005;47(3):191–4.
37. Öztürk E, Yeşilova Z, Ilgan S, Arslan N, Erdil A, Celasun B, et al. A new, practical, low-dose 14C-urea breath test for the diagnosis of *Helicobacter pylori* infection: clinical validation and comparison with the standard method. Eur J Nucl Med Mol Imaging. 2003;30(11):1457–62.
38. Shirin H, Kenet G, Shevah O, Wardi Y, Birkenfeld S, Shahmurov M, et al. Evaluation of a novel continuous real time 13 C urea breath analyser for *Helicobacter pylori*. Aliment Pharmacol Ther. 2001;15(3):389–94.
39. Yañez P, Madrazo-De La Garza A, Pérez-Pérez G, Cabrera L, Muñoz O, Torres J. Comparison of invasive and noninvasive methods for the diagnosis and evaluation of eradication of *Helicobacter pylori* infection in children. Arch Med Res. 2000;31(4):415–21.
40. Mock T, Yatscoff R, Foster R, Hyun JH, Chung IS, Shim CS, et al. Clinical validation of the Helikit: a 13C urea breath test used for the diagnosis of *Helicobacter pylori* infection. Clin Biochem. 1999;32(1):59–63.

41. Yu WK, Chow PKH, Tan SY, Ng EH, Goh ASW, Soo KC, et al. Five micro-curie urea breath test for the diagnosis of *Helicobacter pylori* infection: evaluation in a South-East Asian population. Aust N Z J Surg. 1999;69(1):37–40.
42. Maconi G, Vago L, Galletta G, Imbesi V, Sangaletti O, Parente F, et al. Is routine histological evaluation an accurate test for *Helicobacter pylori* infection? Aliment Pharmacol Ther. 1999;13(3):327–31.
43. Graham DY, Evans DJ, Alpert LC, Klein PD, Evans DG, Opekun AR, et al. *Campylobacter pylori* detected noninvasively by the 13C-urea breath test. The Lancet. 1987;1(8543):1174–7.
44. Thomas JE, Dale A, Harding M, Coward WA, Cole TJ, Sullivan PB, et al. Interpreting the 13C-urea breath test among a large population of young children from a developing country. Pediatr Res. 1999 Aug;46(2):147–51.
45. Atherton J. Non-endoscopic tests in the diagnosis of *Helicobacter pylori* infection. Aliment Pharmacol Ther. 1997;11(1):11–20.
46. Atherton JC, Washington N, Blackshaw E, Greaves JL, Perkins AC, Hawkey J, et al. Effect of a test meal on the intragastric distribution of urea in the 13C-urea breath test for *Helicobacter pylori*. Gut. 1995;36(3):337–40.
47. Dunn BE, Campbell GP, Perez-Perez GI, Blaser MJ. Purification and characterization of urease from *Helicobacter pylori*. Journal of Biological Chemistry. 1990;265(16):9464–9.
48. Perri F, Clemente R, Pastore M, Quitadamo M, Festa V, Bisceglia M, et al. The 13C-urea breath test as a predictor of intragastric bacterial load and severity of *Helicobacter pylori* gastritis. Scand J Clin Lab Invest. 1998;58(1):19–28.
49. Perri F, Clemente R, Pastore M, Quitadamo M, Festa V, Bisceglia M, et al. The 13 C-urea breath test as a predictor of intragastric bacterial load and severity of *Helicobacter pylori* gastritis. Scand J Clin Lab Invest. 1998;58(1):19–28.
50. Vincent P, Michaud L, Martin De Lasalle E, Benon B, Turck D, Gottrand F. 13C-urea breath test and gastric mucosal colonization by *Helicobacter pylori* in children: quantitative relation and usefulness for diagnosis of infection. Helicobacter. 1999;4(4):233–7.
51. Desroches JJ, Lahaie RG, Picard M, Morã¡is J, Dumont A, Gaudreau C, et al. Methodological validation and clinical usefulness of carbon-14-urea breath test for

documentation of presence and eradication of *Helicobacter pylori* infection. J Nucl Med. 1997;38:1141–5.

52. Zagari RM, Pozzato P, Martuzzi C, Fuccio L, Martinelli G, Roda E, et al. 13C-urea breath test to assess *Helicobacter pylori* bacterial load. Helicobacter. 2005;10(6):615–9.
53. Sheu BS, Lee SC, Yang HB, Lin XZ. Quantitative result of 13C urea breath test at 15 minutes may correlate with the bacterial density of *H. pylori* in the stomach. Hepatogastroenterology. 1999;47(27):2057–62.
54. Kobayashi D, Eishi Y, Ohkusa Ã T, Ishige I, Suzuki T, Minami J, et al. Gastric mucosal density of *Helicobacter pylori* estimated by real-time PCR compared with results of urea breath test and histological grading. J Med Microbiol. 2002;51(4):305–11.
55. Epple HJ, Kirstein FW, Bojarski C, Frege J, Fromm M, Riecken EO, et al. 13C-urea breath test in *Helicobacter pylori* diagnosis and eradication. Correlation to histology, origin of "false" results, and influence of food intake. Scand J Gastroenterol. 1997;32(4):308–14.
56. Rothman K, Greenland S, Haneuse S, Lash T. Modern epidemiology [Internet] [Internet]. 4th ed. Philadelphia: Wolters Kluwer; 2021. Available from: https://search-ebscohost-com.login.ezproxy.library.ualberta.ca/login.aspx?direct=true&db=cat03710a&AN=alb.9534916&site=eds-live&scope=site
57. Kumar D, Bal CS, Dattagupta S, Ahuja V, Mathur M, Sharma MP. 14C urea breath test does not predict density of *Helicobacter pylori* in duodenal ulcer disease. Indian J Med Res. 2001;113:129–34.
58. Machado RS, Kawakami E, Da Silva Patrício FR, Reber M. Urease activity does not reflect the degree of colonization by *Helicobacter pylori* in children. Pediatrics International. 2006 Aug;48(4):398–402.
59. Rasheed F, Khalid A, Saadia A, Habib S, Khan SM, Saeed F, et al. Correlation of 13C urea breath test values with *Helicobacter pylori* load among positive patients. Hepatogastroenterology. 2014;61(129):125–8.
60. Lai YC, Yang JC, Huang SH. Pre-treatment urea breath test results predict the efficacy of *Helicobacter pylori* eradication therapy in patients with active duodenal ulcers. World J Gastroenterol. 2004;10(7):991–4.

61. Chang YW, Min SK, Kim KJ, Han YS, Lee JH, Dong SH, et al. Delta (13)C-urea breath test value is a useful indicator for *Helicobacter pylori* eradication in patients with functional dyspepsia. J Gastroenterol Hepatol. 2003;18(6):726–31.
62. Chang MC, Chang YT, Sun CT, Wu MS, Wang *Hp*, Lin JT. Quantitative correlation of *Helicobacter pylori* stool antigen (*Hp*SA) test with 13C-urea breath test (13C-UBT) by the updated Sydney grading system of gastritis. Hepatogastroenterology. 2002;49(44):576–9.
63. Suto H, Azuma T, Ito S, Ito Y, Miyaji H, Yamazaki Y, et al. Endoscopic [13C]-urea breath test for quantification of *Helicobacter pylori* infection. J Gastroenterol Hepatol. 2000;15(2):161–7.
64. Perri F, Clemente R, Festa V, Quitadamo M, Conoscitore P, Niro G, et al. Relationship between the results of pre-treatment urea breath test and efficacy of eradication of *Helicobacter pylori* infection. Ital J Gastroenterol Hepatol. 1998;30(2):146–50.
65. Ellenrieder V, Glasbrenner B, Stoffels C, Weiler S, Bode G, Möller P, et al. Qualitative and semi-quantitative value of a modified 13C-urea breath test for identification of *Helicobacter pylori* infection. Eur J Gastroenterol Hepatol. 1997;9(11):1085–9.
66. Labenz J, Bärsch G, Peitz U, Aygen S, Hennemann O, Tillenburg B, et al. Validity of a novel biopsy urease test (HUT) and a simplified 13C-urea breath test for diagnosis of *Helicobacter pylori* infection and estimation of the severity of gastritis. Digestion. 1996;57(6):391–7.
67. Tiwari A, Rai R, Dahal P, Regmi S. Prevalence of *Helicobacter pylori* in endoscopic gastric biopsies of chronic gastritis patients at a tertiary care centre. JNMA J Nepal Med Assoc. 2020;58(228):564–8.
68. Nai GA, Carolina A, Parizi G, Barbosa RL. Association between *Helicobacter pylori* concentration and the combining frequency of histopathological findings in gastric biopsies specimens. Arq Gastroenterol. 2007;44(3):240–3.
69. Tutar E, Ertem D, Kotiloglu Karaa E, Pehlivanoglu E. Endoscopic and histopathologic findings associated with *H. pylori* infection in very young children. Dig Dis Sci. 2009 Jan;54(1):111–7.
70. Gallo N, Zambon CF, Navaglia F, Basso D, Guariso G, Piva MG, et al. *Helicobacter pylori* infection in children and adults: a single pathogen but a different pathology. Helicobacter. 2003;8(1):21–8.

71. Dzierzanowska-Fangrat K, Crabtree JE, Rozynek E, Dura W, Celińska-Cedro D, Wojda U, et al. *Helicobacter pylori* cagA genotype and density of colonization in relation to gastric inflammation in children. Eur J Gastroenterol Hepatol. 2002;14(2):1303–7.
72. Zsikla V, Hailemariam S, le Baumann M, Mund MT, Schaub N, Meier R, et al. Increased rate of *Helicobacter pylori* infection detected by PCR in biopsies with chronic gastritis. Am J Surg Pathol. 2006;30(2):242–8.
73. Şenocak Taşçl E, Akbaş T. The Relationship between the Sydney Classification and the First-Line Treatment Efficacy in Helicobacter -Associated Gastritis. Med Princ Pract. 2020 Dec 28;29(6):551–7.
74. Molaei M, Foroughi F, Mashayekhi R, Haghazali M, Zojaji H, Jafari F, et al. CagA status and VacA subtypes of *Helicobacter pylori* in relation to histopathologic findings in Iranian population. Indian J Pathol Microbiol. 2010 Jan 1;53(1):24–7.
75. Garg B, Sandhu V, Sood N, Sood A, Malhotra V. Histopathological analysis of chronic gastritis and correlation of pathological features with each other and with endoscopic findings. Pol J Pathol. 2012;63(3):172–8.
76. Jaramillo-Rodríguez Y, Nares-Cisneros J, Martínez-Ordaz VA, Velasco-Rodríguez VM, López Márquez FC, Manríquez-Covarrubias LE. Chronic gastritis associated with *Helicobacter pylori* in Mexican children: histopathological patterns. Pediatr Dev Pathol. 2011 Mar;14(2):93–8.
77. Tummala S, Sheth SG, Goldsmith JD, Goldar-Najafi A, Murphy CK, Osburne MS, et al. Quantifying gastric *Helicobacter pylori* infection: A comparison of quantitative culture, urease breath testing, and histology. Dig Dis Sci. 2007 Feb;52(2):396–401.
78. Misra V, Misra S, Dwivedi M, Singh PA, Bhargava V, Gupta SC. A topographic study of *Helicobacter pylori* density, distribution and associated gastritis. J Gastroenterol Hepatol. 2000;15(7):737–43.
79. Fareed R, Abbas Z, Shah A. Effect of *Helicobacter pylori* density on inflammatory activity in stomach. J Pak Med Assoc. 2000;50(5):148–51.
80. Chen X, Haruma K, Kamada T, Mihara M, Komoto K, Yoshihara M, et al. Factors that affect results of the 13C urea breath test in Japanese patients. Helicobacter. 2000;5(2):98–103.

81. Yamaoka Y, Kodama T, Kita M, Imanishi J, Kashima K, Graham Y, et al. Relation between clinical presentation, *Helicobacter pylori* density, interleukin 1beta and 8 production, and cagA status. Gut. 1999;45(6):804–11.
82. Yamamura F, Yoshikawa N, Akita Y, Mitamura K, Miyasaka N. Relationship between *Helicobacter pylori* infection and histologic features of gastritis in biopsy specimens in gastroduodenal diseases, including evaluation of diagnosis by polymerase chain reaction assay. J Gastroenterol. 1999;34(4):461–6.
83. Atherton JC, Tham KT, Peek RM, Cover TL, Blaser MJ. Density of *Helicobacter pylori* infection in vivo as assessed by quantitative culture and histology. J Infect Dis [Internet]. 1996;174(3):552–8. Available from: https://academic.oup.com/jid/article/174/3/552/902697
84. Khulusi S, Mendall MA, Patel P, Levy J, Badve S, Northfield TC. *Helicobacter pylori* infection density and gastric inflammation in duodenal ulcer and non-ulcer subjects. Gut. 1995;37(3):319–24.
85. Alam K, Schubert TT, Bologna SD, Ma CK. Increased density of *Helicobacter pylori* on antral biopsy is associated with severity of acute and chronic inflammation and likelihood of duodenal ulceration. Am J Gastroenterol. 1992 Apr;87(4):424–8.
86. Logan RP, Polson RJ, Misiewicz JJ, Rao G, Karim NQ, Newell D, et al. Simplified single sample 13Carbon urea breath test for *Helicobacter pylori*: comparison with histology, culture, and ELISA serology. Gut. 1991;32(12):1461–4.
87. Osaki T, Mabe K, Hanawa T, Kamiya S. Urease-positive bacteria in the stomach induce a false-positive reaction in a urea breath test for diagnosis of *Helicobacter pylori* infection. J Med Microbiol. 2008 Jul;57(7):814–9.
88. Seo K, Yang HR, Jeong A. Diagnostic accuracy of the 13C-urea breath test in children: Adjustment of the cut-off value according to age. J Gastroenterol Hepatol. 2005;20:264–9.
89. CANHelp Working Group. Aklavik *H. pylori* Project Progress Report [Internet]. 2017. Available from: http://canhelpworkinggroup.ca/wp-content/uploads/AklavikProgressReport_20171022.pdf
90. NWT Bureau of Statistics. 2006 Census Population and Dwelling Counts [Internet]. 2007. Available from: https://www.statsnwt.ca/census/2006/Pop%20and%20Dwelling%20counts_2006.pdf

91. Yukon Bureau of Statistics. Population and Dwellings: Census 2011 [Internet]. 2011. Available from: https://yukon.ca/sites/yukon.ca/files/ybs/fin-population-dwellings-census-2011.pdf
92. CANHelp Working Group. ISR (Tuktoyaktuk) *H. pylori* Project Progress Report [Internet]. 2017. Available from: http://canhelpworkinggroup.ca/wp-content/uploads/ISR_ProgressReport_20170810.pdf
93. NWT Bureau of Statistics. 2011 Census Population and Dwelling Counts [Internet]. 2012. Available from: https://www.statsnwt.ca/census/2011/Pop%20and%20Dwelling.pdf
94. CANHelp Working Group. Fort McPherson Progress Report [Internet]. 2017. Available from: http://canhelpworkinggroup.ca/wp-content/uploads/FortMcPherson_ProgressReport_20171022-2.pdf
95. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016 Jun 1;15(2):155–63.
96. Ii Y, Hiro S, Nakazuru Y. Use of diagnostic likelihood ratio of outcome to evaluate misclassification bias in the planning of database studies. BMC Med Inform Decis Mak. 2022 Dec 1;22(1).
97. Guyatt G, Rennie D, Meade MO, Cook DJ. Chapter 16: The Process of Diagnosis. In: Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, 3rd ed [Internet]. New York, NY: McGraw-Hill Education; 2015. Available from: jamaevidence.mhmedical.com/content.aspx?aid=1147670269
98. Rogan WJ, Gladen B. Estimating prevalence from the results of a screening test. Am J Epidemiol. 1978;107(1):71–6.
99. Cockburn M, Cox B. The effect of measurement error on the determination of *Helicobacter pylori* prevalence. Epidemiology. 1997;8(2):205–9.
100. Zagari RM, Pozzato P, Martuzzi C, Fuccio L, Martinelli G, Roda E, et al. 13C-Urea Breath Test to Assess *Helicobacter pylori* Bacterial Load. Helicobacter. 2005;10(6):615–9.
101. Correa P, Houghton JM. Carcinogenesis of *Helicobacter pylori*. Gastroenterology. 2007;133(2):659–72.
102. Kuipers EJ. Review article: exploring the link between *Helicobacter pylori* and gastric cancer. Aliment Pharmacol Ther. 1999;13(1):3–11.

103. Yin Y, Liang H, Wei N, Zheng Z. Prevalence of chronic atrophic gastritis worldwide from 2010 to 2020: an updated systematic review and meta-analysis. Ann Palliat Med. 2022 Dec;0(0):0–0.
104. Ghasemi Basir HR, Ghobakhlou M, Akbari P, Dehghan A, Seif Rabiei MA. Correlation between the intensity of *Helicobacter pylori* colonization and severity of gastritis. Gastroenterol Res Pract. 2017;2017:1–5.
105. Serhat S. The relation between *Helicobacter pylori* density and gastritis severity. International Archives of Internal Medicine. 2019 Sep 21;3(2):1–4.
106. Choudhary C, Bhanot U, Agarwal A, Garbyal R. Correlation of *H. pylori* density with grading of chronic gastritis. Indian J Pathol Microbiol. 2001;44(3):325–8.
107. Liou JM, Malfertheiner P, Lee YC, Sheu BS, Sugano K, Cheng HC, et al. Screening and eradication of *Helicobacter pylori* for gastric cancer prevention: The Taipei global consensus. Vol. 69, Gut. BMJ Publishing Group; 2020. p. 2093–112.
108. Kalkan IH, Sapmaz F, Güliter S, Atasoy P. Severe gastritis decreases success rate of *Helicobacter pylori* eradication. Wien Klin Wochenschr. 2016 May 1;128(9–10):329–34.
109. Onal IK, Gokcan H, Benzer E, Bilir G, Oztas E. What is the impact of *Helicobacter pylori* density on the success of eradication therapy: A clinico-histopathological study. Clin Res Hepatol Gastroenterol. 2013 Dec;37(6):642–6.
110. NWT Bureau of Statistics. 2016 Census Population and Dwelling Counts [Internet]. 2017. Available from: https://www.statsnwt.ca/census/2016/Pop%20and%20dwelling%20counts%20Feb-8-2017.pdf
111. Yukon Bureau of Statistics. Population and Dwellings: Census 2016 [Internet]. 2016. Available from: https://yukon.ca/sites/yukon.ca/files/ybs/fin-population-dwellings-census-2016.pdf
112. Shah SC, Piazuelo MB, Kuipers EJ, Li D. AGA Clinical Practice Update on the Diagnosis and Management of Atrophic Gastritis: Expert Review. Gastroenterology. 2021 Oct 1;161(4):1325-1332.e7.
113. Sipponen P, Maaroos HI. Chronic gastritis. Scand J Gastroenterol. 2015 Jun 1;50(6):657–67.

114. Pearce N. Effect measures in prevalence studies. Environ Health Perspect. 2004;112(10):1047–50.
115. Bursac Z, Gauss CH, Williams DK, Hosmer DW. Purposeful selection of variables in logistic regression. Source Code Biol Med. 2008 Dec 16;3(17):1–8.
116. Sinha SK, Martin B, Sargent M, McConnell JP, Bernstein CN. Age at acquisition of *Helicobacter pylori* in a pediatric Canadian First Nations population. Helicobacter. 2002;7(2):76–85.
117. Brown LM. *Helicobacter pylori*: epidemiology and routes of transmission. Epidemiol Rev. 2000;22(2):283–97.