OPEN ACCESS

University of BRISTOL

This electronic thesis or dissertation has been downloaded from Explore Bristol Research, http://research-information.bristol.ac.uk

Author:
Chong, Amanda

Title:
Examining whether Helicobacter pylori has a causal effect on cardiovascular disease and cancer

General rights
Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy
Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

• Your contact details
• Bibliographic details for the item, including a URL
• An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

Examining whether *Helicobacter pylori* has a causal effect on cardiovascular disease and cancer

Amanda Hui Wan Chong

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Masters by Research in Population Health Sciences in the Faculty of Population Health Sciences.

Bristol Medical School, January, 2019

Word count (excluding references and appendix): 15,531

Abstract

Infection with *Helicobacter pylori* (*H.pylori*) is estimated to persist in approximately 50% of the human population. This bacterium has been commonly linked to gastrointestinal diseases such as gastritis and peptic ulcers. Current observational studies have also suggested an association with cardiovascular disease and cancer. However, there has been discordance in these findings potentially influenced by confounding. This study aims to identify single nucleotide polymorphisms (SNPs) associated with *H.pylori*, estimate the causal association of *H.pylori* with cardiovascular disease and cancer traits, and examine the direction of causality. These objectives were explored using the Avon Longitudinal Study of Parents and Children (ALSPAC) and Caerphilly Prospective Study (CaPS) cohorts. The use of genome-wide association meta-analysis was employed, and four highly suggestive SNPs possibly associated with *H.pylori* were identified. Two genome-wide significant SNPs identified in a previous published *H.pylori* genome-wide association study were used as instruments in two-sample Mendelian randomization (MR), with the four suggestive SNPs identified in the meta-analysis being excluded from analysis as they would not qualify as valid instruments, potentially violating MR assumptions. There was evidence of a causal effect of *H. Pylori* on LDL-cholesterol, hip circumference, breast cancer, and heart rate. However, the causal estimates suggested that *H.pylori* might be associated with a decrease in these traits which is in contrast to observational findings. Bidirectional MR revealed little evidence of causal effects of the outcomes on *H.pylori* and sensitivity analyses did not identify directional pleiotropy across instruments for each trait or heterogeneity between instruments. Overall, this study extends the scope of MR to infections and does not suffer from the limitations of observational studies, such as confounding, selection biases and reverse causation. These findings contribute to the understanding of the role of *H.pylori* in cardiovascular disease and cancer.

Dedication

I dedicate this dissertation to my family and partner Joe for their unwavering support.

Acknowledgements

First and foremost, I would like to acknowledge my supervisors, Dr Jie Zheng, Dr Evie Stergiakouli, and Dr Tom Gaunt for their academic support and guidance, and for making this Masters degree a wonderful experience.

I would like to thank the participants of the Avon Longitudinal Study of Parents and Children cohort and the Caerphilly Prospective Study cohort, for contributing to the work presented in my dissertation.

I am also very grateful to the PhD students in my office for their support and encouragement throughout this process.

Finally, but in no means the least, I would like to extend my profound gratitude to my parents and Joe for being a constant source of support and reassurance.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: DATE:

Abbreviations

ADIPOGen = Adiponectin Genetics Consortium

ALSPAC = Avon Longitudinal Study of Parents and Children

ATBC = Alpha-Tocopherol, Beta-Carotene Cancer Prevention

BMI = Body mass index

C4D = Coronary Artery Disease Genetics Consortium

CagA = Cytotoxin-associated gene A

CaPS = Caerphilly Prospective Study

CARDIoGRAM = Coronary ARtery Disease Genome-wide Replication and Meta-analysis

CI = Confidence interval

CRP = C-reactive protein

CVD = Cardiovascular disease

DAG = Directed acyclic graph

DIAGRAM = DIAbetes Genetics Replication and Meta-analysis

DNA = Deoxyribonucleic acid

EA = Effect allele

EAF = Effect allele frequency

EGG = Early Growth Genetics Consortium

ELISA = Enzyme-linked immunosorbent assay

FCGR2A = Fc Fragment of IgG Receptor IIa

GIANT = Genetic Investigation of ANthropometric Traits

GLGC = Global Lipids Genetics Consortium

GWAS = Genome-wide association study

H.pylori = *Helicobacter pylori*

HDL = High-density lipoprotein

HRGene = Heart Rate Consortium

ICAM-1 = Intercellular adhesion molecule-1

IgG = Immunoglobulin G

IL-1 = Interleukin-1

IL-8 = Interleukin-8

ILCCO = International Lung Cancer Consortium

InSIDE = Instrument Strength Independent of Direct Effect

ISGC = International Stroke Genetics Consortium

IVW = Inverse variable weighting

LD = Linkage disequilibrium

LDL = Low-density lipoprotein

MAGIC = Meta-Analyses of Glucose and Insulin-related traits Consortium

MALT = Mucosa-associated lymphoid tissue

MDACC = MD Anderson Cancer Centre

MESA = Multi-Ethnic Study of Atherosclerosis

MR = Mendelian randomization

NA = Not applicable

OA = Other allele

OR = Odds ratio

P = P-value

PanScan = Pancreatic Cancer Cohort Consortium

PC = Principal components

PheWAS = Phenome-wide association studies

QC = Quality check

Q-Q plot = Quantile-quantile plot

RCT = Randomised controlled trial

r_g = Genetic correlation

RR = Relative risk

RS = Rotterdam Study

SD = Standard deviation

SE = Standard error

SHIP = Study of Health in Pomerania

SNP = Single nucleotide protein

TLR = Toll-like receptor

USA = United States of America

VacA = Vacuolating cytotoxin gene A

WIPF1 = WAS/WASL Interacting Protein Family Member 1

Z = Z-score

ZEMPA = Zero Modal Pleiotropy Assumption

β = Beta

Table of Contents

List of Tables

List of Figures

1. Introduction

Since the discovery of *Helicobacter pylori* (*H.pylori*) by Marshall and Warren in 1983(1, 2) by successfully isolating and culturing the bacterium, gastric colonisation with *H.pylori* has been linked to various gastrointestinal diseases(3-5). In particular, strong evidence from observational studies have suggested *H.pylori* to be a disease risk factor for gastric cancer(6-8), and have postulated a potential association with cardiovascular disease(9-11) (Table 1). Interestingly, despite the increasing number of observational epidemiological studies investigating the relationship between both cardiovascular disease and gastric cancer with *H.pylori*, studies have not examined whether *H.pylori* is causally associated with cardiovascular disease and cancer using other methodological approaches.

Table 1. Cardiovascular disease outcomes and gastric cancer subtypes that have been investigated in observational studies in association with *Helicobacter pylori* infection

Disease	Disease subtype	References
Cardiovascular disease	Atherosclerosis	(12), (13), (14), (15)
	Coronary heart disease	(16), (17)
	Stroke	(18), (19), (20)
	Myocardial infarction	(21), (22), (23), (24)
Cancer	Non-cardia gastric cancer	(25), (26), (27)
	Cardia gastric cancer	(28), (29)

Observational studies can be potentially hindered by confounding, reverse causation, selection biases and regression dilution bias(30). Results from observational studies do not always replicate in randomised controlled trials (RCTs), as seen in trials investigating beta carotene supplementation and risk of lung cancer(31), and the protective effects of various vitamin supplements and hormone replacement therapies on cardiovascular disease(32-35). In order to overcome these limitations, genetic epidemiological methodologies such as Mendelian randomization (MR) can be employed. The basic principle of this approach is that common genetic polymorphisms can be used to proxy for a modifiable exposure (e.g.

LDL cholesterol, blood pressure, C-reactive protein), or influence exposure patterns (e.g. propensity to smoke) without suffering from confounding and reverse causation(36) .

The current management approach to mitigate the effects of *H.pylori*-induced diseases has been eradication treatment(37). However, increasing *H.pylori* resistance to once effective combined antibiotic treatments have challenged the feasibility and efficacy of this approach, as studies have shown failed *H.pylori* elimination or reinfection(38, 39). Therefore, the implications of findings using Mendelian randomization can be useful in informing novel public health policies to improve population health through population-level interventions. In a broader sense, this research could contribute to lessening the public health burden of these two diseases, as recent statistics have shown that gastric cancer is the fourth most common cancer(40) and second leading cause of cancer death, with cardiovascular disease as the leading cause of morbidity and mortality(41).

Additionally, with the development of recent genome-wide association studies identifying genetic variants associated with common infections, such as *H.pylori(42)*, using Mendelian randomization is a timely approach. The main objective of this study will be to use Mendelian randomization to assess the causal association of *H.pylori* with cardiovascular disease and cancer. This approach can help consolidate research findings from observational studies, and provide a more robust understanding of the potential causal pathways that may exist.

2. *Helicobacter pylori*

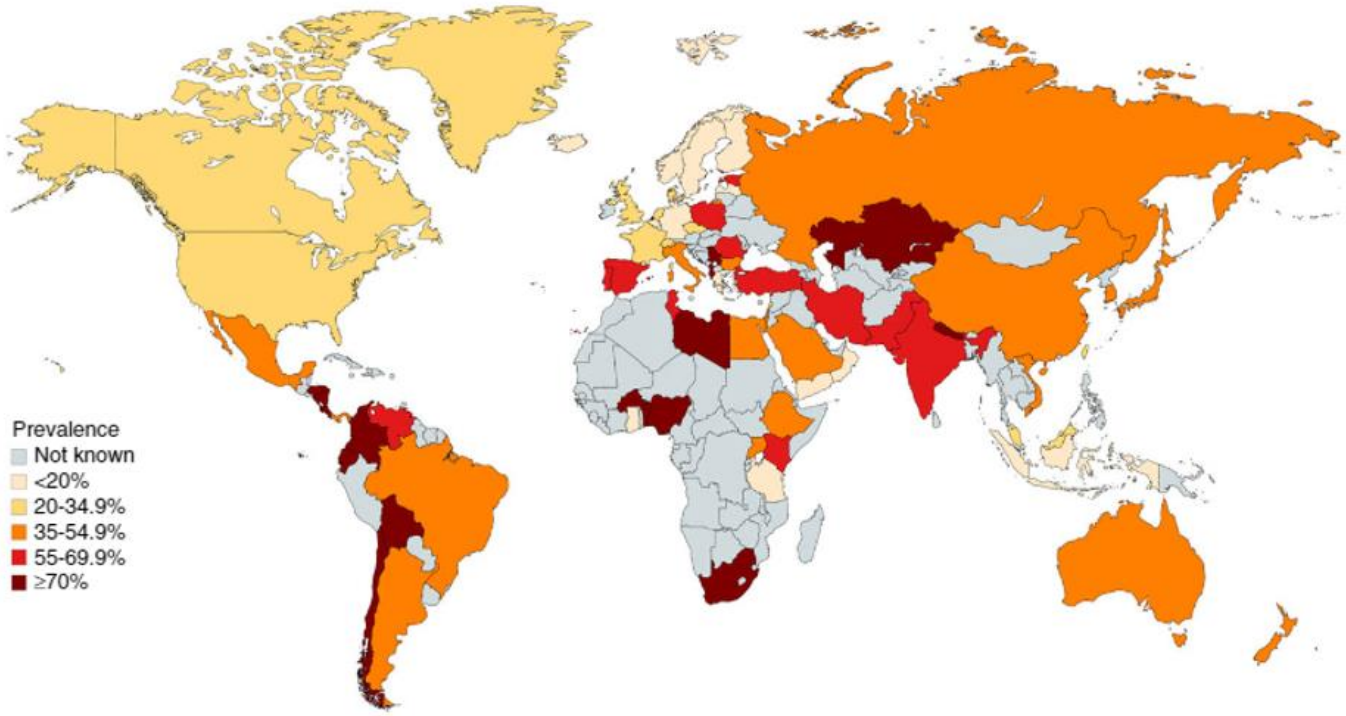
2.1 **Prevalence of bacterium**

The prevalence of *H.pylori* infection is estimated to persist in at least 50% of the world's human population(43, 44). However, the overall prevalence in developed countries is lower than developing countries(45, 46). The acquisition of *H.pylori* infection is predominantly in childhood, with modes of transmission hypothesised to be via person-to-person transmission by familial exposure of large intrafamilial household clustering , oral-to-oral route or faecal-oral route, or by waterborne or zoonotic transmission(47-49). Unless this bacterium is eradicated due to the pathogenesis of *H.pylori*-related diseases such as gastritis and duodenal ulcerations in childhood, it can persist as the child grows into an adult(50-52). The prevalence of *H.pylori* in children in developed countries is approximately 10-40% while it ranges from 80-100% for children in developing countries (53, 54).

This difference in prevalence between developed and developing regions worldwide can be exemplified in a recent systematic review performed by Zamani et al. 2018(55) (Figure 1). Between continents, *H.pylori* showed greater prevalence in Africa (56.7%, 95% confidence interval (CI): 43.6-69.9), Latin America and the Caribbean (59.3%, 95% CI: 52.9-65.6), and Asia (44.7%, 95% CI: 49.4-50.0). This is in

contrast to Northern America (28.5%, 95% CI: 20.7-36.3) and Europe (35.3%, 95% CI: 29.5-41.2) which showed a noticeably lower *H.pylori* prevalence. Large variations between countries within a continent were also observed, for example in Asia between Kazakhstan (79.5%, 95% CI: 74.8-84.2) and Taiwan (22.4%, 95% CI: 16.3-28.4). This association between the prevalence of *H.pylori* infection and the developmental status of countries has been considered to be attributable to socioeconomic factors such as high household density(56, 57), low income(58, 59), and the level of education of the child and parents(58, 60).

Figure 1. Global prevalence of *H.pylori* infection(55)



Interestingly, declining trends in *H.pylori* prevalence have been observed since around the 1950s and have been suggested to be related to the decrease in gastric cancer incidence and mortality rates(61). This decline is thought to reflect the increase in sanitation, level of urbanisation, decrease in household overcrowding, and higher socioeconomic status of countries(62, 63). For example, in Japan gastric cancer has been the leading type of cancer and cancer-related deaths(64). However, due to improved sanitation *H.pylori* prevalence has decreased(65). Children born before 1950 were measured to have a *H.pylori* seroprevalence of approximately 80-90% in contrast to those born after 2000 who have a seroprevalence of less than 2%(65).However, it is important to note that despite the declining trends in *H.pylori* prevalence, *H.pylori* infection and its related disease outcomes still remain an important public health issue(17, 66-68). Its burden

on public health can be highlighted in its role in the pathogenesis of cancer, with chronic *H.pylori* infection estimated to account for 90% of non-cardia gastric cancer cases worldwide(69).

2.2 **Observational studies: Associations with cardiovascular disease and cancer**

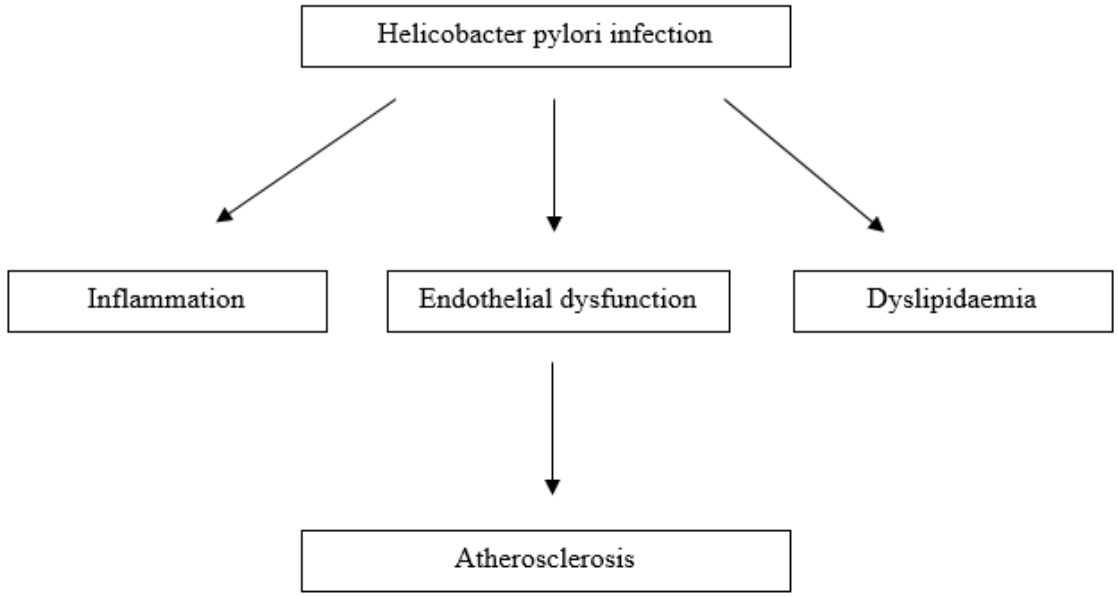
2.2.1 Cardiovascular disease

H.pylori infection has been postulated to be a risk factor for cardiovascular disease. However, an accumulation of inconsistent evidence has put into question the etiological role of *H. pylori*. Mechanisms have been hypothesised to elucidate how *H.pylori* infection increases the risk of cardiovascular disease, with observational epidemiological studies investigating its possible association with atherosclerosis(14, 15, 70), myocardial infarction(18, 71), coronary heart disease(16, 72, 73), strokes(19, 74, 75), and risk factors for cardiovascular diseases(76-78). To highlight the discordance in findings, atherosclerosis, coronary heart disease and stroke have been selected.

2.2.1.1 *Atherosclerosis*

H.pylori has been hypothesised to be involved in the pathophysiological pathways leading to the development of atherosclerosis(15, 70, 79). It has been observed to be more prevalent in subjects with systemic atherosclerosis(80-82), and studies have suggested that infection may trigger ischemic events such as myocardial and cerebral infarction(18, 21, 23, 74) (Figure 2).

Figure 2. Hypothesised biological mechanisms for the association of *H. pylori* infection with atherosclerosis



One possible pathway that *H.pylori* is proposed to encourage atherosclerosis is by affecting lipid metabolism(83). Hoffmeister et al. 2001(84) observed an atherogenic lipid profile in *H.pylori*-positive subjects. These subjects presented with low HDL-cholesterol levels, considerably lower HDL-cholesterol to total cholesterol, higher apolipoprotein B levels, and lower serum apolipoprotein A1 levels, in comparison to the control group who were *H.pylori* seronegative. The observation of lower HDL cholesterol levels could be important as HDL-cholesterol is considered to be antiatherogenic(85, 86), possesses anti-inflammatory and antioxidant properties(87-89), and has the capacity to impede endothelial cell damage(90, 91). A decrease in levels of HDL cholesterol could therefore potentially allow for the development of atherosclerosis. However, other studies investigating this same association have shown small to no notable differences in lipid fractions in *H.pylori*-positive participants(92, 93). Similarly, Mendelian randomization showed limited evidence of HDL cholesterol having a causal effect on atherosclerotic-related diseases such as myocardial infarction (Odds ratio (OR): 0.93; 95% CI: 0.68 – 1.26)(94) and coronary heart disease (OR: 0.81; 95% CI: 0.44 – 1.46)(95). The simplistic pathway of lowered HDL cholesterol driving the development of atherosclerosis has also been challenged, with recent literature arguing that it is a more complex biological mechanism(96).

Chronic *H.pylori* infection has also been implicated in the development of atherosclerosis through systemic and vascular inflammation and endothelial dysfunction(97-99). Oshima et al. 2005(81) supported this hypothesis through the measurement of high-sensitivity C-reactive protein (CRP), a marker of systemic inflammation, and flow-mediated vasodilation. This study illustrated elevated CRP levels in *H.pylori* seropositive subjects greater than the normal reference laboratory value of <1.3 mg/L. Furthermore, flow-mediated vasodilation in *H.pylori*-seropositive subjects was attenuated in comparison to *H.pylori* seronegative subjects. These findings were also similarly observed in a study by Gen et al. 2010(78). However, Mendelian randomization findings challenge the causal association between CRP and coronary heart disease, a disease outcome promoted by the build-up of atherosclerotic plaque. A Mendelian randomization meta-analysis using 47 epidemiological studies showed limited evidence of a causal association with a risk ratio for coronary heart disease of 1.00 (95% CI: 0.90 – 1.13) per one standard deviation (SD) of genetically higher natural log concentration of CRP(100). The interpretation from this study would therefore suggest that CRP does not causally affect coronary heart disease.

The colonisation of *H.pylori* in human carotid atherosclerotic plaque has also been demonstrated in various studies(12, 101-103). Ameriso et al. 2001(14) found *H.pylori* present in 53% of carotid atherosclerotic plaques, and found no trace of *H.pylori* in carotid arteries without atherosclerosis. Additionally, expression of intercellular adhesion module-1 (ICAM-1), a marker of increased endothelial inflammatory activity, was present in 75% of patients with *H.pylori* DNA. However, although this study

supports the notion that direct arterial invasion could encourage the pathogenesis of atherosclerosis and atherosclerotic diseases, other studies have not found any evidence of *H.pylori* presence in atherosclerotic plaque(104-106).

In a recent systematic review(13), the relationship between *H.pylori* infection and atherosclerosis was put into question. This review meta-analysed four included studies investigating the presence of *H.pylori* in the atherosclerotic plaque of patients with atherosclerotic-related vascular disease. In a fixed effects model, the pooled odds ratio was 4.65 (95% CI: 1.99 – 10.85; I^2 = 67.6%)(13). These findings showed limited evidence that *H.pylori* is a risk factor for atherosclerosis, but also strong evidence of heterogeneity between studies. In this study, the power to detect and association was inadequate due to the limited number of included studies, each study was constrained by a small sample size, and sampling methods varied between each study. Therefore, due to lack of concordance between study designs and a need for large sample sizes, the association between *H.pylori* and atherosclerosis still remains unclear.

2.2.1.2 *Coronary heart disease*

Since a pilot study conducted by Mendall et al. 1994(17) reported a higher prevalence of *H.pylori* infection in patients with coronary heart disease than healthy controls, subsequent studies have investigated the association between *H.pylori* infection and coronary heart disease with contradicting results.

Positive associations with *H.pylori* infection were observed by Pellicano et al. 1999(24), Kinjo et al. 2002(21), Kahan et al. 2000(107), and Khodaii et al. 2011(23). These studies observed that *H.pylori* infection was more prevalent in patients diagnosed with myocardial infarction than the control groups. Similarly, it was also observed that higher seropositivity of *H.pylori* was found in patients diagnosed with unstable angina in comparison to the matched control group(24). Building on these observational findings, Tabata et al. 2016(22) reported that Japanese patients burdened by the *H.pylori* pathogen and polymorphisms of the proinflammatory cytokine interleukin-1 (IL-1) beta had substantially higher levels of high sensitivity CRP, and a greater risk of ST-segment elevation myocardial infarction.

In larger studies investigating *H.pylori* seropositivity and coronary heart disease, some evidence or limited evidence of an association were shown(73, 108-112). In all included studies, analyses were adjusted for age, sex, markers of socioeconomic status (e.g. age of ended full time education, estimated household income, region of residence), smoking history, blood pressure, lipid profile, and measures of height and weight (e.g. body mass index). In addition, these studies used similar methods of measuring *H.pylori* from plasma samples and measuring *H.pylori* specific IgG titres by enzyme-linked immunosorbent assay (ELISA) . These studies benefitted from a larger sample size, with the power to

detect an association, and adjustment of related risk factors that could otherwise confound and inflate results. However, although these methods overcome some of the limitations that hinder small observational study findings, these studies could still be biased from unadjusted or unknown confounders.

Similar findings from systematic reviews over the last 12 years illustrated some evidence of an association with coronary heart disease. The most recent systematic review performed by Yu et al. 2017(113) showed some evidence of an association with an odds ratio of 1.96 (95% CI: 1.47 – 2.63). Previous systematic reviews revealed comparable findings with odds ratios of 2.11 (95% CI: 1.70 – 2.62)(79) and 1.87 (95% CI: 1.46 – 2.40)(114).

2.2.1.3 *Stroke*

The hypothesis that chronic *H.pylori* infection is associated with the pathogenesis of ischemic stroke was first investigated by Markus and Mendall in 1998(74). They reported that, after adjusting for socioeconomic status and cardiovascular disease risk factors, *H.pylori* seropositivity showed a positive association with cerebrovascular disease. Additionally, this study also found that for different ischemic stroke subtypes *H.pylori* infection showed a positive association with large vessel disease and small artery occlusion, but was not associated with stroke subtypes caused by cardioembolism. Similarly, in a small case-control study(115) and in a large Taiwanese nationwide population-based retrospective cohort study(20), non-embolic ischemic strokes were shown to be positively associated with *H.pylori* infection.

Three separate systematic reviews using case-control studies estimated the association of *H.pylori* infection on the risk of ischemic stroke, and findings from all showed some evidence of an association with odds ratios of 1.87 (95% CI: 1.46 – 2.40)(114), 1.49 (95% CI: 1.24 – 1.81)(116), and 1.60 (95% CI: 1.21 – 2.11)(117), respectively. These reviews however were limited by the included studies which consisted of small case-control study designs that might have increased small study bias. In a more recent systematic review using large prospective studies, the pooled odds ratio was 0.96 (95% CI: 0.78 – 1.14)(118).

2.2.2 Cancer

In light of increasing evidence from seroepidemiological and nested case-control studies supporting the association of *H.pylori* infection with gastric cancer, the International Agency for Research on Cancer and the World Health Organisation in 1994 classified *Helicobacter pylori* as a group 1 carcinogen (119). Studies investigating the role of *H.pylori* infection with various types of cancers, such as those subsequently discussed, contributed to this evidence.

The acquisition of *H.pylori* has been attributed to the development of gastric adenocarcinoma and gastric mucosa-associated lymphoid tissue (MALT) lymphoma(120-123). It has been observed that amongst

H.pylori infected individuals, approximately 10% will develop peptic ulcers, 1-3% will be diagnosed with gastric adenocarcinoma, and <0.1% will be identified to have MALT lymphoma(124). In addition to bacterial infection, influence from host susceptibility, genetic susceptibility and immune response as well as environmental factors also contribute to the multifactorial aetiology of gastric cancer(125, 126). This then could increase risk to transition to atrophic gastritis, to metaplasia then dysplasia and finally adenocarcinoma(47, 127).

H.pylori infection has been strongly associated with the development of non-cardia gastric cancer(25, 27, 128). In the Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study, Kamangar et al. 2006(128) reported that in Finnish males (Cases = 234; Age-matched controls = 234) there was a 7.9 fold increased risk of non-cardia gastric cancer (95%CI: 3.0-20.9) after adjustment for age, sex, socioeconomic status, gastric cancer risk factors, and host nutrition. These findings are in agreement with a previous meta-analysis conducted by the Helicobacter and Cancer Collaborative Group(129) in 2001 that concluded that *H.pylori* seropositivity increased the risk of non-cardia gastric cancer (OR: 5.9, 95% CI: 3.4-10.3). Additionally, they suggested that *H.pylori* serology collected at least ten years or more before patient cancer diagnosis was a better indicator of the magnitude of *H.pylori* association with non-cardia cancer. This is because retrospective studies tend to underestimate its association due to the loss of *H.pylori* infection in cases with the onset of cancer. This absence of infection can occur as precancerous lesions can modify the environment in the stomach destroying niches in which *H.pylori* colonies thrived. The study also found that infection with *H.pylori* was restricted to non-cardia gastric cancer, and did not increase risk to cardia gastric cancer. This observation has been similarly shown in most studies within Western populations(26, 130-133). However, studies in East-Asian cohorts have shown a positive association between *H.pylori* seropositivity and cardia gastric adenocarcinomas(134, 135).

To understand the discordant findings, Cavaleiro-Pinto et al. 2011(6) performed a meta-analysis to investigate the association between *H.pylori* infection and both non-cardia gastric cancer and cardia gastric cancer in low risk gastric cancer settings (Australia, Germany, Finland, Norway, USA) and countries with a high risk of gastric cancer (China, Japan, Korea). They found that high-risk countries showed a greater positive association with cardia gastric cancer (Relative risk (RR): 1.98, 95% CI: 1.38-2.83) and non-cardia gastric cancer (RR: 3.02, 95% CI: 1.92-4.74) in comparison to low-risk countries. These findings may contribute to the difference in findings, supporting the hypothesis of a heterogenous distribution of cardia gastric cancer that are etiologically distinct from one another(28, 29). Another suggestion for the conflicting findings is the range of definitions for cardia gastric cancer in different countries in terms of its pathological and clinical classification(136, 137).

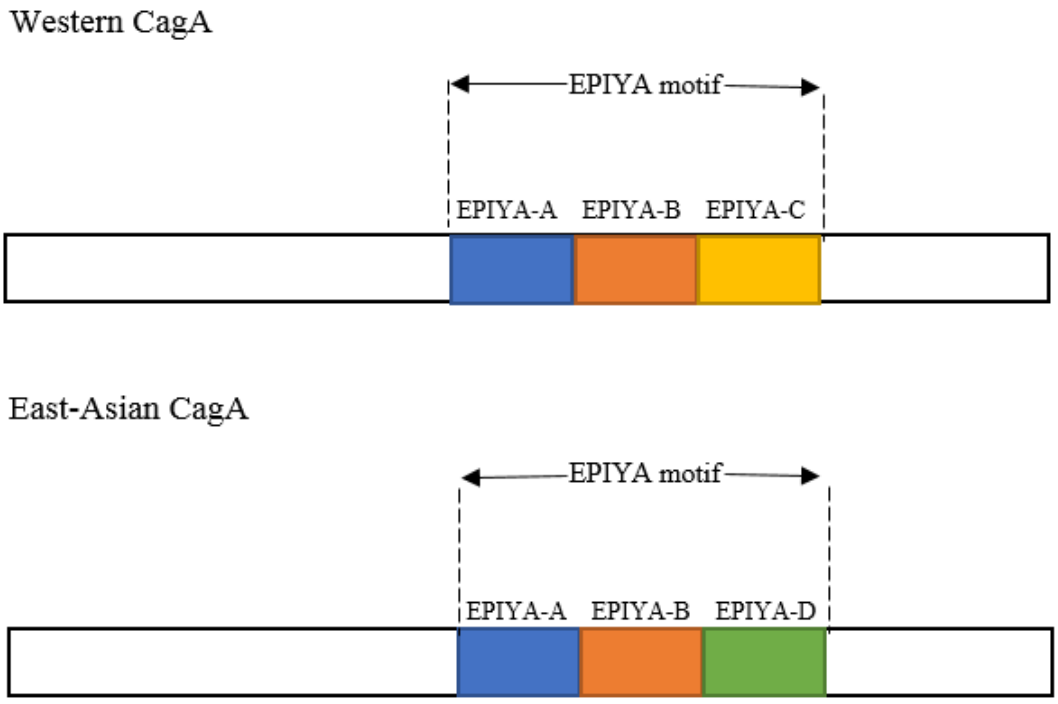
In addition, the association of *H.pylori* infection and the pathogenesis of cancer has also been suggested to be attributable to the high genetic diversity between *H.pylori* strains with strain-specific genotypes conferring different levels of virulence(138, 139). Since the complete genome sequencing of *H.pylori* in 1997(140), high rates of mutation and intraspecies recombination have been observed showing a divergence of *H.pylori* strains(141). Bacterial colonisation of virulent *H.pylori* strains have been seen to be associated with the geographical location of individuals and their ancestral origin (142, 143), and this will be discussed in further detail later. In particular, virulence factors cytotoxin-associated gene A (*cagA*), located in the 40kb chromosomal region known as the *cag* pathogenicity island, and the vacuolating cytotoxin gene A (*vacA*) have been intensely studied based on their presence or absence in *H.pylori* strains(144). Polymorphisms of these genes have shown different levels of cytotoxin activity, with the subsequent chapters, 2.2.2.1 Cytotoxin-associated gene A and 2.2.2.2 Vacuolating cytotoxin gene A, discussing studies investigating the association of these genes with the development of cancer.

2.2.2.1 *Cytotoxin-associated gene A*

The link between cytotoxin-associated gene A (*cagA*) antigen and its ability to confer malignancy has been investigated, with numerous studies showing an increased risk of gastric cancer. This bacterium-derived oncoprotein has also been shown to trigger the development of tumours and gastric carcinomas in transgenic mouse models and in a Mongolian gerbil model(145-149).

The mechanism by which CagA infiltrates host cells is important as the sequence diversity among *cagA* genes influences pathogenicity(150-154). In the carboxyl-terminal regions of CagA antigens are EPIYA (glutamic acid-proline-isoleucine-tyrosine-alanine) motifs which are sites where tyrosine phosphorylation occur(151, 155). These EPIYA motifs can be defined as EPIYA-A, EPIYA-B, EPIYA-C, EPIYA-D, with EPIYA-A and EPIYA-B motifs presents in almost all CagA sequences(156, 157). Interestingly, the addition of the EPIYA-C motif is observed in Western strains, and the EPIYA-D motif is seen in East-Asian strains(143, 156, 158) (Figure 3). Studies have shown that the East-Asian CagA subtype is more virulent than the Western subtype as it is more effective at binding to SHP-2 to cause deregulation and morphological transformations in host cells to activate oncogenic pathways(150, 151, 158). The East-Asian CagA sequence is also shown to induce the production of a greater level of the inflammatory cytokine interleukin-8 (IL-8) than the Western subtype(152, 159), and can act to potentiate oncogenic cell proliferation(160). These findings complement clinical studies that reported a greater association between the East-Asian subtype and chronic active inflammation, atrophic gastritis, and increased risk of gastric cancer in comparison to the Western CagA subtype(153, 154, 161).

Figure 3. Structural diversity of the *H.pylori* cytotoxin-associated gene A (adapted from Hatakeyama et al. (162))

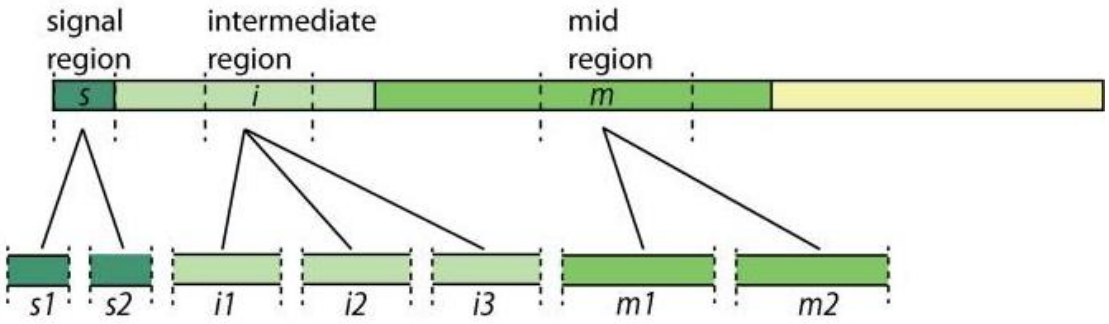


These differences in virulence between the East-Asian CagA sequence (EPIYA motif ABD type) and the Western CagA sequence (EPIYA motif ABC type) could potentially contribute to explaining the pattern in prevalence of gastric cancer. Despite the frequency of *H.pylori* infection decreasing in East-Asian countries(93, 163, 164), studies have reported a higher prevalence of gastric cancer in these countries compared to Western countries(165-167). This could potentially be due to the occurrence of the *cagA* gene in Western populations (e.g. Europe, North America, Australia) to only be present in approximately 60% or less of *H.pylori* strains(168-171). In contrast, East-Asian populations (i.e. Japan, Korea, China) have shown a prevalence of 80% to 100% of CagA-positive *H.pylori* strains(172-174).

2.2.2.2 *Vacuolating cytotoxin gene A*

Polymorphisms among the vacuolating cytotoxin gene A (*vacA*) genotypes have been demonstrated to contribute to different degrees of cytotoxicity and are possibly associated with different levels of gastrointestinal diseases from peptic ulcers to gastric cancer(175-178) (Figure 4).

Figure 4. Genetic structure of *H.pylori* vacuolating cytotoxin gene A (adapted from Palframan et al. (179))



The signal (s) region which occurs as either s1 or s2 allelic types can be further subtyped into s1a or s1b(180, 181). Studies have shown *vacA* s1 strains to be associated with enhanced and chronic gastric mucosal inflammation to cause epithelial cell damage, possibly due to its increased virulence potential to produce a greater proportion of vacuolating cytotoxin(178, 182, 183). It has also been associated with producing higher levels of IL-8 and intestinal metaplasia(184, 185). Furthermore, s1a allelic types have shown a correlation with increased mucosal neutrophil activity and lymphocyte infiltration in comparison to the s1b and s2 strains(178). Studies such as these have led to the notion that *H.pylori* strains bearing the s1 allele could increase risk to gastric cancer more than those *H.pylori* strains containing the s2 allele. Interestingly, many of these studies also investigated the associated of *cagA* and gastric cancer, as most strains of *H.pylori* that carry the *cagA*-positive genotypes also present the more cytotoxic form of the *vacA* allele, s1(183, 184, 186). This contrasts with *cagA*-negative strains which are commonly present with the non-cytotoxic *vacA* s2 allele(187). Although the interplay between CagA and VacA is still not fully known, functional antagonism has been observed that could affect signal transduction in host cells, and this interaction is postulated to have an effect on the severity of disease manifestation(188-190).

The mid (m) region, categorised as m1 or m2, is responsible for determining the cell specificity of the cytotoxin with the different subtypes differing in their receptor binding domains(191-193). In vivo,

Atherton et al. 1997(178) reported that epithelial injury showed the strongest association with the *vacA* mid region genotype, with m1 strains associated with a greater propensity to cause injury in the antrum and corpus of the stomach compared to m2 strains. The m1 strains have also been shown to be cytotoxic to a wider range of host epithelial cells(193, 194). When coupled with other *vacA* genotypes, Western population studies have shown that the *vacA* s1/m1 genotype are associated with gastric ulcers and gastric carcinoma(183, 195, 196). However, the s2/m2 strains have been shown to be almost non-cytotoxic(197). Interestingly, East-Asian strains of *H.pylori* have been observed to carry the *vacA* s1 allele and predominantly the m1 allele(167, 198, 199). This observation is noteworthy as the East-Asian countries of Japan and Korea have been reported to have high incidences of gastric cancer(200, 201). In South Asia and southern regions of East Asia the m2 allele is becoming more prevalent, and in these areas incidences of gastric cancer are shown to be lower(202, 203).

Furthermore, the intermediate (i) region, subtyped into i1,i2 and i3, have been suggested to be a determinant for virulence and pathogenicity(204, 205). Rhead et al. 2007(206) reported that within Western countries, the s1/m1 strains consistently presented type i1 and displayed vacuolating activity. Whilst s2/m2 strains were subtype i2 and did not display vacuolating activity(207). Similar to i2, the rare type i3 was shown to not exhibit vacuolating properties(207). It was also observed that s1/m2 strains varied in their intermediate region and were shown to have either the i1 or i2 allele(206). *vacA* s1/i1/m1 genotype have been strongly associated with the increased risk of gastric cancer and peptic ulcers in Iranian and Italian populations(186, 208). However these studies are discordant to findings from Ogiwara et al. 2009(209). They reported that in Western and Asian populations, there was no increased virulence in strains of the s1/i1/m1 genotype compared to strains of the s1/i1/m2 genotype when looking at their influence on the risk of developing peptic ulcers and gastric cancer.

2.3 **Genetic studies**

The genome-wide association meta-analysis published by Mayerle et al. 2013(42) was the first genetic study to identify genetic variants associated with *Helicobacter pylori* infection. This study used 10,938 participants of European ancestry from two independent population-based cohorts, the Study of Health in Pomerania (SHIP) and the Rotterdam Study (RS). Findings from the meta-analysis revealed two genetic variants strongly associated with *H.pylori* seroprevalence: rs10004195 (OR: 0.70; 95% CI: 0.65 - 0.76) located on the toll-like receptor (*TLR*) gene on 4p14, and rs368433 (OR: 0.73; 95% CI: 0.65 - 0.81) found in an intron of the Fc Fragment Of IgG Receptor IIa (*FCGR2A*) gene on 1q23.3. Genes encoded on the 4p14 region were identified as *TLR1, TLR6,* and *TLR10,* with *TLR1* identified to be the gene with the most plausible biological pathway associated with *H.pylori* seroprevalence. The protein encoded by *TLR1* is an innate immune response receptor which forms a heterodimer with TLR2 to stimulate immune

responses when lipopeptides, present on a wide range of pathogens such bacteria and viruses, are recognised(210-212). To explore the function of TLR1 further, Mayerle et al. 2013(42) recruited additional participants to investigate gene expression levels of *TLR1, TLR6* and *TLR10* on the 4p14 region using whole blood samples. Findings showed that rs10004195, the SNP strongly associated with *H.pylori* seroprevalence on *TLR*, showed evidence of an association with mRNA levels of *TLR1* in SHIP ($P = 2.1 \times 10^{-4}$) and RS ($P = 3.2 \times 10^{-17}$)(42). In addition, *TLR1* was the only gene that was differentially expressed per copy number of the minor rs10004195-A allele (β: -0.23; 95% CI: -0.34 - -0.11)(42). An interpretation of these results suggest that a lower expression of TLR1 may be associated with a protective effect against the acquisition of *H.pylori* infection(213). Conversely, an increased expression of TLR1 may suggest an increased likelihood of acquiring *H.pylori* infection and persistence(213). However, the biological mechanism of *TLR1* and *H.pylori* infection has not been defined, and further research is required to elucidate the pathways. In addition, the effect of the different *H.pylori* strains was not examined. Replication of this study would also be required in other ethnic groups which are observed to have a higher prevalence of virulent *H.pylori* strains(167, 214, 215).

To date, this is the only genetic study to investigate the genetic basis of *H.pylori* seroprevalence. This lack of genetic studies highlights the challenges attributable to performing a GWAS on bacterial infections, such as *H.pylori*(216, 217). *H.pylori* has a high genetic diversity and some strains have been suggested to be asymptomatic while others can result in disease development(139, 141, 153, 181). Classifying individuals as seropositive for cases and seronegative for controls could be an oversimplification of the different *H.pylori* phenotypes that confer different levels of disease risk. Delineating between *H.pylori* phenotypes (e.g. individuals infected with *H.pylori* with presence of *cagA*) could provide a more accurate classification of individuals who are seropositive and have a putative *H.pylori* virulence marker that has been shown to encourage disease development. However, a caveat to this suggestion is that individuals can be infected by multiple strains of *H.pylori(195, 218, 219)*. This can result in different strains colonizing different anatomic regions of the gastric mucosa, resulting in different disease outcomes depending on where *H.pylori* colonizes. Therefore, these limitations illustrate some of the challenges that can arise when performing bacterial GWASs.

3. Limitations of observational studies

Although numerous observational studies, both retrospective and prospective, provide evidence of associations between *H.pylori* seropositivity and both cardiovascular disease and cancer, these studies are prone to confounding, reverse causation, selection biases and regression dilution bias(30). The failures of some of the findings from observational study to replicate in RCTs investigating the same hypotheses is a major concern as research costs to implement RCTs are high.

Confounding can affect the findings of observational studies as unadjusted or unknown factors can influence the apparent effect of the exposure on the outcome(220). For example, failure to include factors like socioeconomic status as a covariate and appropriately controlling for its effect can bias findings and result in an overestimated positive association between *H.pylori* infection and coronary heart disease. This can be seen in a study by Ponzetto et al. 1996(221) who reported a fourfold increased risk of coronary heart disease when not accounting for social class and cardiovascular disease risk factors. Similarly, Aceti et al. 1996(222) reported a fivefold increased risk in coronary heart disease. These accentuated disease risks can be attributable to low socioeconomic status not being adjusted for, as this factor is strongly correlated with coronary heart disease and *H.pylori* infection(73, 76).

Moreover, observational findings can be biased by reverse causation which can generate spurious associations with exposures that have been found to be non-causal when investigated using RCTs. Reverse causation can occur when the disease outcome influences the exposure of interest. For example, a hypothetical situation could be that *H.pylori* IgG titers rise after a myocardial infarction event. This would then suggest that myocardial infarctions drive an increase in *H.pylori* infection, and not the reverse pathway which is the proposed biological mechanism.

Associations in observational studies can also become biased due to selection(30). Study settings can cause selection bias as selected cases and controls may not be representative of the general population. For example, studies conducted in specific workplaces, such as hospitals, or selection according to a specific lifestyle choice, such as participants who are vegan. Selection bias can be exemplified in a study performed by Mendall et al. 1994(17). The study was observed to not be representative of the general population, with the included participants consisting of only white British men aged 45-65 years old from a single general practice clinic. This study showed some evidence of *H.pylori* infection having a positive association with coronary heart disease (OR: 2.15; 95% CI: 1.07 – 4.29). However, these results can be said to be biased due to the selection of participants from only one gender type, one type of ethnicity, from a specific age group, and a small geographical setting. Therefore, to interpret these results as applicable to the general population would be misleading.

Attenuation by error, later renamed ‘regression dilution bias’, is also a limitation that can hinder observational study findings(223). This occurs when random imprecision in measurements of an exposure variable causes an attenuation of the regression slope towards the null. This bias will then result in an underestimation of the regression slope (i.e. beta coefficient) as estimates of the association between an exposure and disease outcome are biased in a downward bias(224). For example, when blood pressure is measured by a sphygmomanometer, random error due to rounding error or variation in day-to-day blood pressure may result in imprecise measurement.

Therefore, although observational studies report associations on *H.pylori* with cardiovascular disease and cancer, evidence of causal associations from these studies can be hindered by the limitations mentioned above. Well-designed randomised controlled trials (RCT) are considered the gold standard to infer causal associations between an exposure and outcome of interest. However, the use of this method can be limited by ethical restraints, feasibility, high cost, and are known to be very time consuming(225, 226). For example, infecting people with *H.pylori* would be unethical given the risk of developing cancer and cardiovascular disease. In addition, vaccinating children against *H.pylori* infection could be performed in a RCT setting, however to see the benefits of vaccination would take years and this would be too costly. Thus, Mendelian randomization provides an alternative approach to infer causality using germline genetics that overcome limitations that are observed in observational studies.

4. Mendelian randomization

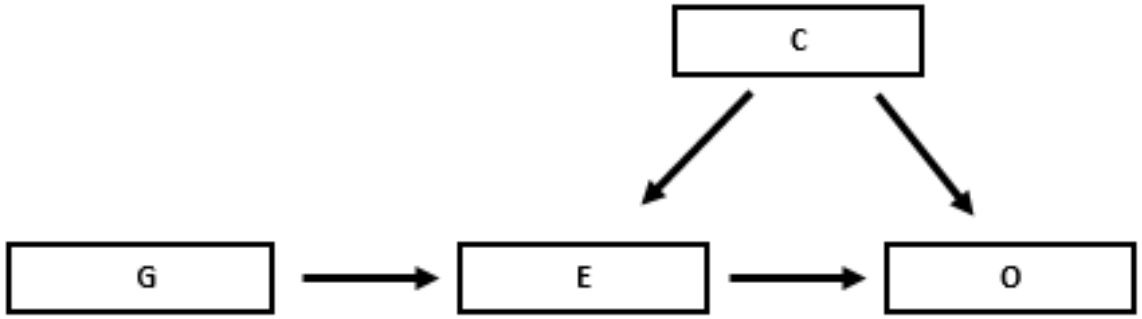
Mendelian randomization uses germline genetic variants as instrumental variables ("proxies") for exposures of interest (e.g. biological traits, environmental factors) to estimate the causal effects of these exposures on disease outcomes(227).

The framework of Mendelian randomization can be intuitively likened to randomized controlled trials as a natural analogue to infer causality(228). Instead of randomizing participants into different levels of treatment as in RCTs, genotypes in Mendelian randomization randomize individuals into different levels of exposure in order to avoid confounding and determine causality. It is important, however, to be mindful that although this analogy allows for a better understanding of Mendelian randomization, this analogy is not perfect(229).

To infer causality using Mendelian randomization methodologies, core assumptions must be met for the results to be valid (Figure 5). These core assumptions are:

1. The genetic variant must be associated with the exposure of interest.
2. The genetic variant should not be associated with confounders of the exposure-outcome association.
3. The genetic variant must be associated with the outcome of interest through the exposure of interest.

Figure 5. Causal directed acyclic graph (DAG) of Mendelian randomization methodology. Genetic variants (G) are used as proxies for a modifiable exposure of interest (E) to examine the causal association between E and an outcome of interest (O), without the limitations of reverse causation and confounding (C).



Mendelian randomization can be used to help overcome the limitations of observational epidemiological studies. Firstly, as germline genetic variants are randomly allocated and independently assorted at conception, genotypes are largely independent of biological, socioeconomic and physiological factors minimizing confounding. Secondly, since germline genetic variants are fixed at conception, they are non-modifiable and cannot be influenced by reverse causation. Thirdly, as the associations of germline genetic variants with modifiable (non-genetic) exposures are normally thought to persist through the lifecourse, associations cannot be attenuated by random imprecision in measurement of the exposure which reduces regression dilution bias. Finally, genetic variants are not generally influenced by factors which determine how participants are selected in a study, overcoming selection bias. This approach therefore allows for more reliable estimates of the causal effects of exposures on outcomes, as compared to conventional observational epidemiological studies(230).

4.1 **Mendelian randomization limitations**

Mendelian randomization has its downfalls and estimates of the causal effect can become biased as a result of limitations such as pleiotropy, confounding due to linkage disequilibrium and population stratification, sample size, lack of robust genetic variants associated with the exposure of interest, and trait heterogeneity in genetic variants, which can lead to misleading findings(30). To mitigate potential pitfalls, suggestions have been discussed in greater depth in other papers to address these issues in order to

provide more reliable, precise causal estimates(227, 229). Limitations specific to examining the causal association of *H.pylori* infection with cardiovascular disease and cancer that must be considered are the lack of robust genetic instruments as proxies for *H.pylori* infection as the exposure of interest, and the limited biological understanding of the genetic variants used as instrumental variables. These limitations can affect both the robustness of the causal inference and the interpretation of MR findings.

4.1.1 Lack of robust genetic instruments

To date, only two SNPs have been identified that show strong evidence of an association with *H.pylori* seroprevalence: rs10004196 (OR: 0.7; 95% CI: 0.65 – 0.76) and rs368433 (OR: 0.73; 95% CI: 0.65 – 0.81)(42). These SNPs were identified in a GWAS meta-analysis by Mayerle et al. 2013(42) which consisted of 10,938 participants in total from the Study of Health in Pomerania cohort and the Rotterdam Study cohort (Studies discussed in detail in Chapter 7 Mendelian randomization). Using a small number of SNPs as proxies for *H.pylori* infection in MR analysis can result in low statistical power to detect a causal effect and imprecision in causal estimates due to large confidence intervals(229). This is because each SNP will only explain a small amount of variance in *H.pylori* as the exposure of interest, and power is a function of sample size, variance explained by the SNP, strength of confounding, causal effect size, and type 1 error rate. The *H.pylori* GWAS meta-analysis performed in this study using the ALSPAC and CaPS cohorts will be implemented in order to identify more SNPs to be used as proxies for *H.pylori* infection in MR analysis. However, these SNPs will need to satisfy the core MR assumptions in order to qualify as valid instruments for causal inference.

4.1.2 Limited biological understanding of the instrumental variables

The limited understanding of the biological function of the *H.pylori* genetic variants used as instrumental variables can be problematic as horizontal pleiotropy can bias the causal inference. This can occur if the *H.pylori* genetic variants affect the disease outcome (i.e. cardiovascular disease and cancer) via an independent biological pathway not through the exposure of interest, therefore violating the 'exclusion restriction criterion'. In order to mitigate against this limitation, measures such as assessing heterogeneity between genetic instruments, as well as tests for pleiotropy using MR-Egger regression, weight median, and weighted mode approaches, can indicate the strength of bias from horizontal pleiotropy(229, 231) (MR methods discussed in detail in Chapter 4.2 Mendelian randomization methods to estimate the causal effect). Furthermore, the complexity of the underlying biological mechanisms in the association between *H.pylori* infection and cardiovascular disease and cancer, respectively, can be misleading if MR interpretations are overly simplified. An improved understanding of the function of *H.pylori* genetic instruments used in MR analysis through molecular biology approaches would be required in order to elucidate the exact mechanisms involved.

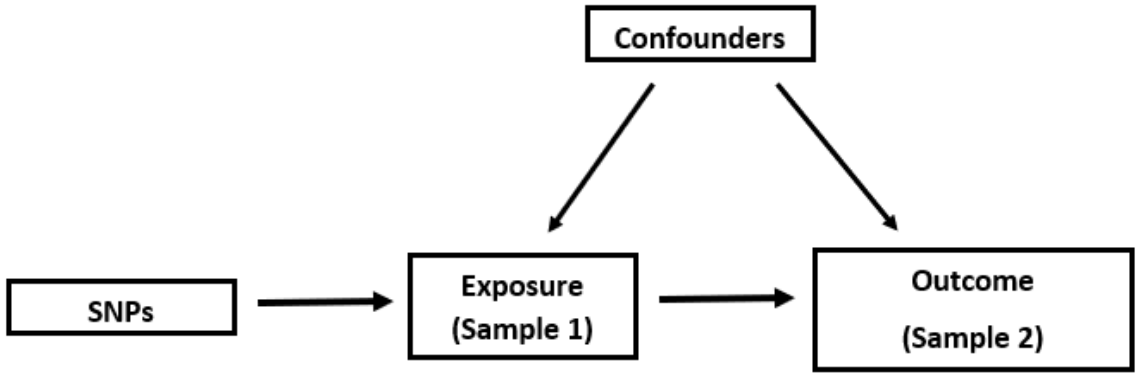
4.2 **Mendelian randomization study designs**

Extensions to the standard one-sample Mendelian randomization design have been developed. These approaches extend the scope of standard MR and are used when one-sample MR might not applicable or appropriate method to infer causality. These extensions include two-sample MR, bidirectional MR, multivariable MR, and factorial MR. Two-sample MR and bidirectional MR are discussed in further detail as these approaches have been utilized in this study.

4.2.1 Two-sample Mendelian randomization

The implementation of two-sample MR has become increasingly common practice as this approach uses summary statistics on gene-exposure and gene-outcome estimates(232-234). The advantage of this method is that individual-level data from a single sample is not necessary, thus making studies possible in cases where individual level data are not available or cannot be shared (235). Instead, two-sample MR, as the name suggests, uses summary statistics from two independent samples for the exposure and outcome variables (Figure 6). In addition, this approach benefits from the impact of weak instrument bias being in the direction of the null which is conservative and will not lead to inflated Type 1 error rates(236), compared to one-sample MR which is biased towards the confounded observational estimate(237).

Figure 6. DAG of two-sample Mendelian randomization paradigm. The causal association between the exposure variable and outcome variable are obtained from separate non-overlapping samples



The main advantage over standard MR is the use of the increasing amount of publicly available summary data from large GWAS consortia(235). By using two independent samples rather than one, the statistical power to detect a causal effect is increased as a result of increasing the overall sample size, and this can lead to a greater precision of the causal estimate. This is particularly advantageous when assessing the causal effects on dichotomous disease outcomes which require larger sample sizes than continuous outcomes(238, 239). The likelihood of 'winner's curse' bias (i.e. chance correlation between genetic variants and confounders in the GWAS discovery stage, generally leading to an overestimation of the SNP-exposure effect) is also likely to be reduced due to increased statistical power(240). Furthermore,

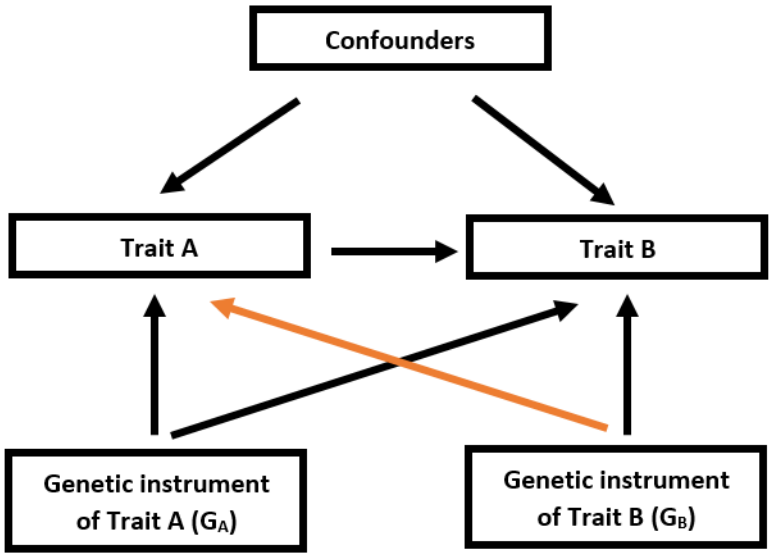
violations to the core MR assumptions can be better detected using sensitivity analyses such as MR-Egger regression, weighted mode and weighted median approaches (Discussed in further detail in Chapter 4.3 Mendelian randomization methods to estimate the causal effect).

In addition to the core MR assumption, additional criteria must be met which require the samples from the gene-exposure and the gene-outcome associations to not overlap(241). However, overlap can occur if cohorts contribute to GWAS analysis of the exposure of interest and outcome of interest. If samples do overlap, the extent of the overlap is important(235). If the overlap is large the causal effect estimates will be biased in the direction of the observational estimate. Furthermore, it is important to consider whether the two samples are homogenous and representative of the same underlying population. For example, ethnicity, age and sex should be checked to ensure that both samples represent the same population.

4.2.2 Bidirectional Mendelian randomization

To determine the direction of causality, bidirectional two-sample MR can be employed using robust genetic instruments from independent GWASs, for both the exposure of interest and outcome of interest. This method evaluates whether the "exposure" variable causally influences the "outcome", or whether the "outcome" variable causally influences the "exposure"(242). Figure 7 illustrates that if trait A is causally associated with trait B, then the genetic instrument G_A is associated with trait A and trait B. However, as shown with the orange arrow, the genetic instrument G_B which is associated with trait B will not be associated with trait A.

Figure 7. DAG of bidirectional Mendelian randomization design strategy



Implementation of bidirectional MR can be demonstrated with a study performed by Timpson et al. 2011(243) examining the causal direction between measured BMI and circulating CRP to evaluate discordant findings in observational studies. Some studies suggested that CRP promotes the development of elevated adiposity, whilst others proposed that obesity is a determinant of an inflammatory state of the body and subsequently a marker of inflammation that includes CRP levels(244-246). This study inferred that the direction of causation was likely to be driven by BMI, with elevated CRP being an inflammatory marker(243). However, there are caveats to this approach as it assumes that a single causal direction underlies the biological mechanism for the causal association(229). This can be an oversimplification of complex biological pathways and may result in misleading interpretations.

4.3 **Mendelian randomization methods to estimate the causal effect**

4.3.1 Inverse-variance weighted

The inverse-variance weighted (IVW) method is the traditional MR method to estimate causal effect. For this method to calculate a consistent estimate of the causal effect either: 1) All genetic variants (i.e. SNPs) must satisfy the instrumental variable assumptions or; 2) The sum of the horizontal pleiotropic effects of each instrument must be zero (i.e. balanced horizontal pleiotropy) and pleiotropic effects are independent in magnitude of the instrument strength across all instruments (i.e. the Instrument Strength Independent of Direct Effect (InSIDE) assumption) (247, 248). When multiple genetic instruments are utilized in MR analysis, this method is essentially a meta-analysis of the Wald ratio estimates of the causal effect from each genetic instrument to provide an overall causal estimate (i.e. IVW estimate)(229). This IVW causal estimate assumes that the Wald ratio estimates from the genetic instruments are uncorrelated, and therefore provide independent evidence on the causal effect(249). The same IVW causal estimate can be obtained using a weighted linear regression of SNP-outcome associations on SNP-exposure associations(249).

To detect and adjust for heterogeneity, the Cochran's Q statistic can be employed when using the IVW method. This test can be applied as it requires the same assumptions to hold that are required to estimate the IVW causal estimate(248). If the Q statistic is observed to be much larger than the degrees of freedom minus one, this is indicative of heterogeneity(250, 251). This could suggest horizontal pleiotropy between genetic instruments and a violation to the exclusion restriction assumption (i.e. The genetic variant is only associated with the outcome of interest through the exposure of interest).

4.3.2 MR-Egger regression

MR-Egger regression provides a useful sensitivity analysis to the IVW method as it does not assume that pleiotropic effects of the SNP-outcome association is zero(252). Under the InSIDE assumption,

mentioned previously, MR-Egger regression is able to test for directional pleiotropy, test for a causal effect, and provide an estimation of the causal effect(249). If the InSIDE assumption holds, this method can provide a consistent causal effect estimate calculated by a genotype-outcome dose response relationship(229, 252). The slope estimate, if the assumption holds, then provides an unbiased estimate for the causal effect. Directional pleiotropy can also be detected as the intercept estimate can be interpreted as the average pleiotropic effect across all genetic instruments(249). Unlike the IVW method, the intercept is not constrained to zero(249). A non-zero intercept can either (or in addition to directional pleiotropy) demonstrate a violation of the InSIDE assumption, highlighting a potential bias to the IVW estimate(249).

The Rucker's Q' statistic is an extension of the Cochran's Q statistic used in the IVW method to assess heterogeneity. If the Q-Q' value is large, this is suggestive of heterogeneity and indicates that there may be directional horizontal pleiotropy between genetic variants(250, 253).

4.3.3 Weighted median

Median-based approaches are a beneficial analysis method as, unlike IVW method, they does not require all instruments to be valid(254). In an unweighted median-based analysis, the causal estimates for each genetic variant is calculated, and the median of the causal estimates is the estimation of the causal effect(249). A consistent causal estimate can be estimated using this approach if at least 50% of the causal estimates provided come from valid instruments(229). However the unweighted median-estimator can result in bias of the causal effect estimate when the precision of the causal estimates from the genetic instruments vary(254). The use of the weighted median approach can therefore be a more suitable method to estimate the causal effect. The method proposes that genetic instruments with more precise causal estimates will then contribute more weight to the MR analysis(249). This approach can then consistently estimate the causal effect if at least 50% of the weight come from valid genetic instruments(254).

4.3.4 Weighted mode

Mode-based approaches provide another sensitivity analysis in addition to MR-Egger regression and weighted median methods. This approach is based on the concept that genetic instruments of similar causal effects will be grouped together, with the group with the largest number of genetic instrument providing the causal effect estimate(255). This method can consistently estimate the causal effect if the Zero Modal Pleiotropy Assumption (ZEMPA) holds(256). This assumption holds if the most common causal estimate comes from valid instruments.

Mode-based approaches can be categorized into simple and weighted. Simple mode is unweighted and is the mode of the empirical density function of causal estimates(255). Weighted modes, however, are

weighted by the inverse variance of the SNP-outcome association(255). For a weighted mode-based approach, varying the weight given to each Wald ratio estimate and giving more weighting to valid instruments will satisfy the ZEMPA assumption and return a consistent causal effect estimate(256).

4.4 **MR-Base**

To take advantage of the increasing amount of publicly available published GWAS summary statistics, MR-Base(257) (http://www.mrbase.org/) was developed with the aim to 1) Curate complete summary statistics from published genome-wide association studies into a centralized database and 2) become an analytical platform that uses these GWAS summary statistics and, through automation, perform two-sample MR association tests and sensitivity analyses. Currently, the database consists of 1673 GWAS from which include 11 billion SNP-trait associations(258). These GWAS summary data comprise of various complex traits and diseases and allows for millions of potential causal associations to be evaluated.

Another application that MR-Base supports is the use of phenome-wide association studies (PheWAS) to identify if certain genetic instruments (i.e. SNPs) are a source of horizontal pleiotropy(258). MR-Base PheWAS allows users to input a reference SNP ID number (rsID) for a given genetic instrument, and it returns a list of traits (starting from the trait with the smallest P-value) identified from genome-wide association studies that show an association with the genetic instrument.

Researchers can benefit from the use of MR-Base through its automation of two-sample MR(258). This increases the efficiency and practicality of performing the analysis, as steps such as harmonization of GWAS summary data, LD clumping to retain independent instruments for the exposure of interest, and sensitivity analyses, such as MR-Egger regression and heterogeneity tests across genetic instruments, can be performed(250, 252, 259). Furthermore, as MR-Base comprises of both a curated, centralized database and an analytical platform, results that are generated from MR-Base can be easily reproduced.

However, findings from MR-Base can be limited by issues such as multiple testing and issues concerning the interpretation of complex MR results as a result of performing a hypothesis-free study(258). Considerations into presenting all results should be performed to avoid the potential for selecting results from analyses based on subjective p-value thresholds, and generating a well-defined analysis plan can mitigate against such biases(258).

4.5 **LD Hub**

To exploit the use of publicly available GWAS summary statistics and tackle the questions regarding the underlying genetic contribution of complex disease risk factors and outcomes, LD Hub (http://ldsc.broadinstitute.org/ldhub/)(260) was developed as a centralised database and web interface.

The LD Hub web interface consists of: 1) A 'Lookup Centre' that allows users to search for existing LD Score regression findings; 2) Access of the centralised GWAS summary statistics database; 3) Automation of cross-trait LD Score regression analysis utilising the GWAS summary statistics in the centralised database. Currently, this database consists of cleaned and harmonised summary data from 36 GWAS consortia, 173 traits with GWAS data, and approximately 1.5 million individuals of European ancestry(260).

The cross-trait LD Score regression analysis (i.e. LD Score correlation analysis) performed in LD Hub is an extension of single-trait LD Score regression which requires individual-level genotype data(261). In brief, the concept behind single-trait LD Score regression is that genetic variants that have high LD scores will have a greater probability of tagging causal variants and have elevated test statistics(262). Conversely, genetic variants with low LD scores will have a smaller probability of tagging causal variants and attenuated test statistics. Single-trait LD Score regression calculates the SNP-trait correlation as a function of the LD Score (i.e. sum of linkage disequilibrium (r^2) measured with all other SNPs)(262). This approach allows polygenicity to be distinguished from confounding (e.g. population stratification, cryptic relatedness). However, flexibility in this approach allows the use of GWAS summary statistics to be employed in cross-trait LD Score regression analysis to estimate the genetic correlation between different complex disease outcomes and traits of interest (Discussed in further detail in Chapter 4.5.1 LD Score correlation analysis). Automation of the cross-trait LD Score regression analysis pipeline allows users in LD Hub to upload their GWAS summary data, perform quality control checks on their GWAS summary data, perform SNP heritability analysis, and then implement genetic correlation analysis(260).

The benefits of using LD Hub is primarily its accessibility to all users, even individuals who do not a computational background(260). The centralised database minimises the time researchers spend cleaning and harmonising GWAS summary statistics, and the user-friendly web interface guides users into how to upload their GWAS summary data. However, a limitation of this software is that currently the centralised database only includes GWAS studies with individuals of European ancestry(260).

4.5.1 LD Score correlation analysis

The method of LD Score regression analysis can be applied to estimate genetic correlations between traits or disease outcomes(261). An understanding of genetic correlation is important as it assesses the shared genetic architecture of complex disease traits and outcomes. Studies have used the LD Score correlation approach to determine the genetic overlap between traits, with studies investigating the genetic correlation between autism spectrum disorders and neuropsychiatric variation(263), and years of education and age at first childbirth(264).

LD Score correlation analysis estimates the genetic correlation (r_g) by regressing the trait-trait correlation (i.e. product of two z-scores obtained from separate GWASs each investigating a specific trait) as a function of the LD Score(261). The slope of the LD Score correlation analysis estimates genetic covariance (i.e. per-SNP heritability), with the intercept of the LD Score correlation measuring the degree of sample overlap between the two GWASs.

5. *Helicobacter pylori* genome-wide association meta-analysis and genetic correlation analysis

5.1 **Objectives**

To identify genetic variants that are strongly associated with *Helicobacter pylori* infection using two independent prospective cohorts, and subsequently use the GWAS summary statistics to perform an exploratory genetic correlation analysis of phenotypes associated with cardiometabolic diseases and cancer.

5.2 **Methods**

5.2.1 Study samples

ALSPAC is a transgenerational prospective study which initially included over 13,000 women and their children recruited in 1990-1992 in the Bristol region of the United Kingdom(265, 266). The study consists of comprehensively measured genetic, phenotypic, epigenetic and metabolomic data on mothers, fathers and children, with ongoing follow-up data(267). In this analysis, we used children from the "Focus@7" Clinic whose blood samples were collected between September 1998 and October 2000. These participants were part of a 10% randomly selection subsample of the ALSPAC cohort (also known as the "Children in Focus"), born between June 1992 to December 1992, and invited to attend clinics(268). The sample included for analysis with *H.pylori* phenotype data consisted of 4651 individuals.

The Caerphilly Prospective Study is a population-based cohort study consisting of unrelated men recruited between 1979 and 1983 when the participants were aged 45-59 years old(269). The study defined the area of selection as men residing in the town of Caerphilly in Wales and five adjacent villages(270). At initial enrolment (Phase 1), the sample size included 2512 participants. Further follow-up data was collected from this cohort every 5 years, resulting in Phase 2 (1984-1988), Phase 3 (1989-1993) and Phase 4 (1993-1997). At each phase ischaemic heart disease outcomes, lifestyle factors, clinical factors and bloods were extensively measured, with additions at Phase 2 of psycho-social factors

and at Phase 3 of cognitive function measurements. The current analysis included 868 Caucasian men with *H.pylori* phenotype data.

5.2.2 Phenotype measurement

Whole blood samples from ALSPAC Focus@7 participants were processed by firstly centrifuging at 3500rpm at 4-5ºC for 10 minutes(268). These samples were then aliquoted and temporarily stored at -20ºC before being stored at -70/80ºC for long-term storage(268). When ready for analysis, EDTA plasma samples were plated out into 96 well plates. *Helicobacter pylori* whole lysate antigen (Meridian Life Sciences, catalogue number: RS2101) was used as an IgG antibody titer and measured using ELISA methods to measure *H.pylori* IgG antibodies in the plasma samples(268, 271). In brief, the assays were implemented by reacting the microtiter plates coated with the *H.pylori* antigen with a sequence of diluted aliquots of human plasma, enzyme-labelled anti-human IgG and enzyme substrate, with each reaction followed by a plate wash. Subsequently, the enzyme-substrate reaction was quantified using a microplate colourimeter to measure optical density. *H.pylori* measurements were recorded in three ways: the measurement of optical density directly read from the ELISA plate; the ratio to standards obtained from the standards measured on each ELISA plate; the standardised z-score from each ratio to standard measure. In the analysis, the last method was selected and these values were calculated by subtracting the ratio to standard and the mean ratio to standard and then dividing by the standard deviation per plate and adding 2(268).

Similarly, frozen plasma samples stored at -20ºC were collected from CaPS participants and measured by commercial ELISA (Helicobacter pylori HM-CAP, Sigma Diagnostics, St Louis, Missouri, USA)(108, 272).

5.2.3 Genotyping and imputation

ALSPAC Focus@7 participants were genotyped using the Illumina HumanHap 550 quad chip genome-wide SNP genotyping platform (Illumina, Inc., San Diego, CA) by 23andMe subcontracting from the Wellcome Trust Sanger Institute, Cambridge, UK and the Laboratory Corporation of America, Burlington, NC, US, and called with Illumina GenomeStudio. PLINK(273)(v1.07) was used to carry out quality control measures on an initial set of 9,912 children, which includes Focus@7 individuals, and 609,203 directly genotyped SNPs. SNPs were removed if they exhibited more than 5% missingness or a Hardy-Weinberg equilibrium P values of less than $5x10^{-7}$. Furthermore, SNPs were removed if they had a minor allele frequency less than 3% and a call rate of <95%. Samples were excluded if they had more than 5% missingness, extreme autosomal heterozygosity, and undetermined X chromosome heterozygosity. Additionally, to determine if the sample displayed evidence of population stratification, multidimensional scaling of genome-wide identity-by-state pairwise distances that clustered outside the

CEU HapMap 2 population were excluded. Additionally, cryptic relatedness was identified using an identity-by-decent estimate 0.1, which approximately corresponds to 10% or more alleles shared identical by decent or relatedness at the level of first cousins.

After quality control, a total of 8,365 unrelated individuals and 500,527 SNPs passed these filters. Autosomal SNPs were imputed against the HapMap(274) CEU population (Residents of Utah with Northern and Western ancestry; release 22) using MACH(275) (v1.0.16). Out of the total 8,365 individuals with genotype data, 4,651 individuals from Focus@7 had *H.pylori* phenotype data, and these individuals were used in the GWAS.

CaPs genotype data was acquired using the Illumina CardioMetabochip(276) which includes approximately 200,000 SNPs from loci that had been identified in previous GWASs investigating various cardiometabolic disease risk factors and outcomes (e.g. blood pressure, lipid levels, type 2 diabetes and myocardial infarction) to have promising associations with the disease risk factor or outcome. Rare variants were also imputed using the 1000 Genomes Project(277) as a template, and genotype data was called with Illumina GenomeStudio(v2010.3). Quality control measures were carried out on an initial set of 1411 individuals and 196,725 directly genotyped SNPs. These measures included removing samples if they had gender ambiguity, a call rate of <95%, sample mix-up, replicate concordance, discordance between reported and genetically-determined ethnicity and cryptic relatedness(278). Furthermore, SNPs were excluded if they had a call rate of <95%.

A total of 1349 individuals passed the quality control measures, and the genome covered by the Metabochip was augmented through imputation using the 1000 Genomes European ancestry reference panel(277). Analyses was limited to 1 million SNPs with imputed $R^2 \geq 0.8$ covering a dense coverage of loci related to cardiometabolic disease, and after imputation 1,309,437 SNPs remained. Out of the total 1349 individuals with genotype data, 868 individuals had *H.pylori* phenotype data, and these individuals were used for analysis.

5.2.4 Association testing

In the ALSPAC cohort, *H.pylori* measures were transformed to SD units. PLINK(273)(v.1.09) was used to carry out the GWAS, with age, sex and the first 10 principal components, to adjust for confounding by population stratification, included in the linear regression model.

In the CaPS cohort, PLINK(v1.09) was used to carry out the GWAS, with age, batch number and the first 10 PCs included in the logistic regression model. LiftOver (http://genome.ucsc.edu/cgi-bin/hgLiftOver) was used to standardise genomic coordinates to be reported on the NCBI build 37 (hg 19), with alleles on

the forward strand. The Haplotype Reference Consortium 1KG reference panel was then used to convert chromosomal coordinates to rsIDs.

In addition, the ALSPAC GWAS included an exclusion list which consisted of a list of individuals who must be excluded from the analysis due to reasons such as withdrawal of consent and loss to follow-up.

5.2.5 Meta-analysis

METAL(279) was employed for meta-analysis to combine the results from the ALSPAC and CaPS GWAS. This analysis used fixed effects model, and only allowed for variants with a high imputation quality score (Rsq < 0.8). The threshold to define genome-wide significant associations was $P < 5 \times 10^{-8}$. Furthermore, as ALSPAC *H.pylori* phenotype data consisted of continuous variables and CaPS *H.pylori* data consisted of binary variables, associations were interpreted based on the P-value.

5.2.6 LD Score correlation analysis

LD Hub(260)(http://ldsc.broadinstitute.org/), a web interface and centralised database for GWAS summary statistics of individuals with European ancestry, was used to perform LD Score correlation analysis. Disease outcomes and risk factors used in the genetic correlation analysis consisted of cardiometabolic outcomes, cancer types and anthropometric traits. This approach used the summary results from my *H.pylori* GWAS meta-analysis previously performed. Quality control measures were then implemented. To standardise the GWAS summary statistics, QC checks such as removal of SNPs with minor allele frequency greater than 1%, removal of strand-ambiguous SNPs, removal of SNPs that show mismatch to SNPs in the 1000 Genomes dataset, and removal SNPs with large effect sizes, were implemented(260). Genetic correlation analysis was subsequently performed. This analysis measures the proportion of genetic overlap between trait A (i.e. *H.pylori*) and trait B (e.g. LDL cholesterol, coronary heart disease).

5.3 **Results**

5.3.1 Genome-wide association meta-analysis

After applying quality control measures, a total of 8,172,101 SNPs were tested for association with *H.pylori* based on phenotype data from 4513 individuals (ALSPAC = 3645; CaPS = 868). ALSPAC phenotype data consisted of continuous values of standardised z-scores from each ratio to standard measure of *H.pylori* IgG antibodies, whilst CaPS phenotype data (580 cases and 288 controls) consisted of binary values. Therefore, in order interpret GWAS meta-analysis findings when combining the two studies, results were interpreted based on P-values.

No SNPs reached genome-wide significance ($P < 5 \times 10^{-8}$), however the meta-analysis did yield four highly suggestive SNPs ($P < 10^{-7}$)(Table 2) (Figure 8) (Figure 9). The SNP with the smallest P-value, rs366337 (P-value = 1.27×10^{-7}) is located on chromosome 19 in the *LILRB2* locus encoding the leukocyte immunoglobulin-like receptor B2. The second leading SNP, rs2177192 (β = 0.475; 95%CI = 0.296 to 0.653; P-value = 1.86×10^{-7}) is located on chromosome 1 in the intron of *ST6GALNAC5*, a gene encoding the protein ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 5. The other suggestive SNPs comprised of rs17502937 (β = -0.443; 95%CI = -0.614 to -0.271; P-value = 4.31×10^{-7}) located in chromosome 13 but not linked to any specific gene, and rs74884614 (β = -0.344; 95%CI = -0.478 to -0.209; P-value = 5.84×10^{-7}) is located on chromosome 2 in the intron of *WIPF1*, which encodes for WAS/WASL interacting protein family member 1.

Table 2. Suggestive SNPs identified in *Helicobacter pylori* GWAS meta-analysis. The meta-analysis used both continuous and binary *H.pylori* outcomes, and therefore interpretation of findings were compared using only the P-values

SNP	Gene	EA	OA	β (95% CI)	se	Weight (N)	p
rs366337	*LILRB2*	A	G	NA	NA	868	1.266×10^{-7}
rs2177192	*ST6GALNAC5*	C	G	0.475 (0.296 to 0.653)	0.091	3645	1.861×10^{-7}
rs17502937	-	T	G	-0.443 (-0.614 to -0.271)	0.088	3645	4.313×10^{-7}
rs74884614	*WIPF1*	C	G	-0.344 (-0.478 to -0.209)	0.069	3645	5.839×10^{-7}

Figure 8. *H.pylori* GWAS meta-analysis results illustrated in a Manhattan plot. This plot is showing the association of all SNPs with *H.pylori* infection. The x-axis illustrates all the available SNPs in the two cohorts used for analysis according to their chromosomal position, against the y-axis (i.e. $-\log_{10}(p)$) demonstrating the association with *H.pylori* infection. The green dots highlight the four highly suggestive SNPs. The solid red line indicates the threshold for genome-wide significance (P-value = 5×10^{-8}).

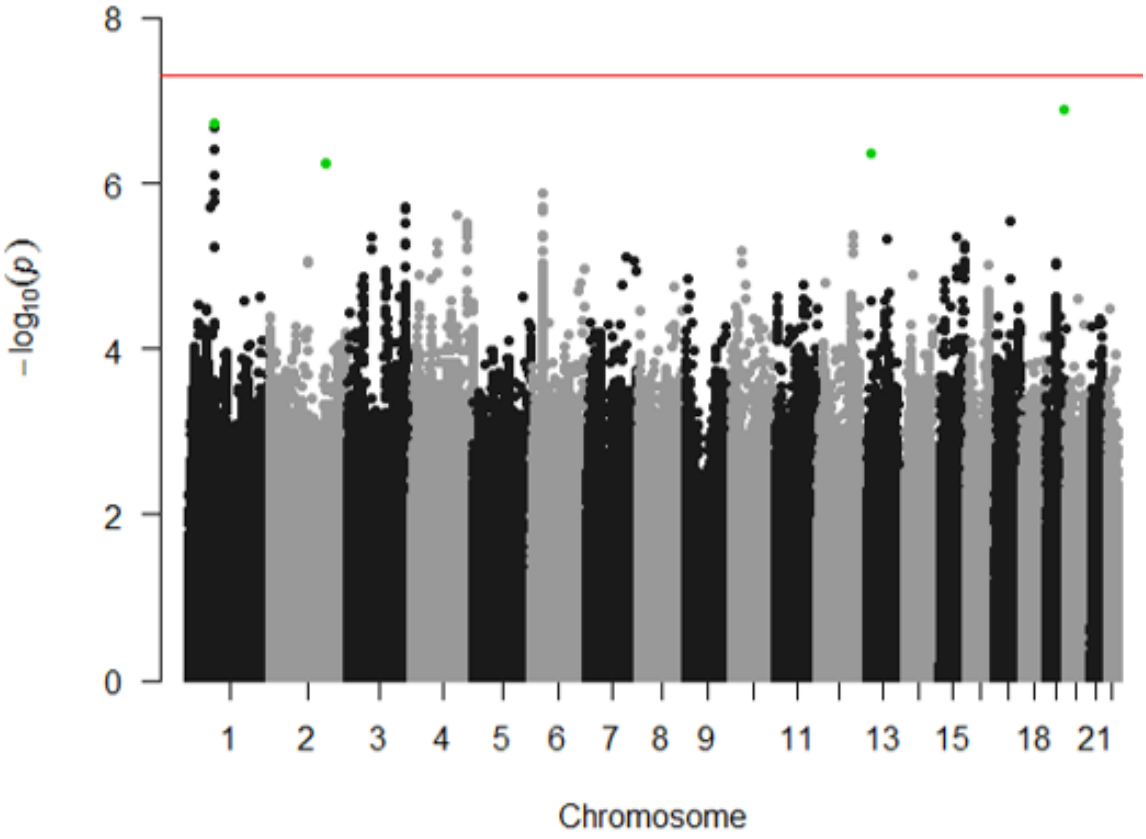
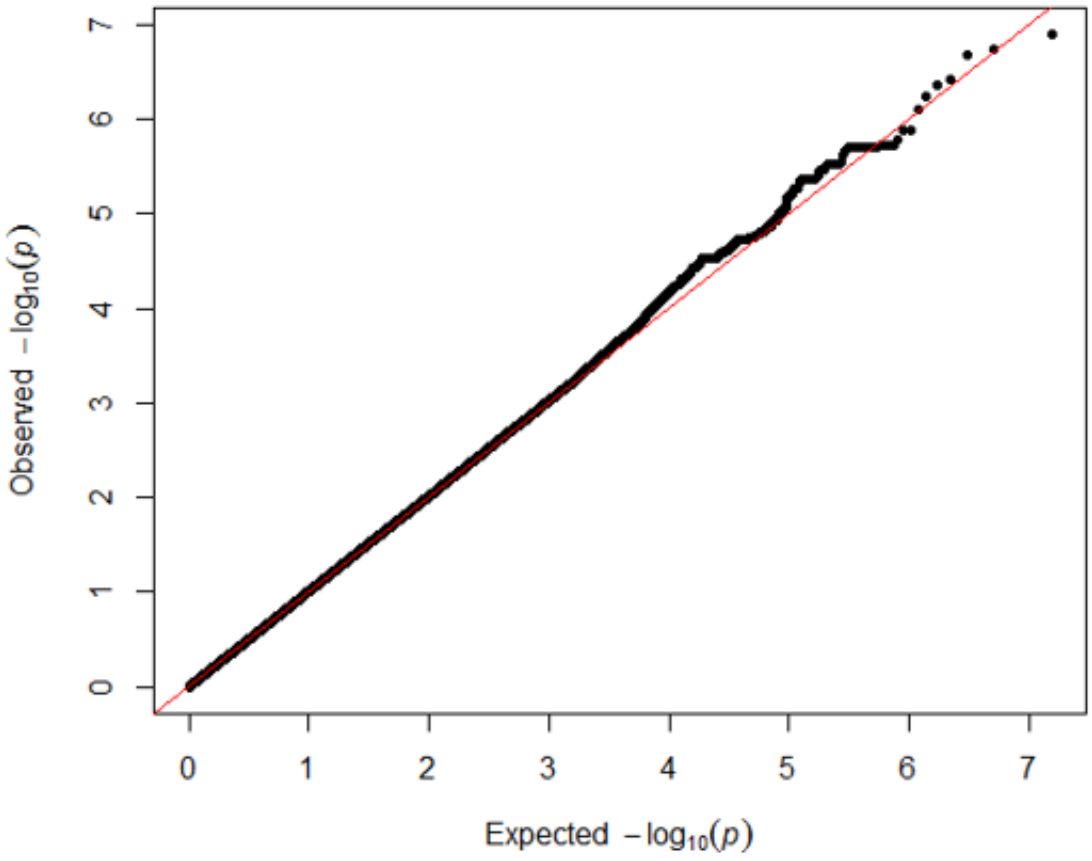


Figure 9. Q-Q plot of *H.pylori* GWAS meta-analysis results. The Q-Q plot shows the distribution of expected P-values (x-axis) compared to the distribution of the observed P-values (y-axis). The plot suggests limited evidence of genetic variants strongly associated with *H.pylori* infection.



5.3.2 LD Score correlation analysis

LD Score regression used 38 anthropometric, cardiometabolic and cancer traits with a Bonferroni correction threshold of P < 0.0013. Results shown in Appendix Table 1 illustrated low SNP heritability Z scores (i.e. Z score < 4) and therefore no strong signals as a result of the limited sample size from the GWAS meta-analysis summary statistics provided (N = 3645).

6. Mendelian randomization

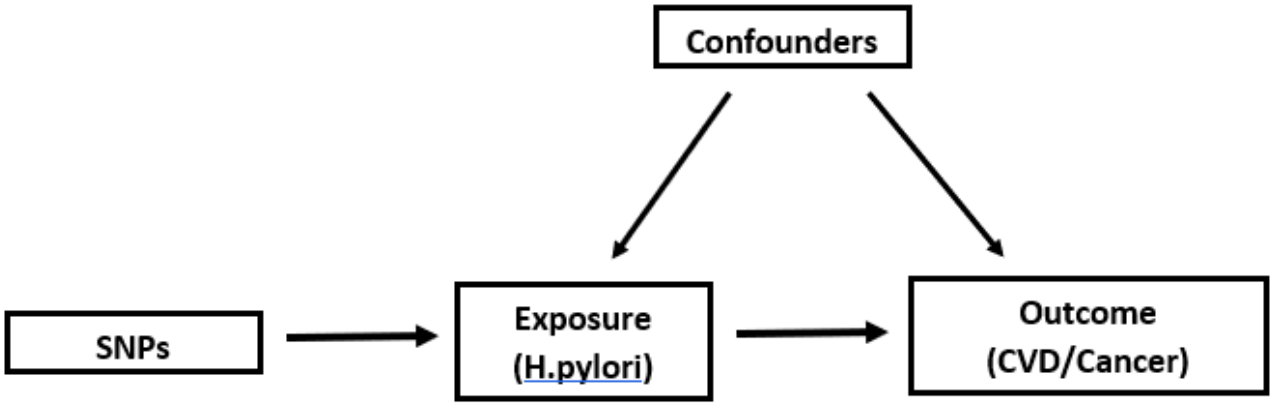
The recent development in genome-wide association studies have extended the scope of Mendelian randomization by providing an opportunity to utilize the identified genetic variants associated with bacterial infection such as *H.pylori(*42*)*, and gut microbiota of various genera such as *Acidaminococcus*, *Escherichia*, *Lactobacillus*, *Bifidobacterium* and *Oscillibacter*(280, 281).

Prior to the first published *H.pylori* GWAS meta-analysis in 2013(42), there were no genetic studies investigating the genetic basis of *H.pylori* seroprevalence. This study identified two genetic variants

strong associated with *H.pylori*, rs10004195 (OR: 0.70; 95% CI: 0.65 - 0.76; P: 1.42 x 10^{-18}) located on the *TLR* locus, and rs368433 (OR: 0.73; 95% CI: 0.65 - 0.81) located in an intron of *FCGR2A*. Both these variants suggest a plausible biological pathway, as discussed in Chapter 2.3 Genetic studies.

To my knowledge, no MR study has been performed to examine the causal association of *H.pylori* with cardiovascular disease and cancer. Therefore, in the absence of studies definitively outlining the causal effect of *H.pylori* on the disease outcomes of interest, Mendelian randomization is a novel and timely approach to traditional observational studies. This study will perform two-sample MR using published genetic variants as proxies for *H.pylori* infection to examine the association of *H.pylori* infection with cardiovascular disease outcomes and cancer types (Figure 10).

Figure 10. DAG of causal associations tested using the two-sample MR paradigm: 1) Causal association between exposure, *H.pylori*, and outcome, cardiovascular disease (CVD) traits, estimated using *H.pylori* single nucleotide polymorphisms (SNPs) as proxies; 2) Causal association between exposure, *H.pylori*, and outcome, cancer types, estimated using *H.pylori* SNPs as proxies for the exposure of interest.



6.1 **Objectives**

To examine the causal association of *Helicobacter pylori* with cardiovascular disease traits and cancer, respectively.

6.2 **Methods**

6.2.1 Instrument selection

To generate a genetic instrument for H.pylori seroprevalence, we used the results from a published H.pylori GWAS meta-analysis(24). To categorise individuals as *H.pylori* seropositive (case) or seronegative (control), seroprevalence was classified as an anti-*H.pylori* IgG titer equal to or greater than 20 U/mL(282). Individual participant data from the Study of Health in Pomerania (SHIP) (n = 3830), and

two cohorts from the Rotterdam Study (RS), RS-I (n = 4542) and RS-II (n = 2566), were included in the GWAS meta-analysis(42).

The SHIP study comprises of two independent population-based cohorts, SHIP and SHIP-TREND, in West Pomerania, north-east of Germany(283). The study design of these two cohorts consists of a two-stage cluster sample of individuals aged 20 to 79 years(283). Participants in the SHIP cohort were recruited from October 1997 to May 2001. Overall, this cohort included 6265 eligible participants(283). Recruitment and baseline measurements for the SHIP-TREND cohort were performed between 2008 and 2012 and consisted of 4420 participants(284).

The Rotterdam Study is a prospective population-based cohort based in Rotterdam, the Netherlands(285). This study consists of 3 cohorts (RS-I, RS-II, RS-III), with eligible participants aged 55 years and over(286). Participant recruitment and baseline data for RS-I (n = 7983) were acquired from 1990-1993, RS-II (n = 3011) obtained between 2000-2001, and RS-III (n = 3932) collected from 2006-2008(286).

Using a fixed-effects meta-analysis model, they identified *rs10004195* ($p = 1.42 \times 10^{-18}$) located on *TLR* and *rs368433* ($p = 2.1 \times 10^{-8}$) located in the intron of *FCGR2A*, showing evidence of genome-wide significance ($p = 5 \times 10^{-8}$)(42). The location of these loci in the *TLR* and *FCGR2A* gene complement the context of *H.pylori* as an infection which triggers an immune response, as *TLR* plays a role in the innate immune system(287) and *FCGR2A* is found on the surface of many immune response cells(288)(Function of genes discussed in detail in 6.2 Genetic studies). Table 3 demonstrates the mentioned genetic variants used as proxies for *H.pylori* as the exposure of interest in two-sample Mendelian randomization analysis.

Linkage disequilibrium clumping was then employed using European samples from the 1000 Genomes Project to remove highly-correlated genetic variants (289). The genetic variants were pruned and clumped using R version 3.5.1 and measured using r^2, a measure of linkage disequilibrium, at a threshold of 0.001 to ensure that the *H.pylori* genetic variants used were independent from one another.

Table 3. *H.pylori* genetic instruments for two-sample MR analysis.

Phenotype	SNP	β	se	eaf	EA	OA	p	Gene
H.pylori	rs10004195	-0.357	0.05	0.247	A	T	1.42×10^{-18}	TLR
H.pylori	rs368433	-0.315	0.08	0.16	G	A	2.1×10^{-8}	FCGR2A

6.2.2 Outcome selection

MR-Base (http://www.mrbase.org/), a centralised database of harmonised summary data from 1094 genome-wide association studies examining various diseases and complex traits, was used to generate a list of traits related to cardiovascular disease and cancers(257). In total, 87 traits were selected and used as outcome data for the two-sample MR analysis using traits from the MR-Base centralised database that were identified in the subcategories 'Anthropometric', 'Cardiovascular', and 'Cancer' (Appendix Table 2). The majority of populations that the GWAS were derived from were homogenous, consisting of Caucasian European ancestry. However, for some traits GWAS data from populations of other ancestry, such as South-East Asian, were included. Principal components were used to mitigate the effect on the causal estimates.

6.2.3 Two-sample Mendelian randomization method

Two-sample MR analysis was performed to estimate the causal effect of *H.pylori* on cardiovascular disease traits and cancers using summary statistics from the Mayerle et al. 2013(42) genome-wide association meta-analysis. R version 3.5.1 was utilised to access the MR-Base centralised database and perform two-sample MR.

In order to perform two-sample MR, specific information from the summary statistics from the *H.pylori* GWAS(42) needed to be extracted. The data frame of the *H.pylori* genetic instruments included: rsID of single-nucleotide polymorphisms (SNPs), beta values, standard error values, effect allele, other allele, effect allele frequency, phenotype, p values and gene name.

Harmonisation of datasets from instrument-exposure and instrument-outcome was then performed. This stage is essential and must be completed correctly to ensure that the causal effect is not biased due to data harmonisation error. The first step in harmonisation was to standardise the direction in the exposure dataset. This included ensuring that the exposure was coded in the same forward (5'-3') positive strand direction, meaning that the exposure-increasing allele is noted as the effect allele(259). If this was not the case, the genetic variant would then need to be 'flipped' (i.e. effect allele, effect estimates and effect allele frequencies) in order to conform to the same direction as the other genetic variants(259). The outcome data (i.e. data of included traits from MR Base) was also required to be coded from the same strand as the exposure dataset. This then ensured that alleles in the outcome dataset matched the alleles in the exposure dataset. Without the proper data harmonisation, effect allele mismatches can result in the causal effect estimate becoming distorted, and interpretation of findings can become misleading(259).

Two-sample MR was then performed to calculate the estimates of the causal effect of the exposure (*H.pylori*) on the cardiovascular disease outcomes and cancer outcomes using Wald ratio and inverse

variance weighted method. Other MR methods such as MR-Egger regression, weighted mode and weighted median could not be used as sensitivity analyses as the genetic instrument consisted of only two genetic variants (*rs10004195* and *rs368433*). More genetic variants are required to perform these MR methods. Wald ratio estimates were calculated for each genetic instrument to estimate the causal effect of *H.pylori* on the outcome variables. IVW method was used when ratio estimates from multiple instrumental variables could be meta-analysed to estimate the causal effect. Sensitivity analyses to test for heterogeneity using Cochran's Q statistic and single SNP analysis were also performed.

PhenoSpD(290) was then implemented as an appropriate method to correct for multiple testing and avoid overcorrecting due to correlated datasets of cardiovascular disease traits and cancer. As the phenotypic correlation matrix from LD Hub(260) was provided, the steps of GWAS data harmonisation and phenotypic correlation estimation was already previously performed to obtain the phenotypic correlation matrix. In brief, GWAS summary statistics of cardiovascular disease outcomes and cancer types were used from MR-Base(257). For each disease traits, included GWASs were selected by largest sample size or largest number of cases for binary traits. All included summary statistics were then harmonised, with all datasets containing 'SNP', 'Effect allele', 'Other allele', 'Z-score', 'P-value', and 'Sample size (N)'. Phenotypic correlation estimation is then calculated in LD Hub(260) by performing bivariate LD score regression, adjusting for sample overlap(290). The regression intercepts, which are the phenotypic correlation between two traits (e.g. phenotypic correlation between LDL cholesterol and coronary heart disease), then form the phenotypic correlation matrix. Spectral decomposition (SpD) approach(291, 292) is then applied to estimate of the number of independent variables among the selected outcome traits using the phenotypic correlation matrix, estimate the number of independent tests, and calculate a more appropriate significance threshold than using Bonferroni correction(290).

An illustrative plot was then generated using a forest plot to illustrate the two-sample MR results. Two-sample MR analysis was performed using the R package of MR-Base (257)(https://github.com/MRCIEU/TwoSampleMR).

Finally, MR-Base PheWAS (http://phewas.mrbase.org/) was employed to determine potential pleiotropic pathways associated with the selected SNPs. This was used to examine if the SNPs were linked to cardiovascular and cancer traits that could violate the "no pleiotropy" assumption.

6.2.4 Sensitivity analyses

6.2.4.1 *Bidirectional Mendelian randomization*

To explore the traits (i.e. heart rate, hip circumference, LDL-cholesterol, and breast cancer) that reached the experiment-wide significance threshold ($P < 0.0007547$) in the preliminary MR analysis, bidirectional MR was employed to examine the direction of causality.

Genetic instruments to act as proxies for these selected exposures of interest were extracted from MR Base(257)(http://www.mrbase.org/) (Appendix Table 3). For heart rate, instruments from the Heart Rate consortium (HRgene) were extracted. 15 robust SNPs (P value $> 5 \times 10^{-8}$) were obtained from a sample size of 92,355 individuals of predominantly European ancestry (European = 92%, Other = 8%). The UK Biobank(293) cohort was used to obtain genetic instruments for hip circumference and breast cancer. In total, this large, population-based prospective study consists of 500,000 participants aged 40-69 years, with extensive genetic and phenotypic data collection between 2006 to 2010(294). 285 robust SNPs associated with hip circumference were included in the analysis, with these genetic instruments acquired from an all European cohort with a sample size of 336,601 individuals. Similarly, an all European sample of 308,780 participants was used to obtain the 11 robust SNPs associated with breast cancer. The GLGC was employed to obtain the 80 robust SNPs associated with LDL-cholesterol used in the analysis. The sample size used to obtain these genetic variants consisted of 173,082 individuals of mixed ancestry, with this consortium made up of cohorts from Europe, East Asia, South Asia and Africa.

Outcome data was obtained from the *H.pylori* GWAS meta-analysis, using the ALSPAC(266) and CaPS(108) cohorts, performed prior to this bidirectional MR(108).

Similar to two-sample MR, LD clumping and harmonisation of the instrument-exposure and instrument-outcome datasets, and MR analysis were performed as described before.

Bidirectional MR was applied to test the hypothesis that the traits of interest (i.e. exposure of interest) would not be in the direction of *H.pylori*, and therefore would not show strong evidence of reverse causation. This MR analysis was able to benefit from all the genetic instruments for each trait having more than two SNPs which therefore allowed the use of MR-Egger regression, weighted median approach, and weighted mode approach to be used for sensitivity analyses. As previously discussed in Chapter 4.2 Mendelian randomization methods to estimate the causal effect, these methods are useful in determining if the IVW causal effect estimate is biased and if the genetic instruments violate instrumental variable assumptions due to limitations such as horizontal pleiotropy.

6.2.4.2 *Test for heterogeneity*

Cochran's Q statistic (IVW) and Rucker's Q' statistic (MR-Egger) (previously discussed in Chapter 4.2 Mendelian randomization methods to estimate the causal effect) were utilised to measure heterogeneity between genetic instruments as sensitivity analyses to detect pleiotropic effects.

6.2.4.3 *Test for directional pleiotropy*

The MR-Egger intercept was used to test for directional pleiotropy by measuring how far the intercept value deviated from zero (i.e. no pleiotropic effect or balanced horizontal pleiotropy). Across all genetic instruments, this value captures the average pleiotropic effect.

6.2.4.4 *Single-SNP analysis*

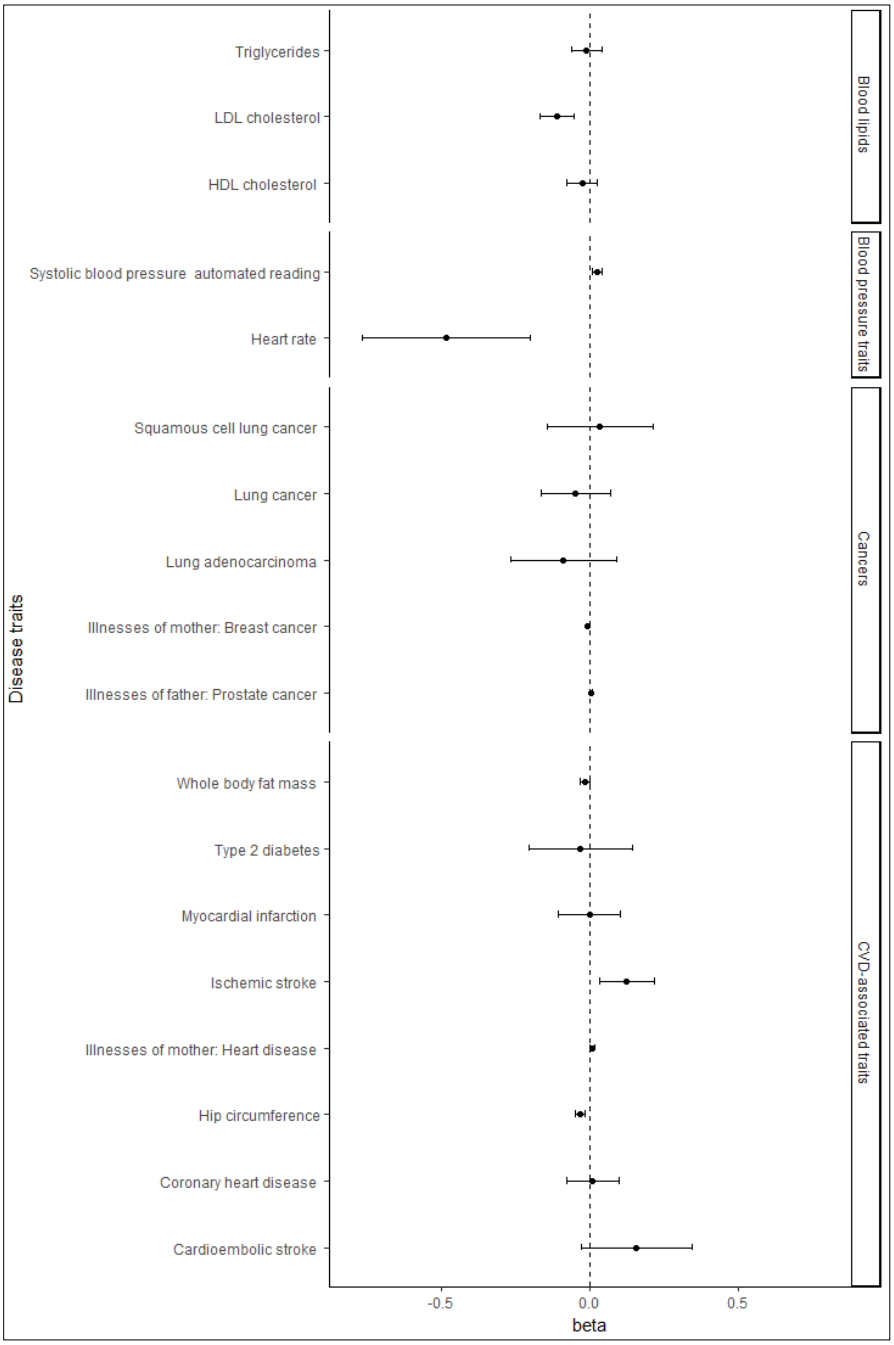
The single-SNP analysis used the Wald method(295) to estimate the causal effect of the exposure (i.e. *H.pylori*) on the disease outcome (i.e. CVD traits or cancer types) from each single genetic instrument.

6.3 **Results**

6.3.1 Two-sample Mendelian randomization analysis

A total of 87 outcomes of interest were analysed to determine which traits were causally associated with *H.pylori* infection. As some of the traits were observed to be corelated, the PhenoSpD approach was employed. PhenoSpD determined that 78 traits out of the 87 were independent of each other, with an experiment-wide significance threshold of P > 0.0007. Using this threshold, four traits were identified to be causally associated with *H.pylori*: LDL cholesterol (β = -0.113; 95% CI = -0.170 to -0.056; P-value = 0.0001), Hip circumference (β = -0.032; 95% CI = -0.049 to -0.016; P-value = 0.0001), Breast cancer (β = -0.008; 95% CI = -0.013 to -0.004), and Heart rate (β = -0.485; 95% CI = -0.767 to -0.203; P-value = 0.0007) (Figure 11) (Appendix Table 4).

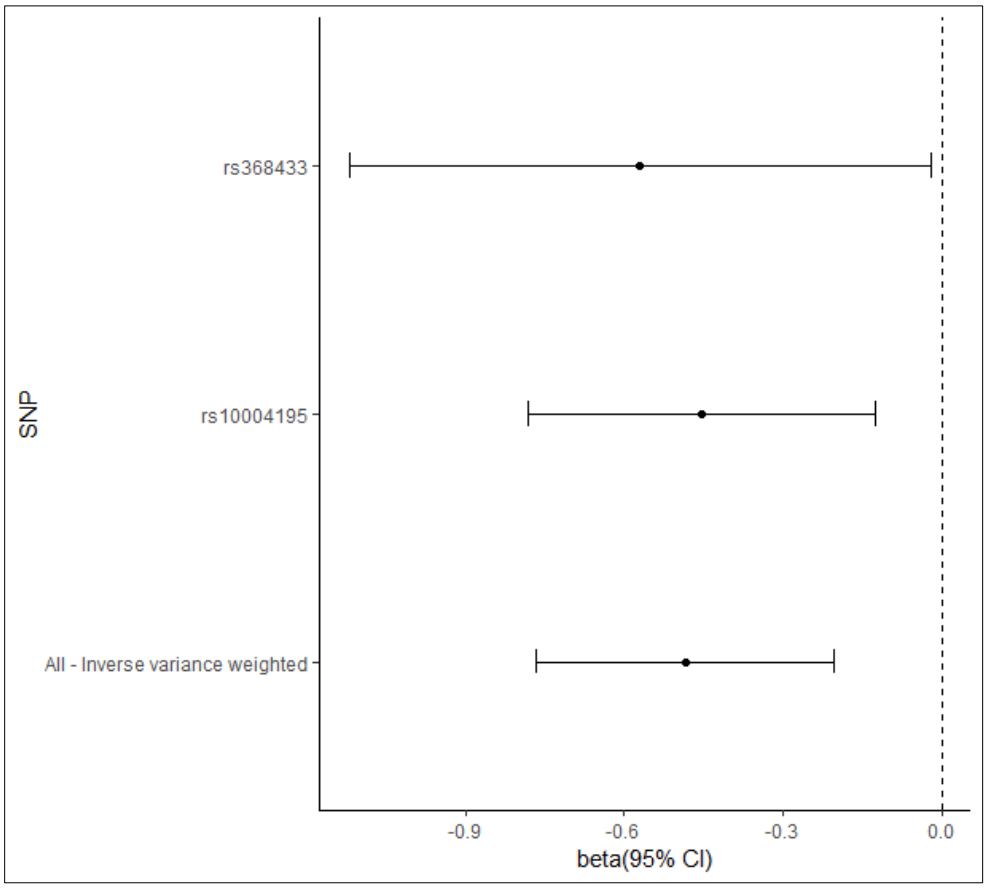
Figure 11. Forest plot illustrating findings from two-sample MR analysis. The plot shows the causal association of *H.pylori* infection with different cardiovascular disease outcomes and cancer types by plotting causal effect estimates and 95% confidence intervals.



Additionally, MR single SNP analysis demonstrated that for heart rate, rs10004195 (β = -0.455; 95% CI = -0.783 to -0.127; P-value = 0.007) and rs368433 (β = -0.570; 95% CI = -1.121 to -0.020; P-value = 0.042) showed some evidence of causality separately, however the IVW causal effect estimate illustrated that together these SNPs showed stronger evidence of a causal association with *H.pylori* in the same direction

(β = -0.485; 95% CI = -0.767 to -0.203; P-value = 0.001)(Figure 12). Sensitivity analysis for heart rate using Cochran's Q test showed limited evidence of heterogeneity between genetic instruments and limited evidence of pleiotropic effects (Q = 0.125; Q P-value = 0.723).

Figure 12. Forest plot of two-sample MR single-SNP analysis: Heart rate results. This analysis illustrates the causal effect estimate of *H.pylori* infection on heart rate for each genetic instruments, and then a combined IVW causal effect estimate



6.3.2 Bidirectional Mendelian randomization analysis

Using the traits that reached the experiment-wide significance threshold in the two-sample MR analysis (i.e. LDL-cholesterol, breast cancer, hip circumference, and heart rate), bidirectional MR was performed to examine the direction of causality.

The odds ratio for one standard deviation unit change of *H.pylori* per one standard deviation (SD) unit change of LDL-cholesterol was 1.036 (95% CI = -0.064 to 0.135) (Appendix Table 5). Similarly, MR-Egger, weighted median and weighted mode methods concurred with this result. Additionally, sensitivity analysis showed little evidence of directional pleiotropy with an MR-Egger intercept estimate of -0.003

(95% CI = -0.012 to 0.006; P-value = 0.569) and the Cochran's Q (Q = 66.336; P-value of Q = 0.696) and Rucker's Q (Q = 66.664; P-value of Q = 0.715) statistics showing little evidence of heterogeneity between genetic instruments. Furthermore, as shown in Appendix Table 6, the single SNP analysis displayed little evidence of association for each SNP separately.

Analysis of hip circumference demonstrated an OR of 1.054 (95% CI = -0.106 to 0.212) for one SD unit change of *H.pylori* per one SD unit change of hip circumference (Appendix Table 5). This is in agreeance with the MR-Egger, weighted median and weighted mode approach (Table 6). The single SNP analysis displayed in Appendix Table 7 showing little evidence of association for each SNP separately. The MR-Egger intercept estimate yielded a value of 0.011 (95% CI: 0.002 to 0.021; P-value = 0.023) suggesting no strong evidence of pleiotropic effects across the genetic instruments. Cochran's Q (Q = 209.486; P-value of Q = 0.748) and Rucker's Q (Q = 204.226; P-value of Q = 0.811) statistics also showed limited evidence of heterogeneity.

Under the assumptions of the IVW method, the OR for one SD unit of change of *H.pylori* per SD unit of change for heart rate was 1.008 (95% CI = -0.020 to 0.037) (Appendix Table 5). This coincides with OR using MR-Egger, weighted median, and weighted mode (Appendix Table 5). Similar to LDL-cholesterol and hip circumference findings, the MR-Egger intercept estimate suggested little evidence of directional pleiotropy (MR-Egger intercept = -0.026; 95% CI = -0.081 to 0.028; P-value = 0.364) or heterogeneity (Cochran's Q: Q = 14.272; P-value of Q = 0.430; Rucker's Q: Q = 13.363; P-value of Q = 0.420). Additionally, in Table 4 single SNP analysis displayed little evidence of association for each SNP separately.

Table 4. Bidirectional MR: Heart rate single-SNP analysis results. The analysis estimated the causal effect of heart rate (exposure of interest) on *H.pylori* infection (outcome of interest) for each genetic instrument

SNP	β	se	p	Lower 95% CI	Upper 95% CI
rs1015451	0.041	0.053	0.437	-0.062	0.144
rs11153730	0.078	0.055	0.152	-0.029	0.185
rs11578508	-0.016	0.057	0.781	-0.128	0.096
rs13030174	0.120	0.078	0.125	-0.033	0.274
rs13245899	0.011	0.052	0.828	-0.091	0.113
rs17287293	-0.018	0.066	0.781	-0.148	0.111
rs17362588	0.062	0.054	0.247	-0.043	0.167
rs174549	0.035	0.058	0.553	-0.080	0.149

rs17796783	-0.030	0.076	0.688	-0.178	0.118
rs2029213	-0.051	0.078	0.514	-0.204	0.102
rs365990	0.011	0.038	0.767	-0.063	0.086
rs3729992	-0.026	0.085	0.763	-0.192	0.141
rs4489968	0.026	0.061	0.673	-0.093	0.145
rs6127471	0.009	0.043	0.838	-0.075	0.093
rs7980799	-0.136	0.053	0.009	-0.239	-0.033

The variable of interest, maternal family history of breast cancer yielded an OR of 0.470 (95% CI = -3.526 to 2.017) for *H.pylori* per one SD higher genetically raised natural log (ln) of breast cancer (Appendix Table 5). Weighted median and weighted mode displayed similar ORs with each other of 0.274 (95% CI = -4.905 to 2.317) and 0.237 (95% CI = -5.770 to 2.891), respectively. Dissimilarly, MR-Egger calculated an OR of 29.93(95% CI = -4.059 to 10.857). Sensitivity analyses illustrated no strong evidence of directional pleiotropy (MR-Egger intercept = -0.029; 95% CI = -0.077 to 0.019; P-value = 0.278) and heterogeneity (Cochran's Q: Q = 7.281; P-value of Q = 0.507; Rucker's Q: Q = 5.898; P-value of Q = 0.552) across and between genetic instruments. Additionally, single SNP analysis did not yield strong evidence of a causal association for any single genetic instrument (Table 5).

Table 5. Bidirectional MR: Breast cancer single-SNP analysis results. The analysis estimated the causal effect of breast cancer (exposure of interest) on *H.pylori* infection (outcome of interest) for each genetic instrument

SNP	β	se	p	Lower 95% CI	Upper 95% CI
rs1078806	0.287	2.842	0.920	-5.283	5.857
rs11836367	-6.383	5.123	0.213	-16.424	3.658
rs1269867	-1.346	5.386	0.803	-11.902	9.210
rs4442975	-3.989	5.413	0.461	-14.598	6.620
rs4784227	-2.528	3.192	0.428	-8.785	3.729
rs58952190	12.253	6.402	0.056	-0.294	24.800
rs78540526	-3.325	4.085	0.416	-11.332	4.681
rs7976725	0.377	5.162	0.942	-9.741	10.495
rs9397437	2.976	4.495	0.508	-5.834	11.785

7. Discussion

7.1 **Objectives and major findings**

Helicobacter pylori infection has been intensely investigated for its association with various gastrointestinal diseases(296-298), with observational studies also suggesting a link to cardiovascular disease and cancer. However, discordance in observational studies, limited by issues such as confounding, have led to no strong body of evidence definitively supporting a causal association of this bacterium with cardiovascular disease and cancer. I used GWAS to identify genetic variants associated with *H.pylori* seroprevalence using the ALSPAC and CaPs cohorts, LD score correlation analysis to identify genetic correlations between *H.pylori* infection and cardiovascular disease traits and cancer, and used published *H.pylori* genetic variants as proxies for Mendelian randomization to examine the causal association of *H.pylori* infection with cardiovascular disease and cancer.

Results from the GWAS meta-analysis identified four highly suggestive SNPs showing some evidence of an association with *H.pylori* infection. The SNP with the lowest P-value is located on *LILRB2*, and the second leading SNP is located on *ST6GALNAC5*. In addition, LD Score correlation analysis was employed but did not yield any reliable estimates of genetic correlation between *H.pylori* and disease traits and outcomes of interest.

My MR findings indicated evidence that an increase in *H.pylori* infection is causally associated with a decrease in LDL cholesterol, hip circumference, breast cancer, and heart rate, respectively. Sensitivity analyses also showed limited evidence of causality in the direction of the identified disease traits to *H.pylori* infection, and no strong evidence of directional pleiotropy and heterogeneity.

7.2 **Genome-wide association meta-analysis findings**

The identification of rs366337 as the SNP with the smallest P-value in the GWAS meta-analysis is potentially of interest due to its location on *LILRB2*. Although this SNP did not reach the genome-wide significance threshold, the function of this gene could suggest a biologically plausible pathway for *H.pylori* infection. Generally, the leukocyte immunoglobulin-like receptor (LILR) family are found among many cell populations in the immune system and are known to regulate immune cell activation induced by external stimuli that encourage innate and adaptive immune responses(299). The immunoregulatory properties of *LILBR2* are specifically in monocytes and dendritic cell functions (i.e. antigen presentation to T cells and pro-inflammatory cytokine production), where it encodes the LILBR2 inhibitory receptor protein(300). Studies have shown the presence of LILBR2 expression in active rheumatoid arthritis (301, 302) and in multiple cancer types such as leukaemia, breast cancer, non-small cell lung cancer and colorectal cancer(303-306). Furthermore, studies have shown evidence that LILBR2

is involved in the propagation of cancer cells(304, 307-309). The role of LILBR2 in infection has also been explored by Brown et al. 2009(310) suggesting that expression of LILBR2 at a lower level as a response to toll-like receptor signalling could be an interesting interplay of immunostimulatory and immunoinhibitory responses to bacterial infection. Through potential modulation of TLR stimulating LILR expression, and the suppression of TLR through LILR activity, this balance to induce inflammation could shed light on a possible biological mechanism for *H.pylori* disease outcomes. Furthermore, the interaction of *TLR* with *LILBR2* and their expression on the same cell types is interesting as *TLR* was identified in the Mayerele et al. 2013(42) *H.pylori* GWAS meta-analysis. This study showed strong evidence of genome-wide significance with a genetic variant located on the *TLR* locus (as discussed in Chapter 2.3 Genetic studies).

Furthermore, the identification of a highly suggestive SNP on *ST6GALNAC5* is also potentially relevant and could be explored further as literature has shown evidence of this gene associated with cardiovascular disease and cancer(311-313). This gene encodes sialyltransferase 7e and, although the function of sialyltransferases are not clearly defined, its interaction with glycosylation patterns suggest that its role includes recognition, proliferation, adhesion and differentiation of cells(314-316). The elevation of sialyltransferase activity has been shown to be associated with atherosclerosis and coronary artery disease in blood cells and serum sialic acid levels(317-320). Additionally, in an inbred Iranian population, two mutations in *ST6GALNAC5* were suggested to drive the pathogenesis of coronary artery disease(311). *ST6GALNAC5* has also been proposed to be a mediator of cancer cell passage from breast cancer metastasis to the brain resulting in increased adhesion of cancer cells to the brain endothelial cells, with cancer cells also able to across the blood-brain barrier(313).

Given the lack of defined biological mechanisms associating *H.pylori* infection with the identified genetic variants on *LILRB2* and *ST6GALNAC5*, the relevance of these findings are only suggestive. The use of animal modelling and cell biology methods to elucidate the potential biological pathways of these gene with *H.pylori* infection would be required.

7.2.1 Strengths and limitations

The strength of the GWAS meta-analysis was the overall sample size. It enabled me to identify genetic variants associated with *H.pylori* infection by increasing statistical power. In addition, it provided the opportunity to collaborate with researchers of a large consortium on GWAS of *H.pylori* infection and make plans to include the GWAS that I performed in their meta-analysis. This collaboration is going to lead to a dramatic increase in sample size and provide new insights into the genetics of *H.pylori* infection.

However, this meta-analysis that I performed was limited by the use of continuous variables (i.e. ALSPAC) and binary variables (i.e. CaPS). This is due to the differences in units used in each phenotype dataset as the continuous trait was measured in SD units and the binary trait was measured in log odds ratio. To transform the ALSPAC data from continuous to binary would have been preferable for analysis, however there is currently no internationally defined cut-off for *H.pylori* IgG antibodies measured from plasma samples that can standardise individuals as *H.pylori* seropositive (case) and *H.pylori* seronegative (control). Studies have demonstrated highly varied thresholds with non-consistent cut-off threshold(4, 42, 321-323).

In contrast, when implementing LD Score correlation analysis, the overall sample size was considered to be a limitation as the analysis did not have enough statistical power to identify genetic correlations. By repeating this analysis in a large sample size, this would increase statistical power to detect genetic correlations between *H.pylori* infection and disease outcomes of interest.

7.2.2 Implications

The implications of the GWAS meta-analysis findings is that it provide new insight into the genetic contribution of *H.pylori* infection, and provide evidence about the biological mechanisms that may underlie disease pathogenesis. The evidence supporting *LILBR2* and its association with cancer(303, 307, 309), and the correlation of *ST6GALNAC5* with coronary artery disease and cancer, respectively, could suggest a genetic basis for these diseases(311, 313). Furthermore, if further research supports the role of these genetic variants in disease development, the implementation of genetic testing and *H.pylori* eradication therapy could be used to identify individuals at risk and treat them accordingly.

7.3 **Mendelian randomization findings**

Mendelian randomization findings showed no direct causal relationship with cardiovascular disease outcomes such as coronary heart disease and stroke. However, *H.pylori* infection was shown to have a negative causal association with LDL cholesterol, hip circumference, heart rate and maternal history of breast cancer. These findings challenge the current literature as they are not in agreement with the biological pathway proposed by observational studies as they suggest that chronic *H.pylori* infection encourages cardiovascular disease outcomes(17, 18, 76, 103) and cancer(3, 43, 124, 138). These MR results therefore put into question the role of *H.pylori* infection in the development of cardiovascular disease and cancer.

One explanation for these findings could be that, due to the limited number of robust genetic instruments used in the analysis, these causal effect estimates are not representative of the true underlying biological mechanisms. Limitations due to this issue are discussed in further detail. In brief, two genetic instruments

can result in a lack of statistical power to detect causal associations between *H.pylori* infection and disease outcomes of interest. In addition, violations to instrumental variable assumptions cannot be tested using sensitivity analyses and, as a result, issues such as horizontal pleiotropy could be biasing MR findings.

Another explanation could suggest that the MR findings are true causal effect estimates. Although observational studies have shown positive associations linking *H.pylori* infection to cardiovascular disease and cancer, heterogeneity in these findings could reflect the limitations that hinder these study designs. A major limitation of observational studies is the inadequate adjustment of factors that can influence *H.pylori* infection and bias findings, resulting in spurious associations or inflated estimates of the real association that may exist. For example, observational studies investigating *H.pylori* infection have shown inflated positive associations with cardiovascular disease outcomes as a result of confounding(221, 222). The inability to make appropriate adjustments for unknown confounders are also a limitation for observational studies. Furthermore, studies with small samples sizes can also distort findings(17, 74, 324). An interpretation of these MR findings could suggest that *H.pylori* could play a beneficial role as a gastric pathogen. Literature investigating allergies, asthma and celiac disease have observationally shown an inverse association with chronic *H.pylori* infection, highlighting its potential as a useful bacterium(325-328).

7.3.1 Strengths and limitations

This method benefitted from a combination of different strengths. Firstly, the main advantage of performing MR analysis is its ability to overcome the limitations that hinder observational studies. As discussed in Chapter 4 Mendelian randomization, the use of germline genetic variants as proxies for an exposure of interest (in this case, *H.pylori*) minimises the potential for confounding, and avoids reverse causation, regression dilution bias and selection bias of participants recruited into a study(36). Secondly, the use of summary statistics from large GWAS consortia available on MR-Base(257) in a two-sample setting increases the sample size of the disease outcome sample enhancing the statistical power to detect causal associations.

The main limitation of the analysis was the lack of robust genetic variants associated with *H.pylori* infection. Unfortunately the four highly suggestive SNPs identified in the GWAS meta-analysis did not show strong enough evidence of an association with *H.pylori* infection, and were not used in MR. To utilise these SNPs as genetic instruments in the MR analysis could have resulted in violations to the instrumental variable assumptions. Thus, only two SNPs from a published GWAS meta-analysis were used as proxies for *H.pylori* infection as the exposure of interest and this decreased the statistical power to detect causal associations. In addition, sensitivity analyses such as MR-Egger regression, mode-based

approaches, and median-based approaches, were unable to be employed due to the limited number of instruments. These approaches would have been beneficial to test for directional horizontal pleiotropy across genetic instruments, and heterogeneity between instruments. As a consequence, violations to the instrumental variable assumptions and bias of the causal effect estimates could not be tested. Therefore the Wald ratio estimates and IVW causal effect estimates should be interpreted cautiously.

7.3.2 Implications

The importance of these MR findings can be highlighted in the possible benefit of *H.pylori* as a common flora. This challenges the current clinical guidelines of treating individuals with various *H.pylori* eradication therapies to prevent diseases associated with *H.pylori*(329, 330). This is especially the case with prescribing eradication therapies to individuals identified as at risk to developing cancer(331). In a RCT in a region that has a high risk of developing gastric cancer, the trial found that after a follow-up of 7.5 years incidence of gastric cancer development was similar between participants randomised to different *H.pylori* eradication treatments and participants randomised to the placebo group(332). This study example illustrates that using other study designs that are more robust at inferring causality, instead of relying on observational studies as evidence to justify the clinical relevance of eliminating *H.pylori* infection, is necessary. A reassessment of clinical guidelines may be required to determine which individuals may benefit from chronic *H.pylori* infection and those that would benefit from eradication therapy.

7.4 **Conclusion**

In summary, I found highly suggestive genetic variants that showed evidence of association with *H.pylori* infection, and MR findings that indicated that an increase in *H.pylori* infection is causally associated with a decrease in some cardiovascular traits and cancer. Further investigation is required to provide a greater understanding of the underlying biological pathways, so that appropriate *H.pylori* interventions can be implemented that benefit individuals infected with *H.pylori*.

.

8. References

1. Warren JR, Marshall B. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. The Lancet. 1983;321(8336):1273-5.
2. Marshall B, Warren JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. The Lancet. 1984;323(8390):1311-5.
3. Sipponen P, Hyvärinen H. Role of Helicobacter pylori in the pathogenesis of gastritis, peptic ulcer and gastric cancer. Scandinavian Journal of Gastroenterology. 1993;28(sup196):3-6.
4. Nomura A, Stemmermann GN, Chyou P-H, Perez-Perez GI, Blaser MJ. Helicobacter pylori infection and the risk for duodenal and gastric ulceration. Annals of internal medicine. 1994;120(12):977-81.
5. Ohata H, Kitauchi S, Yoshimura N, Mugitani K, Iwane M, Nakamura H, et al. Progression of chronic atrophic gastritis associated with Helicobacter pylori infection increases risk of gastric cancer. International journal of cancer. 2004;109(1):138-43.
6. Cavaleiro-Pinto M, Peleteiro B, Lunet N, Barros H. Helicobacter pylori infection and gastric cardia cancer: systematic review and meta-analysis. Cancer causes & control. 2011;22(3):375-87.
7. Bornschein J, Selgrad M, Warnecke M, Kuester D, Wex T, Malfertheiner P. H. pylori infection is a key risk factor for proximal gastric cancer. Digestive diseases and sciences. 2010;55(11):3124-31.
8. den Hoed CM, Kuipers EJ. Gastric cancer: how can we reduce the incidence of this disease? Current gastroenterology reports. 2016;18(7):34.
9. Jamkhande PG, Gattani SG, Farhat SA. Helicobacter pylori and cardiovascular complications: a mechanism based review on role of Helicobacter pylori in cardiovascular diseases. Integrative medicine research. 2016;5(4):244-9.
10. Kucukazman M, Yeniova O, Dal K, Yavuz B. Helicobacter pylori and cardiovascular disease. Eur Rev Med Pharmacol Sci. 2015;19(19):3731-41.
11. Rogha M, Nikvarz M, Pourmoghaddas Z, Shirneshan K, Dadkhah D, Pourmoghaddas M. Is helicobacter pylori infection a risk factor for coronary heart disease? ARYA atherosclerosis. 2012;8(1):5.
12. Farsak B, Yildirir A, Akyön Y, Pinar A, Öç M, Böke E, et al. Detection of Chlamydia pneumoniae andHelicobacter pylori DNA in Human Atherosclerotic Plaques by PCR. Journal of Clinical Microbiology. 2000;38(12):4408-11.
13. Ramezani-Binabaj M, Mobasher-Jannat A, Safiabadi M, Saburi A, Rezaee-Zavareh MS, Karimi-Sari H, et al. The Relationship between Helicobacter pylori Infection and Atherosclerosis: A Meta-analysis. Erciyes Medical Journal/Erciyes Tip Dergisi. 2016;38(3).

14. Ameriso SF, Fridman EA, Leiguarda RC, Sevlever GE. Detection of Helicobacter pylori in human carotid atherosclerotic plaques. Stroke. 2001;32(2):385-91.
15. Mayr M, Kiechl S, Mendall MA, Willeit J, Wick G, Xu Q. Increased risk of atherosclerosis is confined to CagA-positive Helicobacter pylori strains: prospective results from the Bruneck study. Stroke. 2003;34(3):610-5.
16. Al-Ghamdi A, Jiman-Fatani AA, El-Banna H. Role of Chlamydia pneumoniae, helicobacter pylori and cytomegalovirus in coronary artery disease. Pak J Pharm Sci. 2011;24(2):95-101.
17. Mendall MA, Goggin PM, Molineaux N, Levy J, Toosy T, Strachan D, et al. Relation of Helicobacter pylori infection and coronary heart disease. Heart. 1994;71(5):437-9.
18. Wincup PH, Mendall MA, Perry IJ, Strachan DP, Walker M. Prospective relations between Helicobacter pylori infection, coronary heart disease and stroke in middle-aged men. Heart. 1997;77(3):294.
19. Majka J, Róg T, Konturek PC, Konturek SJ, Bielański W, Kowalsky M, et al. Influence of chronic Helicobacter pylori infection on ischemic cerebral stroke risk factors. Medical Science Monitor. 2002;8(10):CR675-CR84.
20. Huang WS, Tseng CH, Lin CL, Tsai CH, Kao CH. Helicobacter pylori infection increases subsequent ischemic stroke risk: a nationwide population-based retrospective cohort study. QJM: An International Journal of Medicine. 2014;107(12):969-75.
21. Kinjo K, Sato H, Sato H, Shiotani I, Kurotobi T, Ohnishi Y, et al. Prevalence of Helicobacter pylori infection and its link to coronary risk factors in Japanese patients with acute myocardial infarction. Circulation journal. 2002;66(9):805-10.
22. Tabata N, Sueta D, Akasaka T, Arima Y, Sakamoto K, Yamamoto E, et al. Helicobacter pylori seropositivity in patients with Interleukin-1 polymorphisms is significantly associated with ST-Segment Elevation Myocardial Infarction. PloS one. 2016;11(11):e0166240.
23. Khodaii Z, Vakili H, Ghaderian SMH, Najar RA, Panah AST. Association of Helicobacter pylori infection with acute myocardial infarction. Coronary artery disease. 2011;22(1):6-11.
24. Pellicano R, Arena V, Rizzetto M, Ponzetto A, Mazzarello MG, Ferrari M, et al. Acute myocardial infarction andHelicobacter pylori seropositivity. International Journal of Clinical and Laboratory Research. 1999;29(4):141.
25. Hansen S, Melby KK, Aase S, Jellum E, Vollset SE. Helicobacter pylori infection and risk of cardia cancer and non-cardia gastric cancer: a nested case-control study. Scandinavian journal of gastroenterology. 1999;34(4):353-60.

26. Machida-Montani A, Sasazuki S, Inoue M, Natsukawa S, Shaura K, Koizumi Y, et al. Association of Helicobacter pylori infection and environmental factors in non-cardia gastric cancer in Japan. Gastric Cancer. 2004;7(1):46-53.
27. Brenner H, Arndt V, Stegmaier C, Ziegler H, Rothenbacher D. Is Helicobacter pylori infection a necessary condition for noncardia gastric cancer? American journal of epidemiology. 2004;159(3):252-8.
28. Mukaisho K-i, Nakayama T, Hagiwara T, Hattori T, Sugihara H. Two distinct etiologies of gastric cardia adenocarcinoma: interactions among pH, Helicobacter pylori, and bile acids. Frontiers in microbiology. 2015;6:412.
29. Hansen S, Vollset SE, Derakhshan MH, Fyfe V, Melby KK, Aase S, et al. Two distinct aetiologies of cardia cancer; evidence from premorbid serological markers of gastric atrophy and H. pylori status. Gut. 2007.
30. Ebrahim S, Smith GD. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? Human genetics. 2008;123(1):15-33.
31. Alpha-Tocopherol Beta Carotene Cancer Prevention Study G. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. New England Journal of Medicine. 1994;330(15):1029-35.
32. Beral V, Banks E, Reeves G. Evidence from randomised trials on the long-term effects of hormone replacement therapy. The Lancet. 2002;360(9337):942-4.
33. Lee IM, Cook NR, Gaziano JM, Gordon D, Ridker PM, Manson JE, et al. Vitamin E in the primary prevention of cardiovascular disease and cancer: the Women's Health Study: a randomized controlled trial. Jama. 2005;294(1):56-65.
34. Heart Protection Study G. MRC/BHF Heart Protection Study of antioxidant vitamin supplementation in 20,536 high-risk individuals: a randomized placebo-controlled trial MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomized placebo-controlled trial. Lancet. 2002;360:23-33.
35. Investigators GI-P. Dietary supplementation with n-3 polyunsaturated fatty acids and vitamin E after myocardial infarction: results of the GISSI-Prevenzione trial. The Lancet. 1999;354(9177):447-55.
36. Davey Smith G, Ebrahim S. Mendelian randomization: genetic variants as instruments for strengthening causal inference in observational studies. 2008.
37. Malfertheiner P, Megraud F, O'Morain CA, Gisbert JP, Kuipers EJ, Axon AT, et al. Management of Helicobacter pylori infection—the Maastricht V/Florence consensus report. Gut. 2017;66(1):6-30.
38. Beek D, De Craen AJM. A systematic review of Helicobacter pylori eradication therapy—the impact of antimicrobial resistance on eradication rates. Alimentary pharmacology & therapeutics. 1999;13(8):1047-55.

39. Bock H, Koop H, Lehn N, Heep M. Rifabutin-based triple therapy after failure of Helicobacter pylori eradication treatment: preliminary experience. Journal of clinical gastroenterology. 2000;31(3):222-5.
40. Van Cutsem E, Sagaert X, Topal B, Haustermans K, Prenen H. Gastric cancer. The Lancet. 2016;388(10060):2654-64.
41. Roth GA, Johnson C, Abajobir A, Abd-Allah F, Abera SF, Abyu G, et al. Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. Journal of the American College of Cardiology. 2017;70(1):1-25.
42. Mayerle J, den Hoed CM, Schurmann C, Stolk L, Homuth G, Peters MJ, et al. Identification of genetic loci associated with Helicobacter pylori serologic status. Jama. 2013;309(18):1912-20.
43. Hatakeyama M. Oncogenic mechanisms of the Helicobacter pylori CagA protein. Nature Reviews Cancer. 2004;4(9):688.
44. Peek Jr RM, Blaser MJ. Helicobacter pylori and gastrointestinal tract adenocarcinomas. Nature Reviews Cancer. 2002;2(1):28.
45. Rothenbacher D, Brenner H. Burden of Helicobacter pylori and H. pylori-related diseases in developed countries: recent developments and future implications. Microbes and Infection. 2003;5(8):693-703.
46. Hunt RH, Xiao SD, Megraud F, Leon-Barua R, Bazzoli F, Van der Merwe S, et al. Helicobacter pylori in developing countries. World gastroenterology organisation global guideline. J Gastrointestin Liver Dis. 2011;20(3):299-304.
47. Correa P. Human gastric carcinogenesis: a multistep and multifactorial process—first American Cancer Society award lecture on cancer epidemiology and prevention. Cancer research. 1992;52(24):6735-40.
48. Nabwera HM, Logan RP. Epidemiology of Helicobacter pylori: transmission, translocation and extragastric reservoirs. Journal of physiology and pharmacology: an official journal of the Polish Physiological Society. 1999;50(5):711-22.
49. Parsonnet J, Shmuely H, Haggerty T. Fecal and oral shedding of Helicobacter pylori from healthy infected adults. Jama. 1999;282(23):2240-5.
50. Malaty HM, El-Kasabany A, Graham DY, Miller CC, Reddy SG, Srinivasan SR, et al. Age at acquisition of Helicobacter pylori infection: a follow-up study from infancy to adulthood. The Lancet. 2002;359(9310):931-5.
51. Malaty HM, Graham DY, Wattigney WA, Srinivasan SR, Osato M, Berenson GS. Natural history of Helicobacter pylori infection in childhood: 12-year follow-up cohort study in a biracial community. Clinical Infectious Diseases. 1999;28(2):279-82.

52. Pérez-Pérez GI, Sack RB, Reid R, Santosham M, Croll J, Blaser MJ. Transient and persistent Helicobacter pylori colonization in Native American children. Journal of clinical microbiology. 2003;41(6):2401-7.
53. Rowland M, Drumm B. Clinical signicicance of Helicobacter infection in children. British medical bulletin. 1998;54(1):95-103.
54. Perez-Perez GI, Rothenbacher D, Brenner H. Epidemiology of Helicobacter pylori infection. Helicobacter. 2004;9:1-6.
55. Zamani M, Ebrahimtabar F, Zamani V, Miller WH, Alizadeh-Navaei R, Shokri-Shirvani J, et al. Systematic review with meta-analysis: the worldwide prevalence of Helicobacter pylori infection. Alimentary pharmacology & therapeutics. 2018;47(7):868-76.
56. Wizla-Derambure N, Michaud L, Ategbo S, Vincent P, Ganga-Zandzou S, Turck D, et al. Familial and community environmental risk factors for Helicobacter pylori infection in children and adolescents. Journal of pediatric gastroenterology and nutrition. 2001;33(1):58-63.
57. McCallion WA, Murray LJ, Bailie AG, Dalzell AM, O'Reilly DP, Bamford KB. Helicobacter pylori infection in children: relation with current household living conditions. Gut. 1996;39(1):18-21.
58. Jafri W, Yakoob J, Abid S, Siddiqui S, Awan S, Nizami SQ. Helicobacter pylori infection in children: population-based age-specific prevalence and risk factors in a developing country. Acta Paediatrica. 2010;99(2):279-82.
59. Malaty HM, Graham DY. Importance of childhood socioeconomic status on the current prevalence of Helicobacter pylori infection. Gut. 1994;35(6):742-5.
60. Bastos J, Peleteiro B, Pinto H, Marinho A, Guimarães JT, Ramos E, et al. Prevalence, incidence and risk factors for Helicobacter pylori infection in a cohort of Portuguese adolescents (EpiTeen). Digestive and Liver Disease. 2013;45(4):290-5.
61. Torre LA, Siegel RL, Ward EM, Jemal A. Global cancer incidence and mortality rates and trends—an update. Cancer Epidemiology and Prevention Biomarkers. 2016;25(1):16-27.
62. Hooi JKY, Lai WY, Ng WK, Suen MMY, Underwood FE, Tanyingoh D, et al. Global prevalence of Helicobacter pylori infection: systematic review and meta-analysis. Gastroenterology. 2017;153(2):420-9.
63. Peleteiro B, Bastos A, Ferro A, Lunet N. Prevalence of Helicobacter pylori infection worldwide: a systematic review of studies with national coverage. Digestive diseases and sciences. 2014;59(8):1698-709.
64. Inoue M, Tsugane S. Epidemiology of gastric cancer in Japan. Postgraduate medical journal. 2005;81(957):419-24.

65. Inoue M. Changing epidemiology of Helicobacter pylori in Japan. Gastric Cancer. 2017;20(1):3-7.
66. Axon A. Helicobacter pylori and public health. Helicobacter. 2014;19(s1):68-73.
67. Goh KL, Chan WK, Shiota S, Yamaoka Y. Epidemiology of Helicobacter pylori infection and public health implications. Helicobacter. 2011;16:1-9.
68. Moayyedi P, Hunt RH. Helicobacter pylori public health implications. Helicobacter. 2004;9:67-72.
69. Plummer M, Franceschi S, Vignat J, Forman D, de Martel C. Global burden of gastric cancer attributable to Helicobacter pylori. International journal of cancer. 2015;136(2):487-90.
70. Franceschi F, Sepulveda AR, Gasbarrini A, Pola P, Silveri NG, Gasbarrini G, et al. Cross-reactivity of anti-CagA antibodies with vascular wall antigens: possible pathogenic link between Helicobacter pylori infection and atherosclerosis. Circulation. 2002;106(4):430-4.
71. Zhu J, Quyyumi AA, Muhlestein JB, Nieto FJ, Horne BD, Zalles-Ganley A, et al. Lack of association of Helicobacter pylori infection with coronary artery disease and frequency of acute myocardial infarction or death. American journal of cardiology. 2002;89(2):155-8.
72. Koenig W, Rothenbacher D, Hoffmeister A, Miller M, Bode G, Adler G, et al. Infection with Helicobacter pylori is not a major independent risk factor for stable coronary heart disease: lack of a role of cytotoxin-associated protein A–positive strains and absence of a systemic inflammatory response. Circulation. 1999;100(23):2326-31.
73. Wald NJ, Law MR, Morris JK, Bagnall AM. Helicobacter pylori infection and mortality from ischaemic heart disease: negative result from a large, prospective study. Bmj. 1997;315(7117):1199-201.
74. Markus HS, Mendall MA. Helicobacter pylori infection: a risk factor for ischaemic cerebrovascular disease and carotid atheroma. Journal of Neurology, Neurosurgery & Psychiatry. 1998;64(1):104-7.
75. Grau AJ, Buggle F, Lichy C, Brandt T, Becher H, Rudi J. Helicobacter pylori infection as an independent risk factor for cerebral ischemia of atherothrombotic origin. Journal of the neurological sciences. 2001;186(1-2):1-5.
76. Patel P, Mendall MA, Carrington D, Strachan DP, Leatham E, Molineaux N, et al. Association of Helicobacter pylori and Chlamydia pneumoniae infections with coronary heart disease and cardiovascular risk factors. Bmj. 1995;311(7007):711-4.
77. Niemelä S, Karttunen T, Korhonen T, Läärä E, Karttunen R, Ikäheimo M, et al. Could Helicobacter pylori infection increase the risk of coronary heart disease by modifying serum lipid concentrations? Heart. 1996;75(6):573-5.

78. Gen R, Demir M, Ataseven H. Effect of Helicobacter pylori eradication on insulin resistance, serum lipids and low-grade inflammation. Southern medical journal. 2010;103(3):190-6.
79. Zhang S, Guo Y, Ma Y, Teng Y. Cytotoxin-associated gene-A-seropositive virulent strains of Helicobacter pylori and atherosclerotic diseases: a systematic review. 2008.
80. Birnie DH, Holme ER, McKay IC, Hood S, McColl KEL, Hillis WS. Association between antibodies to heat shock protein 65 and coronary atherosclerosis. Possible mechanism of action of Helicobacter pylori and other bacterial infections in increasing cardiovascular risk. European heart journal. 1998;19(3):387-94.
81. Oshima T, Ozono R, Yano Y, Oishi Y, Teragawa H, Higashi Y, et al. Association of Helicobacter pyloriinfection with systemic inflammation and endothelial dysfunction in healthy male subjects. Journal of the American College of Cardiology. 2005;45(8):1219-22.
82. Lenzi C, Palazzuoli A, Giordano N, Alegente G, Gonnelli C, Campagna MS, et al. H pylori infection and systemic antibodies to CagA and heat shock protein 60 in patients with coronary heart disease. World Journal of Gastroenterology: WJG. 2006;12(48):7815.
83. Kanbay M, Gür G, Yücel M, Yilmaz U, Muderrisoglu H. Helicobacter pylori seroprevalence in patients with coronary artery disease. Digestive diseases and sciences. 2005;50(11):2071-4.
84. Hoffmeister A, Rothenbacher D, Bode G, Persson K, März W, Nauck MA, et al. Current infection with Helicobacter pylori, but not seropositivity to Chlamydia pneumoniae or cytomegalovirus, is associated with an atherogenic, modified lipid profile. Arteriosclerosis, thrombosis, and vascular biology. 2001;21(3):427-32.
85. Assmann G, Gotto AM. HDL cholesterol and protective factors in atherosclerosis. Circulation. 2004;109(23 suppl 1):III-8.
86. Kontush A, Chapman MJ. Antiatherogenic small, dense HDL—guardian angel of the arterial wall? Nature Reviews Cardiology. 2006;3(3):144.
87. Park S-H, Park JHY, Kang J-S, Kang Y-H. Involvement of transcription factors in plasma HDL protection against TNF-α-induced vascular cell adhesion molecule-1 expression. The international journal of biochemistry & cell biology. 2003;35(2):168-82.
88. Clay MA, Pyle DH, Rye K-A, Vadas MA, Gamble JR, Barter PJ. Time sequence of the inhibition of endothelial adhesion molecule expression by reconstituted high density lipoproteins. Atherosclerosis. 2001;157(1):23-9.
89. Brites F, Martin M, Guillas I, Kontush A. Antioxidative activity of high-density lipoprotein (HDL): mechanistic insights into potential clinical benefit. BBA clinical. 2017;8:66-77.
90. Eren E, Yılmaz N, Aydin O, Ellidağ HY. Anticipatory role of high density lipoprotein and endothelial dysfunction: an overview. The open biochemistry journal. 2014;8:100.

91. Kratzer A, Giral H, Landmesser U. High-density lipoproteins as modulators of endothelial cell functions: alterations in patients with coronary artery disease. Cardiovascular research. 2014;103(3):350-61.
92. Paximadas S, Pagoni ST, Kosmidis M, Tsarouchas X, Christou M, Chatziantonakis N, et al. W12-P-050 The lipid profile in adults with negative or positive antibody Helicobacter pylori. Atherosclerosis Supplements. 2005;6(1):74.
93. Lim SH, Kwon J-W, Kim N, Kim GH, Kang JM, Park MJ, et al. Prevalence and risk factors of Helicobacter pylori infection in Korea: nationwide multicenter study over 13 years. BMC gastroenterology. 2013;13(1):104.
94. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. The Lancet. 2012;380(9841):572-80.
95. Holmes MV, Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, et al. Mendelian randomization of blood lipids for coronary heart disease. European heart journal. 2014;36(9):539-50.
96. Siddiqi HK, Kiss D, Rader D. HDL-cholesterol and cardiovascular disease: rethinking our approach. Current opinion in cardiology. 2015;30(5):536-42.
97. Kalia N, Jones C, Bardhan KD, Reed MWR, Atherton JC, Brown NJ. Effects of genotypically different strains of Helicobacter pylori on human microvascular endothelial cells in vitro. Digestive diseases and sciences. 2001;46(1):54-61.
98. Kutluana U, Simsek I, Akarsu M, Kupelioglu A, Karasu S, Altekin E. Is there a possible relation between atrophic gastritis and premature atherosclerosis? Helicobacter. 2005;10(6):623-9.
99. Sung JJ, Sanderson JE. Hyperhomocysteinaemia, Helicobacter pylori, and coronary heart disease. Heart. 1996;76(4):305-7.
100. Collaboration CRPCHDG. Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data. Bmj. 2011;342:d548.
101. Kaplan M, Yavuz SS, Cinar B, Koksal V, Kut MS, Yapici F, et al. Detection of Chlamydia pneumoniae and Helicobacter pylori in atherosclerotic plaques of carotid artery by polymerase chain reaction. International journal of infectious diseases. 2006;10(2):116-23.
102. Kowalski M, Rees W, Konturek PC, Grove R, Scheffold T, Meixner H, et al. Detection of Helicobacter pylori specific DNA in human atheromatous coronary arteries and its association to prior myocardial infarction and unstable angina. Digestive and Liver Disease. 2002;34(6):398-402.
103. Kowalski M. Helicobacter pylori (H. pylori) infection in coronary artery disease: influence of H. pylori eradication on coronary artery lumen after percutaneous transluminal coronary angioplasty. The

detection of H. pylori specific DNA in human coronary atherosclerotic plaque. Journal of physiology and pharmacology: an official journal of the Polish Physiological Society. 2001;52(1 Suppl 1):3-31.
104. Blasi F, Denti F, Erba M, Cosentini R, Raccanelli R, Rinaldi A, et al. Detection of Chlamydia pneumoniae but not Helicobacter pylori in atherosclerotic plaques of aortic aneurysms. Journal of Clinical Microbiology. 1996;34(11):2766-9.
105. Dore MP, Sepulveda AR, Bacciu PP, Blasi F, Simula L, Marras L, et al. Detection of Chlamydiae pneumoniae but not Helicobacter pylori DNA in atherosclerosis plaques. Digestive diseases and sciences. 2003;48(5):945-51.
106. Malnick SDH, Goland S, Kaftoury A, Schwarz H, Pasik S, Mashiach A, et al. Evaluation of carotid arterial plaques after endarterectomy for Helicobacter pylori infection. The American journal of cardiology. 1999;83(11):1586-7.
107. Kahan T, Lundman P, Olsson G, Wendt M. Greater than normal prevalence of seropositivity for Helicobacter pylori among patients who have suffered myocardial infarction. Coronary artery disease. 2000;11(7):523-6.
108. Strachan DP, Mendall MA, Carrington D, Butland BK, Yarnell JWG, Sweetnam PM, et al. Relation of Helicobacter pylori infection to 13-year mortality and incident ischemic heart disease in the caerphilly prospective heart disease study. Circulation. 1998;98(13):1286-90.
109. Danesh J, Youngman L, Clark S, Parish S, Peto R, Collins R. Helicobacter pylori infection and early onset myocardial infarction: case-control and sibling pairs study. Bmj. 1999;319(7218):1157-62.
110. Haider AW, Wilson PWF, Larson MG, Evans JC, Michelson EL, Wolf PA, et al. The association of seropositivity to Helicobacter pylori, Chlamydia pneumoniae, and cytomegalovirus with risk of cardiovascular disease: a prospective study. Journal of the American College of Cardiology. 2002;40(8):1408-13.
111. Strandberg TE, Tilvis RS, Vuoristo M, Lindroos M, Kosunen TU. Prospective study of Helicobacter pylori seropositivity and cardiovascular diseases in a general elderly population. Bmj. 1997;314(7090):1317.
112. Ridker PM, Hennekens CH, Buring JE, Kundsin R, Shih J. Baseline IgG antibody titers to Chlamydia pneumoniae, Helicobacter pylori, herpes simplex virus, and cytomegalovirus and the risk for cardiovascular disease in women. Annals of internal medicine. 1999;131(8):573-7.
113. Yu XJ, Yang X, Feng L, Wang LL, Dong QJ. Association between Helicobacter pylori infection and angiographically demonstrated coronary artery disease: A meta‐analysis. Experimental and therapeutic medicine. 2017;13(2):787-93.
114. Pasceri V, Patti G, Cammarota G, Pristipino C, Richichi G, Di Sciascio G. Virulent strains of Helicobacter pylori and vascular diseases: a meta-analysis. American heart journal. 2006;151(6):1215-22.

115. Sawayama Y, Ariyama I, Hamada M, Otaguro S, Machi T, Taira Y, et al. Association between chronic Helicobacter pylori infection and acute ischemic stroke: Fukuoka Harasanshin Atherosclerosis Trial (FHAT). Atherosclerosis. 2005;178(2):303-9.
116. Cremonini F, Gabrielli M, Gasbarrini G, Pola P, Gasbarrini A. The relationship between chronic H. pylori infection, CagA seropositivity and stroke: meta-analysis. Atherosclerosis. 2004;173(2):253-9.
117. Wang ZW, Li Y, Huang LY, Guan QK, Xu DW, Zhou WK, et al. Helicobacter pylori infection contributes to high risk of ischemic stroke: evidence from a meta-analysis. Journal of neurology. 2012;259(12):2527-37.
118. Yu M, Zhang Y, Yang Z, Ding J, Xie C, Lu N. Association between Helicobacter pylori infection and stroke: a meta-analysis of prospective observational studies. Journal of Stroke and Cerebrovascular Diseases. 2014;23(9):2233-9.
119. International Agency for Research on C. Evaluation of the Carcinogenic Risk to Humans. Schistosomes, Liver Flukes and Helicobacter pylori. IARC Monographs; 1994.
120. Eslick GD, Lim LLY, Byles JE, Xia HHX, Talley NJ. Association of Helicobacter pylori infection with gastric carcinoma: a meta-analysis. The American journal of gastroenterology. 1999;94(9):2373.
121. Hansson L-E, Engstrand L, Nyrén O, Evans Jr DJ, Lindgren A, Bergström R, et al. Helicobacter pylori infection: independent risk indicator of gastric adenocarcinoma. Gastroenterology. 1993;105(4):1098-103.
122. Wotherspoon AC, Ortiz-Hidalgo C, Falzon MR, Isaacson PG. Helicobacter pylori-associated gastritis and primary B-cell gastric lymphoma. The Lancet. 1991;338(8776):1175-6.
123. Eck M, Schmausser B, Haas R, Greiner A, Czub S, Muller-Hermelink HK. MALT-type lymphoma of the stomach is associated with Helicobacter pylori strains expressing the CagA protein. Gastroenterology. 1997;112(5):1482-6.
124. Peek JRM, Miller GG, Tham KT, Perez-Perez GI, Zhao X, Atherton JC, et al. Heightened inflammatory response and cytokine expression in vivo to cagA+ Helicobacter pylori strains. Laboratory investigation; a journal of technical methods and pathology. 1995;73(6):760-70.
125. Forman D, Burley VJ. Gastric cancer: global pattern of the disease and an overview of environmental risk factors. Best practice & research Clinical gastroenterology. 2006;20(4):633-49.
126. Brenner H, Rothenbacher D, Arndt V. Epidemiology of stomach cancer. Cancer Epidemiology: Springer; 2009. p. 467-77.
127. Sipponen P, Marshall BJ. Gastritis and gastric cancer: Western countries. Gastroenterology Clinics. 2000;29(3):579-92.

128. Kamangar F, Dawsey SM, Blaser MJ, Perez-Perez GI, Pietinen P, Newschaffer CJ, et al. Opposing risks of gastric cardia and noncardia gastric adenocarcinomas associated with Helicobacter pylori seropositivity. Journal of the National Cancer Institute. 2006;98(20):1445-52.
129. Helicobacter, Cancer Collaborative G. Gastric cancer and Helicobacter pylori: a combined analysis of 12 case control studies nested within prospective cohorts. Gut. 2001;49(3):347-53.
130. Wu AH, Crabtree JE, Bernstein L, Hawtin P, Cockburn M, Tseng Cc, et al. Role of Helicobacter pylori CagA+ strains and risk of adenocarcinoma of the stomach and esophagus. International journal of cancer. 2003;103(6):815-21.
131. Vicari JJ, Peek RM, Falk GW, Goldblum JR, Easley KA, Schnell J, et al. The seroprevalence of cagA-positive Helicobacter pylori strains in the spectrum of gastroesophageal reflux disease. Gastroenterology. 1998;115(1):50-7.
132. Weston AP, Badr AS, Topalovski M, Cherian R, Dixon A, Hassanein RS. Prospective evaluation of the prevalence of gastric Helicobacter pylori infection in patients with GERD, Barrett's esophagus, Barrett's dysplasia, and Barrett's adenocarcinoma. The American journal of gastroenterology. 2000;95(2):387-94.
133. Vieth M, Masoud B, Meining A, Stolte M. Helicobacter pylori infection: protection against Barrett's mucosa and neoplasia? Digestion. 2000;62(4):225-31.
134. Limburg PJ, Qiao Y-L, Mark SD, Wang G-Q, Perez-Perez GI, Blaser MJ, et al. Helicobacter pylori seropositivity and subsite-specific gastric cancer risks in Linxian, China. Journal of the National Cancer Institute. 2001;93(3):226-33.
135. Kamangar F, Qiao YL, Blaser MJ, Sun XD, Katki H, Fan JH, et al. Helicobacter pylori and oesophageal and gastric cancers in a prospective study in China. British journal of cancer. 2007;96(1):172.
136. Dawsey SM, Mark SD, Taylor PR, Limburg PJ. Gastric cancer and H pylori. Gut. 2002;51(3):457-8.
137. Marsman WA, Tytgat GNJ, Ten Kate FJW, Van Lanschot JJB. Differences and similarities of adenocarcinomas of the esophagus and esophagogastric junction. Journal of surgical oncology. 2005;92(3):160-8.
138. Wroblewski LE, Peek RM, Wilson KT. Helicobacter pylori and gastric cancer: factors that modulate disease risk. Clinical microbiology reviews. 2010;23(4):713-39.
139. Blaser MJ, Berg DE. Helicobacter pylori genetic diversity and risk of human disease. The Journal of clinical investigation. 2001;107(7):767-73.
140. Tomb J-F, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, et al. Corrections: The complete genome sequence of the gastric pathogen Helicobacter pylori. Nature. 1997;389(6649):412.

141. Salama N, Guillemin K, McDaniel TK, Sherlock G, Tompkins L, Falkow S. A whole-genome microarray reveals genetic diversity among Helicobacter pylori strains. Proceedings of the national Academy of Sciences. 2000;97(26):14668-73.
142. Suzuki R, Shiota S, Yamaoka Y. Molecular epidemiology, population genetics, and pathogenic role of Helicobacter pylori. Infection, Genetics and Evolution. 2012;12(2):203-13.
143. Suzuki H, Mori H. Different Pathophysiology of Gastritis between East and West? An Asian Perspective. Inflammatory Intestinal Diseases. 2016;1(3):123-8.
144. Cover TL. Helicobacter pylori diversity and gastric cancer risk. MBio. 2016;7(1):e01869-15.
145. Gwack J, Shin A, Kim C-S, Ko KP, Kim Y, Jun JK, et al. CagA-producing Helicobacter pylori and increased risk of gastric cancer: a nested case–control study in Korea. British journal of cancer. 2006;95(5):639.
146. Plummer M, van Doorn L-J, Franceschi S, Kleter B, Canzian F, Vivas J, et al. Helicobacter pylori cytotoxin-associated genotype and gastric precancerous lesions. Journal of the National Cancer Institute. 2007;99(17):1328-34.
147. Nomura AMY, Lee J, Stemmermann GN, Nomura RY, Perez-Perez GI, Blaser MJ. Helicobacter pylori CagA seropositivity and gastric carcinoma risk in a Japanese American population. The Journal of infectious diseases. 2002;186(8):1138-44.
148. Kikuchi S, Crabtree JE, Forman D, Kurosawa M. Association between infections with CagA-positive or-negative strains of Helicobacter pylori and risk for gastric cancer in young adults. The American journal of gastroenterology. 1999;94(12):3455.
149. Rudi J, Kolb C, Maiwald M, Zuna I, Von Herbay A, Galle PR, et al. Serum antibodies against Helicobacter pylori proteins VacA and CagA are associated with increased risk for gastric adenocarcinoma. Digestive diseases and sciences. 1997;42(8):1652-9.
150. Azuma T, Yamakawa A, Yamazaki S, Fukuta K, Ohtani M, Ito Y, et al. Correlation between variation of the 3′ region of the cagA gene in Helicobacter pylori and disease outcome in Japan. The Journal of infectious diseases. 2002;186(11):1621-30.
151. Naito M, Yamazaki T, Tsutsumi R, Higashi H, Onoe K, Yamazaki S, et al. Influence of EPIYA-repeat polymorphism on the phosphorylation-dependent biological activity of Helicobacter pylori CagA. Gastroenterology. 2006;130(4):1181-90.
152. Argent RH, Hale JL, El-Omar EM, Atherton JC. Differences in Helicobacter pylori CagA tyrosine phosphorylation motif patterns between western and East Asian strains, and influences on interleukin-8 secretion. Journal of medical microbiology. 2008;57(9):1062-7.

153. Jones KR, Joo YM, Jang S, Yoo Y-J, Lee HS, Chung I-S, et al. Polymorphism in the CagA EPIYA motif impacts development of gastric cancer. Journal of clinical microbiology. 2009;47(4):959-68.
154. Satomi S, Yamakawa A, Matsunaga S, Masaki R, Inagaki T, Okuda T, et al. Relationship between the diversity of the cagA gene of Helicobacter pylori and gastric cancer in Okinawa, Japan. Journal of gastroenterology. 2006;41(7):668-73.
155. Stein M, Bagnoli F, Halenbeck R, Rappuoli R, Fantl WJ, Covacci A. c-Src/Lyn kinases activate Helicobacter pylori CagA through tyrosine phosphorylation of the EPIYA motifs. Molecular microbiology. 2002;43(4):971-80.
156. Sicinschi LA, Correa P, Peek RM, Camargo MC, Piazuelo MB, Romero-Gallo J, et al. CagA C-terminal variations in Helicobacter pylori strains from Colombian patients with gastric precancerous lesions. Clinical Microbiology and Infection. 2010;16(4):369-78.
157. Azuma T. Helicobacter pylori CagA protein variation associated with gastric cancer in Asia. Journal of gastroenterology. 2004;39(2):97-103.
158. Higashi H, Tsutsumi R, Fujita A, Yamazaki S, Asaka M, Azuma T, et al. Biological activity of the Helicobacter pylori virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. Proceedings of the National Academy of Sciences. 2002;99(22):14428-33.
159. Yuan X-y, Yan J-J, Yang Y-c, Wu C-m, Hu Y, Geng J-l. Helicobacter pylori with East Asian-type cagPAI genes is more virulent than strains with Western-type in some cagPAI genes. brazilian journal of microbiology. 2017;48(2):218-24.
160. Lee KE, Khoi PN, Xia Y, Park JS, Joo YE, Kim KK, et al. Helicobacter pylori and interleukin-8 in gastric cancer. World Journal of Gastroenterology: WJG. 2013;19(45):8192.
161. Azuma T, Yamazaki S, Yamakawa A, Ohtani M, Muramatsu A, Suto H, et al. Association between diversity in the Src homology 2 domain-containing tyrosine phosphatase binding site of Helicobacter pylori CagA protein and gastric atrophy and cancer. Journal of Infectious Diseases. 2004;189(5):820-7.
162. Hatakeyama M. Structure and function of Helicobacter pylori CagA, the first-identified bacterial protein involved in human cancer. Proceedings of the Japan Academy, Series B. 2017;93(4):196-219.
163. Kamada T, Haruma K, Ito M, Inoue K, Manabe N, Matsumoto H, et al. Time trends in Helicobacter pylori infection and atrophic gastritis over 40 years in Japan. Helicobacter. 2015;20(3):192-8.
164. Xu C, Yan M, Sun Y, Joo J, Wan X, Yu C, et al. Prevalence of Helicobacter pylori infection and its relation with body mass index in a Chinese population. Helicobacter. 2014;19(6):437-42.

165. Yang L. Incidence and mortality of gastric cancer in China. World journal of gastroenterology: WJG. 2006;12(1):17.
166. Yamaoka Y, Kato M, Asaka M. Geographic differences in gastric cancer incidence can be explained by differences between Helicobacter pylori strains. Internal medicine. 2008;47(12):1077-83.
167. Fock KM, Ang TL. Epidemiology of Helicobacter pylori infection and gastric cancer in Asia. Journal of gastroenterology and hepatology. 2010;25(3):479-86.
168. Cover TL, Glupczynski Y, Lage AP, Burette A, Tummuru MK, Perez-Perez GI, et al. Serologic detection of infection with cagA+ Helicobacter pylori strains. Journal of clinical microbiology. 1995;33(6):1496-500.
169. Chiurillo MA, Moran Y, Cañas M, Valderrama E, Granda N, Sayegh M, et al. Genotyping of Helicobacter pylori virulence-associated genes shows high diversity of strains infecting patients in western Venezuela. International Journal of Infectious Diseases. 2013;17(9):e750-e6.
170. Bastos J, Peleteiro B, Barros R, Alves L, Severo M, de Fátima Pina M, et al. Sociodemographic determinants of prevalence and incidence of Helicobacter pylori infection in Portuguese adults. Helicobacter. 2013;18(6):413-22.
171. Eusebi LH, Zagari RM, Bazzoli F. Epidemiology of Helicobacter pylori infection. Helicobacter. 2014;19(s1):1-5.
172. Zhou W, Yamazaki S, Yamakawa A, Ohtani M, Ito Y, Keida Y, et al. The diversity of vacA and cagA genes of Helicobacter pylori in East Asia. FEMS Immunology & Medical Microbiology. 2004;40(1):81-7.
173. Yamazaki S, Yamakawa A, Okuda T, Ohtani M, Suto H, Ito Y, et al. Distinct diversity of vacA, cagA, and cagE genes of Helicobacter pylori associated with peptic ulcer in Japan. Journal of clinical microbiology. 2005;43(8):3906-16.
174. Wong BCY, Yin Y, Berg DE, Xia HHX, Zhang JZ, Wang WH, et al. Distribution of distinct vacA, cagA and iceA alleles in Helicobacter pylori in Hong Kong. Helicobacter. 2001;6(4):317-24.
175. Atherton JC, Cao P, Peek RM, Tummuru MKR, Blaser MJ, Cover TL. Mosaicism in vacuolating cytotoxin alleles of Helicobacter pylori association of specific vacA types with cytotoxin production and peptic ulceration. Journal of Biological Chemistry. 1995;270(30):17771-7.
176. Winter JA, Letley DP, Cook KW, Rhead JL, Zaitoun AAM, Ingram RJM, et al. A role for the vacuolating cytotoxin, VacA, in colonization and Helicobacter pylori–induced metaplasia in the stomach. The Journal of infectious diseases. 2014;210(6):954-63.
177. Abadi ATB, Lee YY. Helicobacter pylori vacA as marker for gastric cancer and gastroduodenal diseases: one but not the only factor. Journal of clinical microbiology. 2014;52(12):4451-.

178. Atherton JC, Peek RM, Tham KT, Cover TL, Blaser MJ. Clinical and pathological importance of heterogeneity in vacA, the vacuolating cytotoxin gene of Helicobacter pylori. Gastroenterology. 1997;112(1):92-9.
179. Palframan SL, Kwok T, Gabriel K. Vacuolating cytotoxin A (VacA), a key toxin for Helicobacter pylori pathogenesis. Frontiers in cellular and infection microbiology. 2012;2:92.
180. McClain MS, Cao P, Iwamoto H, Vinion-Dubiel AD, Szabo G, Shao Z, et al. A 12-amino-acid segment, present in type s2 but not type s1 Helicobacter pylori VacA proteins, abolishes cytotoxin activity and alters membrane channel formation. Journal of bacteriology. 2001;183(22):6499-508.
181. van Doorn L-J, Figueiredo C, Sanna R, Pena S, Midolo P, Ng EKW, et al. Expanding allelic diversity of Helicobacter pylori vacA. Journal of clinical microbiology. 1998;36(9):2597-603.
182. Kidd M, Lastovica AJ, Atherton JC, Louw JA. Heterogeneity in the Helicobacter pylori vacA and cagA genes: association with gastroduodenal disease in South Africa? Gut. 1999;45(4):499-502.
183. Miehlke S, Kirsch C, Agha-Amiri K, Günther T, Lehn N, Malfertheiner P, et al. The Helicobacter pylori vacA s1, m1 genotype and cagA is associated with gastric carcinoma in Germany. International journal of cancer. 2000;87(3):322-7.
184. Zambon CF, Navaglia F, Basso D, Rugge M, Plebani M. Helicobacter pylori babA2, cagA, and s1 vacA genes work synergistically in causing intestinal metaplasia. Journal of clinical pathology. 2003;56(4):287-91.
185. Siddique I, Al-Qabandi A, Al-Ali J, Alazmi W, Memon A, Mustafa AS, et al. Association between Helicobacter pylori genotypes and severity of chronic gastritis, peptic ulcer disease and gastric mucosal interleukin-8 levels: Evidence from a study in the Middle East. Gut pathogens. 2014;6(1):41.
186. Basso D, Zambon CF, Letley DP, Stranges A, Marchet A, Rhead JL, et al. Clinical relevance of Helicobacter pylori cagA and vacA gene polymorphisms. Gastroenterology. 2008;135(1):91-9.
187. Cover TL, Blanke SR. Helicobacter pylori VacA, a paradigm for toxin multifunctionality. Nature Reviews Microbiology. 2005;3(4):320.
188. Yokoyama K, Higashi H, Ishikawa S, Fujii Y, Kondo S, Kato H, et al. Functional antagonism between Helicobacter pylori CagA and vacuolating toxin VacA in control of the NFAT signaling pathway in gastric epithelial cells. Proceedings of the National Academy of Sciences. 2005;102(27):9661-6.
189. Argent RH, Thomas RJ, Letley DP, Rittig MG, Hardie KR, Atherton JC. Functional association between the Helicobacter pylori virulence factors VacA and CagA. Journal of Medical Microbiology. 2008;57(2):145-50.
190. Oldani A, Cormont M, Hofman V, Chiozzi V, Oregioni O, Canonici A, et al. Helicobacter pylori counteracts the apoptotic action of its VacA toxin by injecting the CagA protein into gastric epithelial cells. PLoS pathogens. 2009;5(10):e1000603.

191. Pagliaccia C, De Bernard M, Lupetti P, Ji X, Burroni D, Cover TL, et al. The m2 form of the Helicobacter pylori cytotoxin has cell type-specific vacuolating activity. Proceedings of the National Academy of Sciences. 1998;95(17):10212-7.
192. Pan Z-J, Berg DE, van der Hulst RWM, Su W-W, Raudonikiene A, Xiao S-D, et al. Prevalence of vacuolating cytotoxin production and distribution of distinct vacA alleles in Helicobacter pylori from China. Journal of Infectious Diseases. 1998;178(1):220-6.
193. Ji X, Fernandez T, Burroni D, Pagliaccia C, Atherton JC, Reyrat J-M, et al. Cell Specificity of Helicobacter pyloriCytotoxin Is Determined by a Short Region in the Polymorphic Midregion. Infection and immunity. 2000;68(6):3754-7.
194. Amieva MR, El–Omar EM. Host-bacterial interactions in Helicobacter pylori infection. Gastroenterology. 2008;134(1):306-23.
195. Figueiredo C, Van Doorn LJ, Nogueira C, Soares JM, Pinho C, Figueira P, et al. Helicobacter pylori genotypes are associated with clinical outcome in Portuguese patients and show a high prevalence of infections with multiple strains. Scandinavian journal of gastroenterology. 2001;36(2):128-35.
196. González CA, Figueiredo C, Lic CB, Ferreira RM, Pardo ML, Liso JMR, et al. Helicobacter pylori cagA and vacA genotypes as predictors of progression of gastric preneoplastic lesions: a long-term follow-up in a high-risk area in Spain. The American journal of gastroenterology. 2011;106(5):867.
197. Letley DP, Rhead JL, Twells RJ, Dove B, Atherton JC. Determinants of non-toxicity in the gastric pathogen Helicobacter pylori. Journal of Biological Chemistry. 2003;278(29):26734-41.
198. Wen S, Moss SF. Helicobacter pylori virulence factors in gastric carcinogenesis. Cancer letters. 2009;282(1):1-8.
199. Covacci A, Telford JL, Del Giudice G, Parsonnet J, Rappuoli R. Helicobacter pylori virulence and genetic geography. Science. 1999;284(5418):1328-33.
200. Yamaoka Y, Kodama T, Kita M, Imanishi J, Kashima K, Graham DY. Relationship of vacA genotypes of Helicobacter pylori to cagA status, cytotoxin production, and clinical outcome. Helicobacter. 1998;3(4):241-53.
201. Yamaoka Y, Kodama T, Gutierrez O, Kim JG, Kashima K, Graham DY. Relationship between Helicobacter pylori iceA, cagA, and vacA status and clinical outcome: studies in four different countries. Journal of clinical microbiology. 1999;37(7):2274-9.
202. Uchida T, Nguyen LT, Takayama A, Okimoto T, Kodama M, Murakami K, et al. Analysis of virulence factors of Helicobacter pylori isolated from a Vietnamese population. BMC microbiology. 2009;9(1):175.

203. Sahara S, Sugimoto M, Vilaichone R-K, Mahachai V, Miyajima H, Furuta T, et al. Role of Helicobacter pylori cagA EPIYA motif and vacA genotypes for the development of gastrointestinal diseases in Southeast Asian countries: a meta-analysis. BMC infectious diseases. 2012;12(1):223.
204. González-Rivera C, Algood HMS, Radin JN, McClain MS, Cover TL. The intermediate region of Helicobacter pylori VacA is a determinant of toxin potency in a Jurkat T cell assay. Infection and immunity. 2012:IAI-00052.
205. Yordanov D, Boyanova L, Markovska R, Gergova G, Mitov I. Significance of Helicobacter pylori vacA intermediate region genotyping—a Bulgarian study. Diagnostic microbiology and infectious disease. 2012;74(3):253-7.
206. Rhead JL, Letley DP, Mohammadi M, Hussein N, Mohagheghi MA, Hosseini ME, et al. A new Helicobacter pylori vacuolating cytotoxin determinant, the intermediate region, is associated with gastric cancer. Gastroenterology. 2007;133(3):926-36.
207. Chung C, Olivares A, Torres E, Yilmaz O, Cohen H, Perez-Perez G. Diversity of VacA intermediate region among Helicobacter pylori strains from several regions of the world. Journal of clinical microbiology. 2010;48(3):690-6.
208. Hussein NR, Mohammadi M, Talebkhan Y, Doraghi M, Letley DP, Muhammad MK, et al. Differences in virulence markers between Helicobacter pylori strains from Iraq and those from Iran: potential importance of regional differences in H. pylori-associated disease. Journal of clinical microbiology. 2008;46(5):1774-9.
209. Ogiwara H, Sugimoto M, Ohno T, Vilaichone R-K, Mahachai V, Graham DY, et al. Role of deletion located between the intermediate and middle regions of the Helicobacter pylori vacA gene in cases of gastroduodenal diseases. Journal of clinical microbiology. 2009;47(11):3493-500.
210. Hawn TR, Misch EA, Dunstan SJ, Thwaites GE, Lan NTN, Quy HT, et al. A common human TLR1 polymorphism regulates the innate immune response to lipopeptides. European journal of immunology. 2007;37(8):2280-9.
211. Buwitt-Beckmann U, Heine H, Wiesmüller K-H, Jung G, Brock R, Akira S, et al. TLR1-and TLR6-independent recognition of bacterial lipoeptides. Journal of Biological Chemistry. 2006.
212. Takeuchi O, Sato S, Horiuchi T, Hoshino K, Takeda K, Dong Z, et al. Cutting edge: role of Toll-like receptor 1 in mediating immune response to microbial lipoproteins. The Journal of Immunology. 2002;169(1):10-4.
213. El-Omar EM. Helicobacter pylori susceptibility in the GWAS era. JAMA. 2013;309(18):1939-40.
214. Van Doorn LJ, Figueiredo C, Mégraud F, Pena S, Midolo P, Queiroz DMDM, et al. Geographic distribution of vacA allelic types of Helicobacter pylori. Gastroenterology. 1999;116(4):823-30.

215. Miwa H, Go MF, Sato N. H. pylori and gastric cancer: the Asian enigma. The American journal of gastroenterology. 2002;97(5):1106.
216. Chen PE, Shapiro BJ. The advent of genome-wide association studies for bacteria. Current opinion in microbiology. 2015;25:17-24.
217. Read TD, Massey RC. Characterizing the genetic basis of bacterial phenotypes using genome-wide association studies: a new direction for bacteriology. Genome medicine. 2014;6(11):109.
218. Taylor NS, Fox JG, Akopyants NS, Berg DE, Thompson N, Shames B, et al. Long-term colonization with single and multiple strains of Helicobacter pylori assessed by DNA fingerprinting. Journal of Clinical Microbiology. 1995;33(4):918-23.
219. Jorgensen M, Daskalopoulos G, Warburton V, Mitchell HM, Hazell SL. Multiple strain colonization and metronidazole resistance in Helicobacter pylori-infected patients: identification from sequential and multiple biopsy specimens. Journal of Infectious Diseases. 1996;174(3):631-5.
220. Jepsen P, Johnsen SP, Gillman MW, Sørensen HT. Interpretation of observational studies. Heart. 2004;90(8):956-60.
221. Ponzetto A, La Rovere MT, Sanseverino P, Bazzoli F. Association of Helicobacter pylori infection with coronary heart disease. Study confirms previous findings. BMJ: British Medical Journal. 1996;312(7025):251.
222. Aceti A, Mazzacurati G, Amendolea M, Pennica A, Zechini B, Trappolini M, et al. Relation of C reactive protein to cardiovascular risk factors. H pylori and C pneumoniae infections may account for most acute coronary syndromes. BMJ: British Medical Journal. 1996;313(7054):428.
223. Peto R. Two properties of multiple regression analysis and regression to the mean (and regression from the mean). The natural history of chronic bronchitis and emphysema: an eight year study of early chronic obstructive lung disease in working men in London Oxford University Press, Oxford. 1976:218-23.
224. Hutcheon JA, Chiolero A, Hanley JA. Random measurement error and regression dilution bias. Bmj. 2010;340:c2289.
225. Sibbald B, Roland M. Understanding controlled trials. Why are randomised controlled trials important? BMJ: British Medical Journal. 1998;316(7126):201.
226. Bennett DA, Holmes MV. Mendelian randomisation in cardiovascular research: an introduction for clinicians. Heart. 2017:heartjnl-2016.
227. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Statistics in medicine. 2008;27(8):1133-63.

228. Nitsch D, Molokhia M, Smeeth L, DeStavola BL, Whittaker JC, Leon DA. Limits to causal inference based on Mendelian randomization: a comparison with randomized controlled trials. American journal of epidemiology. 2006;163(5):397-403.
229. Zheng J, Baird D, Borges M-C, Bowden J, Hemani G, Haycock P, et al. Recent developments in Mendelian randomization studies. Current epidemiology reports. 2017;4(4):330-45.
230. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? International journal of epidemiology. 2003;32(1):1-22.
231. Davey Smith G, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. International journal of epidemiology. 2004;33(1):30-42.
232. Corbin LJ, Richmond RC, Wade KH, Burgess S, Bowden J, Smith GD, et al. Body mass index as a modifiable risk factor for type 2 diabetes: Refining and understanding causal estimates using Mendelian randomisation. Diabetes. 2016:db160418.
233. Xu L, Borges MC, Hemani G, Lawlor DA. The role of glycaemic and lipid risk factors in mediating the effect of BMI on coronary heart disease: a two-step, two-sample Mendelian randomisation study. Diabetologia. 2017;60(11):2210-20.
234. Noyce AJ, Kia DA, Hemani G, Nicolas A, Price TR, De Pablo-Fernandez E, et al. Estimating the causal influence of body mass index on risk of Parkinson disease: A Mendelian randomisation study. PLoS medicine. 2017;14(6):e1002314.
235. Lawlor DA. Commentary: Two-sample Mendelian randomization: opportunities and challenges. International journal of epidemiology. 2016;45(3):908.
236. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International journal of epidemiology. 2015;44(2):512-25.
237. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Human molecular genetics. 2014;23(R1):R89-R98.
238. Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. International journal of epidemiology. 2014;43(3):922-9.
239. Gao C, Patel CJ, Michailidou K, Peters U, Gong J, Schildkraut J, et al. Mendelian randomization study of adiposity-related traits and risk of breast, ovarian, prostate, lung and colorectal cancer. International Journal of Epidemiology. 2016;45(3):896-908.
240. Taylor AE, Davies NM, Ware JJ, VanderWeele T, Smith GD, Munafò MR. Mendelian randomization in health research: using appropriate genetic variants and avoiding biased estimates. Economics & Human Biology. 2014;13:99-106.

241. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. Genetic epidemiology. 2016;40(7):597-608.
242. Welsh P, Polisecki E, Robertson M, Jahn S, Buckley BM, de Craen AJM, et al. Unraveling the directional link between adiposity and inflammation: a bidirectional Mendelian randomization approach. The Journal of Clinical Endocrinology & Metabolism. 2010;95(1):93-9.
243. Timpson NJ, Nordestgaard BG, Harbord RM, Zacho J, Frayling TM, Tybjærg-Hansen A, et al. C-reactive protein levels and body mass index: elucidating direction of causation through reciprocal Mendelian randomization. International journal of obesity. 2011;35(2):300.
244. Bochud M, Marquant F, Marques-Vidal P-M, Vollenweider P, Beckmann JS, Mooser V, et al. Association between C-reactive protein and adiposity in women. The Journal of Clinical Endocrinology & Metabolism. 2009;94(10):3969-77.
245. Greenfield JR, Samaras K, Jenkins AB, Kelly PJ, Spector TD, Gallimore JR, et al. Obesity is an important determinant of baseline serum C-reactive protein concentration in monozygotic twins, independent of genetic influences. Circulation. 2004;109(24):3022-8.
246. Calabro P, Chang DW, Willerson JT, Yeh ETH. Release of C-reactive protein in response to inflammatory cytokines by human adipocytes: linking obesity to vascular inflammation. Journal of the American College of Cardiology. 2005;46(6):1112-3.
247. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genetic epidemiology. 2013;37(7):658-65.
248. Bowden J, Spiller W, Del Greco M F, Sheehan N, Thompson J, Minelli C, et al. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. International journal of epidemiology. 2018;47(4):1264-78.
249. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. European journal of epidemiology. 2017;32(5):377-89.
250. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. Statistics in medicine. 2017;36(11):1783-802.
251. Greco M FD, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. Statistics in medicine. 2015;34(21):2926-40.
252. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I 2 statistic. International journal of epidemiology. 2016;45(6):1961-74.

253. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. Human molecular genetics. 2018.
254. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genetic epidemiology. 2016;40(4):304-14.
255. Hemani G, Bowden J, Haycock PC, Zheng J, Davis O, Flach P, et al. Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. bioRxiv. 2017:173682.
256. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. International journal of epidemiology. 2017;46(6):1985-98.
257. Hemani G, Zheng J, Wade KH, Laurin C, Elsworth B, Burgess S, et al. MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations. BioRxiv. 2016:078972.
258. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife. 2018;7:e34408.
259. Hartwig FP, Davies NM, Hemani G, Davey Smith G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. Oxford University Press; 2016.
260. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics. 2017;33(2):272-9.
261. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. Nature genetics. 2015;47(11):1236.
262. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature genetics. 2015;47(3):291.
263. Robinson EB, St Pourcain B, Anttila V, Kosmicki JA, Bulik-Sullivan B, Grove J, et al. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. Nature genetics. 2016;48(5):552.
264. Barban N, Jansen R, De Vlaming R, Vaez A, Mandemakers JJ, Tropf FC, et al. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. Nature genetics. 2016;48(12):1462.

265. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort profile: the 'children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. International journal of epidemiology. 2013;42(1):111-27.
266. Golding J, Pembrey M, Jones R. ALSPAC–the avon longitudinal study of parents and children. Paediatric and perinatal epidemiology. 2001;15(1):74-87.
267. Pearson H. Coming of age. Nature. 2012;484(7393):155.
268. Mitchell RE, Jones HJ, Yolken RH, Ford G, Jones-Brando L, Ring SM, et al. Longitudinal serological measures of common infection in the Avon Longitudinal Study of Parents and Children cohort. Wellcome Open Research. 2018;3.
269. Group CSC. Caerphilly and Speedwell collaborative heart disease studies. Journal of Epidemiology and Community Health (1979-). 1984:259-62.
270. Davey Smith G, Frankel S, Yarnell J. Sex and death: are they related? Findings from the Caerphilly cohort study. Bmj. 1997;315(7123):1641-4.
271. Dickerson FB, Boronow JJ, Stallings C, Origoni AE, Ruslanova I, Yolken RH. Association of serum antibodies to herpes simplex virus 1 with cognitive deficits in individuals with schizophrenia. Archives of general psychiatry. 2003;60(5):466-72.
272. Stone AFM, Risley P, Markus HS, Butland BK, Strachan DP, Elwood PC, et al. Ischaemic heart disease and Cag A strains ofHelicobacter pylori in the Caerphilly heart disease study. Heart. 2001;86(5):506-9.
273. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics. 2007;81(3):559-75.
274. International HapMap C. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449(7164):851.
275. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genetic epidemiology. 2010;34(8):816-34.
276. Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. PLoS genetics. 2012;8(8):e1002793.
277. Siva N. 1000 Genomes project. Nature Publishing Group; 2008.
278. Shah T, Engmann J, Dale C, Shah S, White J, Giambartolomei C, et al. Population genomics of cardiometabolic traits: design of the University College London-London School of Hygiene and Tropical Medicine-Edinburgh-Bristol (UCLEB) Consortium. PloS one. 2013;8(8):e71345.

279. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-1.
280. Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, et al. Genetic determinants of the gut microbiome in UK twins. Cell host & microbe. 2016;19(5):731-43.
281. Davenport ER, Cusanovich DA, Michelini K, Barreiro LB, Ober C, Gilad Y. Genome-wide association studies of the human gut microbiota. PLoS One. 2015;10(11):e0140301.
282. Rehnberg-Laiho L, Salomaa A, Rautelin H, Koskela P, Sarna S, Kosunen TU. Accelerated decline in Helicobacter pylori seroprevalence rate during the screen and treat project in Vammala, Finland, as demonstrated in 29-to 45-year-old pregnant women. Apmis. 2004;112(1):34-8.
283. Völzke H, Alte D, Schmidt CO, Radke D, Lorbeer R, Friedrich N, et al. Cohort profile: the study of health in Pomerania. International journal of epidemiology. 2010;40(2):294-307.
284. Jürgens C, Völzke H, Tost F. Study of Health in Pomerania (SHIP-Trend). Der Ophthalmologe. 2014;111(5):443-7.
285. Hofman A, Breteler MMB, van Duijn CM, Krestin GP, Pols HA, Stricker BHC, et al. The Rotterdam Study: objectives and design update. European journal of epidemiology. 2007;22(11):819-29.
286. Hofman A, van Duijn CM, Franco OH, Ikram MA, Janssen HLA, Klaver CCW, et al. The Rotterdam Study: 2012 objectives and design update. European journal of epidemiology. 2011;26(8):657.
287. Akira S, Takeda K, Kaisho T. Toll-like receptors: critical proteins linking innate and acquired immunity. Nature immunology. 2001;2(8):675.
288. Cheung N-KV, Sowers R, Vickers AJ, Cheung IY, Kushner BH, Gorlick R. FCGR2A polymorphism is correlated with clinical outcome after immunotherapy of neuroblastoma with anti-GD2 antibody and granulocyte macrophage colony-stimulating factor. Journal of clinical oncology. 2006;24(18):2885-90.
289. Burgess S, Zuber V, Valdes-Marquez E, Sun BB, Hopewell JC. Mendelian randomization with fine-mapped genetic data: Choosing from large numbers of correlated instrumental variables. Genetic epidemiology. 2017;41(8):714-25.
290. Zheng J, Richardson T, Millard L, Hemani G, Raistrick C, Vilhjalmsson B, et al. PhenoSpD: an atlas of phenotypic correlations and a multiple testing correction for the human phenome. bioRxiv. 2017:148627.
291. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. The American Journal of Human Genetics. 2004;74(4):765-9.
292. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity. 2005;95(3):221.

293. Elliott P, Peakman TC. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. International Journal of Epidemiology. 2008;37(2):234-44.
294. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS medicine. 2015;12(3):e1001779.
295. Wald A. The fitting of straight lines if both variables are subject to error. The Annals of Mathematical Statistics. 1940;11(3):284-300.
296. Kusters JG, van Vliet AHM, Kuipers EJ. Pathogenesis of Helicobacter pylori infection. Clinical microbiology reviews. 2006;19(3):449-90.
297. Graham DY. History of Helicobacter pylori, duodenal ulcer, gastric ulcer and gastric cancer. World Journal of Gastroenterology: WJG. 2014;20(18):5191.
298. Ford AC, Talley NJ. Does Helicobacter pylori really cause duodenal ulcers? Yes. BMJ: British Medical Journal (Online). 2009;339.
299. Katz HR. Inhibition of inflammatory responses by leukocyte Ig-like receptors. Advances in immunology. 2006;91:251-72.
300. Sloane DE, Tedla N, Awoniyi M, MacGlashan DW, Borges L, Austen KF, et al. Leukocyte immunoglobulin-like receptors: novel innate receptors for human basophil activation and inhibition. Blood. 2004;104(9):2832-9.
301. Huynh OA, Hampartzoumian T, Arm JP, Hunt J, Borges L, Ahern M, et al. Down-regulation of leucocyte immunoglobulin-like receptor expression in the synovium of rheumatoid arthritis patients after treatment with disease-modifying anti-rheumatic drugs. Rheumatology. 2007;46(5):742-51.
302. Tedla N, An H, Borges L, Vollmer-Conna U, Bryant K, Geczy C, et al. Expression of activating and inhibitory leukocyte immunoglobulin-like receptors in rheumatoid synovium: correlations to disease activity. Tissue antigens. 2011;77(4):305-16.
303. Liu J, Wang L, Gao W, Li L, Cui X, Yang H, et al. Inhibitory receptor immunoglobulin-like transcript 4 was highly expressed in primary ductal and lobular breast cancer and significantly correlated with IL-10. Diagnostic pathology. 2014;9(1):85.
304. Liu X, Yu X, Xie J, Zhan M, Yu Z, Xie L, et al. ANGPTL2/LILRB2 signaling promotes the propagation of lung cancer cells. Oncotarget. 2015;6(25):21004.
305. Deng M, Lu Z, Zheng J, Wan X, Chen X, Hirayasu K, et al. A motif in LILRB2 critical for Angptl2 binding and activation. Blood. 2014:blood-2014.

306. He J, Xu J, Yu X, Zhu H, Zeng Y, Fan D, et al. Overexpression of ANGPTL2 and LILRB2 as predictive and therapeutic biomarkers for metastasis and prognosis in colorectal cancer. Int J Clin Exp Pathol. 2018;11(5):2281-94.
307. Kang X, Kim J, Deng M, John S, Chen H, Wu G, et al. Inhibitory leukocyte immunoglobulin-like receptors: immune checkpoint proteins and tumor sustaining factors. Cell cycle. 2016;15(1):25-40.
308. Zheng J, Umikawa M, Cui C, Li J, Chen X, Zhang C, et al. Inhibitory receptors bind ANGPTLs and support blood stem cells and leukaemia development. Nature. 2012;485(7400):656.
309. Chen H-M, van der Touw W, Wang YS, Kang K, Mai S, Zhang J, et al. Blocking immunoinhibitory receptor LILRB2 reprograms tumor-associated myeloid cells and promotes antitumor immunity. The Journal of clinical investigation. 2018;128(12).
310. Brown DP, Jones DC, Anderson KJ, Lapaque N, Buerki RA, Trowsdale J, et al. The inhibitory receptor LILRB4 (ILT3) modulates antigen presenting cell phenotype and, along with LILRB2 (ILT4), is upregulated in response to Salmonella infection. BMC immunology. 2009;10(1):56.
311. InanlooRahatloo K, Parsa AFZ, Huse K, Rasooli P, Davaran S, Platzer M, et al. Mutation in ST6GALNAC5 identified in family with coronary artery disease. Scientific reports. 2014;4:3595.
312. Hebbar P, Nizam R, Melhem M, Alkayal F, Elkum N, John SE, et al. Genome-wide association study identifies novel recessive genetic variants for high TGs in an Arab population. Journal of lipid research. 2018;59(10):1951-66.
313. Bos PD, Zhang XHF, Nadal C, Shu W, Gomis RR, Nguyen DX, et al. Genes that mediate breast cancer metastasis to the brain. Nature. 2009;459(7249):1005.
314. Varki A. Sialic acids as ligands in recognition phenomena. The FASEB Journal. 1997;11(4):248-55.
315. Varki A. Sialic acids in human health and disease. Trends in molecular medicine. 2008;14(8):351-60.
316. Schauer R. Sialic acids as regulators of molecular and cellular interactions. Current opinion in structural biology. 2009;19(5):507-14.
317. Gopaul KP, Crook MA. Sialic acid: a novel marker of cardiovascular disease? Clinical biochemistry. 2006;39(7):667-81.
318. Nigam PK, Narain VS, Kumar A. Sialic acid in cardiovascular diseases. Indian Journal of Clinical Biochemistry. 2006;21(1):54-61.
319. Altay M, Karakoç MA, Çakır N, Yılmaz Demirtaş C, Cerit ET, Aktürk M, et al. Serum Total Sialic Acid Level is Elevated in Hypothyroid Patients as an Atherosclerotic Risk Factor. Journal of clinical laboratory analysis. 2017;31(2):e22034.

320. Wakabayashi I, Sakamoto K, Yoshimoto S, Kakishita E. Serum sialic acid concentration and atherosclerotic risk factors. Journal of atherosclerosis and thrombosis. 1994;1(2):113-7.
321. Forman D, Newell DG, Fullerton F, Yarnell JW, Stacey AR, Wald N, et al. Association between infection with Helicobacter pylori and risk of gastric cancer: evidence from a prospective investigation. Bmj. 1991;302(6788):1302-5.
322. Crabtree JE, Mahony MJ, Taylor JD, Heatley RV, Littlewood JM, Tompkins DS. Immune responses to Helicobacter pylori in children with recurrent abdominal pain. Journal of clinical pathology. 1991;44(9):768-71.
323. Graham DY, Genta RM, Graham DP, Crabtree JE. Serum CagA antibodies in asymptomatic subjects and patients with peptic ulcer: lack of correlation of IgG antibody in patients with peptic ulcer or asymptomatic Helicobacter pylori gastritis. Journal of clinical pathology. 1996;49(10):829-32.
324. Parsonnet J, Friedman GD, Orentreich N, Vogelman H. Risk for gastric cancer in people with CagA positive or CagA negative Helicobacter pylori infection. Gut. 1997;40(3):297-301.
325. Chen Y, Blaser MJ. Helicobacter pylori colonization is inversely associated with childhood asthma. The Journal of infectious diseases. 2008;198(4):553-60.
326. Amedei A, Codolo G, Del Prete G, de Bernard M, D'Elios MM. The effect of Helicobacter pylori on asthma and allergy. Journal of asthma and allergy. 2010;3:139.
327. Lebwohl B, Blaser MJ, Ludvigsson JF, Green PHR, Rundle A, Sonnenberg A, et al. Decreased risk of celiac disease in patients with Helicobacter pylori colonization. American journal of epidemiology. 2013;178(12):1721-30.
328. Reibman J, Marmor M, Filner J, Fernandez-Beros M-E, Rogers L, Perez-Perez GI, et al. Asthma is inversely associated with Helicobacter pylori status in an urban population. PloS one. 2008;3(12):e4060.
329. Chey WD, Wong BCY. American College of Gastroenterology guideline on the management of Helicobacter pylori infection. The American journal of gastroenterology. 2007;102(8):1808.
330. Ansari S, Yamaoka Y. Current understanding and management of Helicobacter pylori infection: an updated appraisal. F1000Research. 2018;7.
331. Ford AC, Forman D, Hunt RH, Yuan Y, Moayyedi P. Helicobacter pylori eradication therapy to prevent gastric cancer in healthy asymptomatic infected individuals: systematic review and meta-analysis of randomised controlled trials. Bmj. 2014;348:g3174.
332. Wong BC-Y, Lam SK, Wong WM, Chen JS, Zheng TT, Feng RE, et al. Helicobacter pylori eradication to prevent gastric cancer in a high-risk region of China: a randomized controlled trial. Jama. 2004;291(2):187-94.

Appendix

Table 1. LD Score correlation results using the LD Hub web interface and centralised database for disease outcomes. The analysis used my *H.pylori* GWAS meta-analysis result and summary-level GWAS statistics for disease outcomes of interest (i.e. cardiometabolic, cancer and anthropometric traits).

Trait 1	Trait 2	rg	se	z	p
H.pylori	Fasting proinsulin	-0.390	0.510	-0.764	0.445
H.pylori	Lung cancer	-0.379	0.460	-0.823	0.410
H.pylori	Body fat	-0.339	0.306	-1.108	0.268
H.pylori	Child birth length	-0.307	0.403	-0.764	0.445
H.pylori	Lung cancer (all)	-0.249	0.463	-0.537	0.591
H.pylori	Lung cancer (squamous cell)	-0.202	0.631	-0.320	0.749
H.pylori	Lung adenocarcinoma	-0.182	0.573	-0.318	0.750
H.pylori	Squamous cell lung cancer	-0.147	0.435	-0.339	0.735
H.pylori	HbA1C	-0.117	0.334	-0.349	0.727
H.pylori	HOMA-IR	-0.109	0.381	-0.285	0.776
H.pylori	Childhood obesity	-0.030	0.220	-0.136	0.892
H.pylori	Waist-to-hip ratio	-0.024	0.184	-0.130	0.897
H.pylori	Adiponectin	0.011	0.626	0.018	0.986
H.pylori	Fasting glucose main effect	0.033	0.265	0.125	0.900
H.pylori	Child birth weight	0.039	0.354	0.110	0.912
H.pylori	Body mass index	0.082	0.193	0.423	0.672
H.pylori	Difference in height between adolescence and adulthood; age 14	0.088	0.565	0.156	0.876
H.pylori	Waist circumference	0.100	0.202	0.495	0.621

H.pylori	Type 2 Diabetes	0.102	0.307	0.334	0.739
H.pylori	Hip circumference	0.118	0.193	0.610	0.542
H.pylori	Fasting insulin main effect	0.118	0.326	0.362	0.717
H.pylori	Birth weight	0.166	0.289	0.574	0.566
H.pylori	HOMA-B	0.174	0.299	0.582	0.561
H.pylori	Obesity class 1	0.183	0.395	0.464	0.642
H.pylori	Coronary artery disease	0.213	0.285	0.746	0.456
H.pylori	Height_2010	0.222	0.237	0.934	0.350
H.pylori	Overweight	0.282	0.390	0.723	0.470
H.pylori	Infant head circumference	0.326	0.501	0.651	0.515
H.pylori	Difference in height between childhood and adulthood; age 8	0.342	0.530	0.646	0.518
H.pylori	2hr glucose adjusted for BMI	0.570	0.671	0.849	0.396
H.pylori	Height; Females at age 10 and males at age 12	0.610	0.940	0.648	0.517

Table 2. List of outcome traits from MR-Base used in two-sample MR analysis

Consortium	Number of SNPs	Population	Sample size	Sex	Trait
ADIPOGen	2,675,209	Mixed	39,883	M/F	Adiponectin
BioBank Japan Project	2,178,019	Japanese	10,112	M/F	C-reactive protein
MESA	2,390,490	European	2,431	M/F	Percent emphysema
HRgene consortium	2,516,790	Mixed	92,355	M/F	Heart rate
NA	425,707	East Asian	907	M/F	Gallbladder cancer
GIANT	2,760,790	European	60,586	M	Weight

EGG	2,442,739	European	13,848	M/F	Childhood obesity
ISGC	2,421,920	Mixed	29,633	M/F	Ischemic stroke
ISGC	2,421,920	Mixed	21,185	M/F	Cardioembolic stroke
ISGC	2,421,920	Mixed	21,143	M/F	Large vessel disease
ISGC	2,421,920	Mixed	20,675	M/F	Small vessel disease
DIAGRAM	2,915,012	Mixed	110,452	M/F	Type 2 diabetes
GLGC	2,447,442	Mixed	187,167	M/F	HDL cholesterol
GLGC	2,437,752	Mixed	173,082	M/F	LDL cholesterol
GLGC	2,439,433	Mixed	177,861	M/F	Triglycerides
GIANT	2,547,573	Mixed	245,746	M/F	Waist circumference
CARDIoGRAMplusC4D	9,455,779	Mixed	184,305	M/F	Coronary heart disease
GIANT	2,562,516	Mixed	224,459	M/F	Waist-to-hip ratio
MAGIC	2,401,709	European	15,234	M/F	2hr glucose
MAGIC	2,576,680	European	46,368	M/F	HbA1C
MAGIC	2,434,142	European	4,213	M/F	AUCins/AUCglu
MAGIC	2,433,997	European	4,324	M/F	AUCins
MAGIC	2,425,234	European	5,318	M/F	Corrected insulin response
MAGIC	2,426,095	European	5,130	M/F	Insulin disposition index
MAGIC	2,427,303	European	4,447	M/F	Incremental insulin at 30 minutes
MAGIC	2,432,136	European	4,409	M/F	Insulin at 30 minutes
MAGIC	2,418,000	European	4,769	M/F	Insulin sensitivity index
MAGIC	2,496,074	European	10,701	M/F	Fasting proinsulin
MAGIC	2,456,946	European	46,186	M/F	HOMA-B
MAGIC	2,458,074	European	46,186	M/F	HOMA-IR

MAGIC	2.635.762	European	58,074	M/F	Fasting glucose
MAGIC	2,634,889	European	51,750	M/F	Fasting insulin
MAGIC	63,984	European	42,854	M/F	2hr glucose
MDACC	818,978	European	2,830	M/F	Melanoma
CARDIoGRAMplusC4D	9,289,492	Mixed	171,875	M/F	Myocardial infarction
NA	468,788	European	4,881	M/F	Neuroblastoma
PanScan1	521,863	European	3,835	M/F	Pancreatic cancer
NA	11,760,646	European	20,687	M/F	Apolipoprotein A-I
NA	11,813,266	European	20,690	M/F	Apolipoprotein B
GIANT	19,848,14	European	16,068	M/F	Extreme body mass index
GIANT	1,939,901	European	10,255	M/F	Extreme waist-to-hip ratio
GIANT	2,331,456	European	72,546	M/F	Obesity class 2
GIANT	2,250,779	European	50,364	M/F	Obesity class 3
GIANT	2,435,045	European	158,855	M/F	Overweight
ILCCO	8,881,354	European	18,336	M/F	Lung adenocarcinoma
ILCCO	8,945,893	European	27,209	M/F	Lung cancer
ILCCO	8,893,750	European	18,313	M/F	Squamous cell lung cancer
NA	3,228,665	European	100,716	M/F	Body fat
Neale Lab	10,894,596	European	337,159	M/F	Non-cancer illness code self-reported: high cholesterol
Neale Lab	10,894,596	European	337,030	M/F	Current tobacco smoking
Neale Lab	10,894,596	European	337,159	M/F	Treatment/medication code: warfarin
Neale Lab	10,894,596	European	310,749	M/F	Past tobacco smoking
Neale Lab	10,894,596	European	305,723	M/F	Exposure to tobacco smoke at home

Neale Lab	10,894,596	European	292,053	M/F	Illnesses of father: Heart disease
Neale Lab	10,894,596	European	292,053	M/F	Illnesses of father: Prostate cancer
Neale Lab	10,894,596	European	292,053	M/F	Illnesses of father: Lung cancer
Neale Lab	10,894,596	European	292,053	M/F	Illnesses of father: Chronic bronchitis/emphysema
Neale Lab	10,894,596	European	292,053	M/F	Illnesses of father: High blood pressure
Neale Lab	10,894,596	European	292,053	M/F	Illnesses of father: Diabetes
Neale Lab	10,894,596	European	308,780	M/F	Illnesses of mother: Heart disease
Neale Lab	10,894,596	European	337,154	M/F	Number of self-reported cancers
Neale Lab	10,894,596	European	308,780	M/F	Illnesses of mother: Breast cancer
Neale Lab	10,894,596	European	308,780	M/F	Illnesses of mother: Chronic bronchitis/emphysema
Neale Lab	10,894,596	European	308,780	M/F	Illnesses of mother: High blood pressure
Neale Lab	10,894,596	European	308,780	M/F	Illnesses of mother: Diabetes
Neale Lab	10,894,596	European	259,921	M/F	Illnesses of siblings: Heart disease
Neale Lab	10,894,596	European	259,921	M/F	Illnesses of siblings: Stroke
Neale Lab	10,894,596	European	259,921	M/F	Illnesses of siblings: High blood pressure
Neale Lab	10,894,596	European	259,921	M/F	Illnesses of siblings: Diabetes
Neale Lab	10,894,596	European	336,024	M/F	Smoking status: Previous
Neale Lab	10,894,596	European	336,067	M/F	Ever smoked
Neale Lab	10,894,596	European	101,726	M/F	Pack years of smoking
Neale Lab	10,894,596	European	336,107	M/F	Body mass index (BMI)
Neale Lab	10,894,596	European	83,133	M/F	Tobacco smoking: Ex-smoker
Neale Lab	10,894,596	European	331,117	M/F	Body fat percentage

Neale Lab	10,894,596	European	330,762	M/F	Whole body fat mass
Neale Lab	10,894,596	European	331,291	M/F	Whole body fat-free mass
Neale Lab	10,894,596	European	331,315	M/F	Whole body water mass
Neale Lab	10,894,596	European	23,205	M/F	Number of cigarettes currently smoked daily (current cigarette smokers)
Neale Lab	10,894,596	European	317,756	M/F	Diastolic blood pressure automated reading
Neale Lab	10,894,596	European	317,754	M/F	Systolic blood pressure automated reading
Neale Lab	10,894,596	European	336,601	M/F	Hip circumference
Neale Lab	10,894,596	European	336,683	M/F	Vascular/heart problems diagnosed by doctor: High blood pressure
Neale Lab	10,894,596	European	336,782	M/F	Blood clot DVT bronchitis emphysema asthma rhinitis eczema allergy diagnosed by doctor: Emphysema/chronic bronchitis
Neale Lab	10,894,596	European	180,203	M/F	Medication for cholesterol blood pressure diabetes or take exogenous hormones: Cholesterol lowering medication
Neale Lab	10,894,596	European	180,203	M/F	Medication for cholesterol blood pressure diabetes or take exogenous hormones: None of the above
Neale Lab	10,894,596	European	180,203	M/F	Medication for cholesterol blood pressure diabetes or take exogenous hormones: Blood pressure medication
Neale Lab	10,894,596	European	180,203	M/F	Medication for cholesterol blood pressure diabetes or take exogenous hormones: Hormone replacement therapy

Neale Lab	10,894,596	European	154,702	M/F	Medication for cholesterol blood pressure or diabetes: Cholesterol lowering medication
Neale Lab	10,894,596	European	154,702	M/F	Medication for cholesterol blood pressure or diabetes: None of the above
Neale Lab	10,894,596	European	154,702	M/F	Medication for cholesterol blood pressure or diabetes: Blood pressure medication
Neale Lab	10,894,596	European	154,702	M/F	Medication for cholesterol blood pressure or diabetes: Insulin
Neale Lab	10894596	European	337,159	M/F	Non-cancer illness code self-reported: emphysema/chronic bronchitis

Table 3. *H.pylori* genetic instruments for bidirectional MR analysis

Phenotype	SNP	β	se	eaf	EA	OA
LDL cholesterol (N = 173,082)	rs2419604	0.030	0.004	0.318	A	G
	rs646776	0.160	0.004	0.788	T	C
	rs10893499	0.052	0.005	0.144	A	G
	rs10832962	0.032	0.004	0.719	T	C
	rs267733	0.033	0.005	0.863	A	G
	rs174583	0.052	0.004	0.625	C	T
	rs3184504	0.027	0.004	0.534	C	T
	rs1169288	0.038	0.004	0.334	C	A
	rs2642438	0.035	0.004	0.745	G	A
	rs2587534	0.039	0.004	0.528	A	G

	rs10903129	0.033	0.004	0.537	G	A
	rs12748152	0.050	0.007	0.071	T	C
	rs4942486	0.024	0.004	0.462	T	C
	rs8017377	0.030	0.004	0.459	A	G
	rs2495495	0.034	0.006	0.135	T	C
	rs11591147	0.497	0.018	0.983	G	T
	rs7551981	0.047	0.004	0.595	T	G
	rs12066643	0.039	0.006	0.881	C	T
	rs7534572	0.041	0.006	0.690	G	C
	rs247616	0.055	0.004	0.707	C	T
	rs2000999	0.065	0.005	0.185	A	G
	rs6504872	0.027	0.004	0.472	T	C
	rs1801689	0.103	0.014	0.037	C	A
	rs2886232	0.045	0.006	0.120	T	C
	rs314253	0.024	0.004	0.665	T	C
	rs6511720	0.221	0.006	0.902	G	T
	rs2738459	0.053	0.006	0.555	A	C
	rs2228603	0.104	0.007	0.929	C	T
	rs4970712	0.034	0.004	0.806	C	A
	rs2965157	0.189	0.011	0.979	T	C
	rs7254892	0.485	0.012	0.968	G	A
	rs75687619	0.174	0.016	0.024	T	G
	rs12721109	0.446	0.018	0.983	G	A
	rs676388	0.027	0.004	0.463	C	T

	rs364585	0.025	0.004	0.633	G	A
	rs2328223	0.030	0.005	0.249	C	A
	rs6016373	0.035	0.004	0.627	A	G
	rs6065311	0.042	0.004	0.460	C	T
	rs1800961	0.069	0.011	0.966	C	T
	rs10490626	0.051	0.007	0.921	G	A
	rs2030746	0.021	0.004	0.398	T	C
	rs16831243	0.038	0.006	0.181	T	C
	rs10195252	0.024	0.004	0.582	T	C
	rs1367117	0.119	0.004	0.288	A	G
	rs72902576	0.093	0.013	0.963	T	G
	rs1250229	0.024	0.004	0.789	C	T
	rs5763662	0.077	0.012	0.025	T	C
	rs11563251	0.035	0.006	0.125	T	C
	rs4253776	0.031	0.006	0.124	G	A
	rs6544713	0.081	0.004	0.294	T	C
	rs6709904	0.055	0.009	0.887	A	G
	rs2710642	0.024	0.004	0.619	A	G
	rs9875338	0.027	0.004	0.612	G	A
	rs17404153	0.034	0.005	0.856	G	T
	rs7640978	0.039	0.007	0.895	C	T
	rs6818397	0.022	0.004	0.413	T	G
	rs4530754	0.028	0.004	0.582	A	G
	rs6882076	0.046	0.004	0.666	C	T

	rs12916	0.073	0.004	0.431	C	T
	rs6909746	0.026	0.004	0.608	C	T
	rs112201728	0.068	0.010	0.058	T	C
	rs1564348	0.048	0.005	0.145	C	T
	rs16891156	0.097	0.017	0.018	C	A
	rs2315065	0.110	0.016	0.087	A	C
	rs3757354	0.038	0.004	0.790	C	T
	rs13206249	0.038	0.006	0.784	G	A
	rs1408272	0.052	0.008	0.947	T	G
	rs10947332	0.050	0.006	0.132	A	G
	rs2390536	0.022	0.004	0.368	A	G
	rs4722551	0.039	0.005	0.170	C	T
	rs2073547	0.049	0.005	0.194	G	A
	rs2737252	0.031	0.004	0.744	G	A
	rs2954029	0.056	0.004	0.532	A	T
	rs7832643	0.034	0.004	0.405	T	G
	rs13277801	0.034	0.004	0.347	C	T
	rs9987289	0.071	0.007	0.925	G	A
	rs1883025	0.030	0.004	0.757	C	T
	rs579459	0.067	0.005	0.215	C	T
	rs3780181	0.045	0.007	0.947	A	G
	rs964184	0.086	0.008	1.000	G	C
Hip circumference (N = 336,601)	rs9378684	0.020	0.003	0.201	T	C
	rs4467770	0.017	0.003	0.731	A	G

	rs76040172	-0.034	0.005	0.054	A	G
	rs10887571	0.015	0.002	0.453	T	C
	rs1182199	-0.028	0.003	0.305	A	C
	rs7442885	-0.020	0.003	0.210	G	C
	rs66679256	0.016	0.002	0.446	T	C
	rs55650227	-0.020	0.003	0.190	C	G
	rs1294437	0.019	0.003	0.355	T	C
	rs10144067	0.017	0.002	0.592	T	C
	rs1727901	0.019	0.003	0.735	T	C
	rs28479795	0.024	0.003	0.222	T	C
	rs4872142	-0.021	0.003	0.186	G	C
	rs41271299	0.041	0.005	0.052	T	C
	rs57636386	-0.037	0.004	0.083	C	T
	rs1449630	0.014	0.002	0.570	G	A
	rs72656010	-0.024	0.004	0.131	C	T
	rs12300276	0.016	0.003	0.237	A	G
	rs845084	0.018	0.003	0.258	A	G
	rs13389219	0.026	0.002	0.394	T	C
	rs34373881	-0.015	0.003	0.278	A	G
	rs12607512	0.013	0.002	0.448	G	A
	rs34517439	0.043	0.004	0.126	A	C
	rs252749	-0.023	0.003	0.247	A	G
	rs1127100	0.015	0.003	0.648	C	T
	rs6080646	-0.013	0.002	0.497	A	G

	rs7707394	-0.019	0.003	0.354	A	G
	rs543874	0.044	0.003	0.208	G	A
	rs2814943	0.055	0.003	0.140	A	G
	rs4776970	-0.020	0.003	0.357	T	A
	rs12619178	-0.016	0.002	0.403	T	C
	rs3807566	-0.016	0.002	0.438	T	G
	rs7982447	0.016	0.003	0.206	C	T
	rs7516554	0.016	0.002	0.400	T	C
	rs7116641	0.022	0.003	0.317	G	T
	rs62396185	-0.037	0.003	0.256	C	G
	rs34811474	-0.021	0.003	0.232	A	G
	rs2737250	-0.021	0.003	0.351	G	A
	rs9808900	0.027	0.003	0.199	T	G
	rs1955695	-0.019	0.002	0.625	G	A
	rs675162	0.018	0.002	0.481	G	A
	rs588660	0.019	0.002	0.586	A	G
	rs390192	-0.014	0.002	0.522	G	A
	rs3803286	-0.018	0.003	0.666	G	A
	rs35057083	0.014	0.003	0.687	T	C
	rs28366156	-0.030	0.004	0.131	C	T
	rs34769775	-0.017	0.003	0.298	T	C
	rs962554	-0.022	0.003	0.284	C	T
	rs13264909	-0.016	0.002	0.428	T	A
	rs113866544	0.033	0.005	0.068	C	T

	rs2129869	-0.021	0.003	0.218	T	A
	rs79969674	0.027	0.005	0.074	T	C
	rs2187449	0.016	0.003	0.767	A	G
	rs11109097	-0.016	0.003	0.697	T	C
	rs4660586	-0.020	0.003	0.738	T	C
	rs1014291	-0.016	0.002	0.426	T	G
	rs6567160	0.050	0.003	0.234	C	T
	rs40071	-0.017	0.003	0.179	C	T
	rs6707036	-0.015	0.003	0.338	G	A
	rs28377268	0.025	0.004	0.108	T	G
	rs894347	-0.018	0.002	0.395	G	A
	rs4430895	0.024	0.002	0.480	T	C
	rs6907872	0.015	0.003	0.302	T	C
	rs1428120	-0.014	0.002	0.574	T	G
	rs4482463	-0.030	0.005	0.924	A	C
	rs750090	-0.017	0.003	0.356	C	T
	rs4741546	-0.018	0.002	0.397	T	C
	rs11150461	-0.015	0.003	0.728	G	C
	rs2861690	-0.017	0.002	0.389	G	C
	rs2034768	-0.017	0.002	0.512	G	A
	rs2897968	0.014	0.002	0.605	A	G
	rs1296328	-0.016	0.002	0.560	C	A
	rs62425398	0.024	0.004	0.107	A	C
	rs6973656	0.019	0.002	0.397	G	A

	rs245775	0.015	0.003	0.729	G	A
	rs2244786	0.014	0.003	0.357	A	G
	rs1285997	0.023	0.003	0.709	G	C
	rs8192675	0.017	0.003	0.287	C	T
	rs2307111	-0.028	0.002	0.393	C	T
	rs4843158	0.020	0.003	0.683	C	G
	rs41284816	0.079	0.009	0.020	T	G
	rs12254441	-0.015	0.003	0.373	T	C
	rs12877270	0.015	0.002	0.438	A	G
	rs12519997	-0.014	0.002	0.560	A	G
	rs1528450	0.017	0.002	0.597	C	T
	rs10938397	0.023	0.002	0.434	G	A
	rs10236214	0.018	0.003	0.641	T	C
	rs2102278	0.016	0.003	0.322	G	A
	rs12140153	-0.025	0.004	0.097	T	G
	rs4777541	0.019	0.003	0.765	T	C
	rs4966012	0.015	0.003	0.677	G	C
	rs3746759	-0.017	0.003	0.202	G	T
	rs9788550	-0.021	0.003	0.248	C	G
	rs10118701	0.018	0.003	0.317	G	A
	rs2013002	0.018	0.002	0.587	C	T
	rs9512696	0.018	0.003	0.662	G	A
	rs8023263	0.015	0.002	0.529	T	G
	rs72892910	0.036	0.003	0.170	T	G

	rs1918249	0.016	0.003	0.760	A	T
	rs6470771	-0.018	0.003	0.169	C	A
	rs62246314	0.025	0.004	0.101	A	G
	rs10237317	0.014	0.002	0.417	G	A
	rs7145337	-0.016	0.003	0.710	T	C
	rs1118151	0.017	0.003	0.722	G	T
	rs6739755	-0.015	0.002	0.603	G	A
	rs7740107	-0.023	0.003	0.736	A	T
	rs4722398	0.022	0.004	0.136	T	C
	rs4240326	-0.028	0.002	0.550	G	A
	rs2371767	0.020	0.003	0.273	C	G
	rs12680342	-0.019	0.003	0.229	G	T
	rs113364497	-0.016	0.003	0.251	T	C
	rs10100245	0.020	0.002	0.566	A	G
	rs74749286	0.031	0.004	0.108	A	G
	rs1458156	0.013	0.002	0.488	T	C
	rs12209223	0.022	0.004	0.102	A	C
	rs55932154	-0.024	0.004	0.115	G	A
	rs6821305	0.015	0.002	0.397	C	A
	rs6840236	0.017	0.002	0.462	C	T
	rs6601527	-0.016	0.002	0.588	A	C
	rs10756798	-0.015	0.003	0.645	T	C
	rs12528644	0.023	0.003	0.283	A	C
	rs9323375	0.016	0.003	0.232	A	T

	rs2253310	0.021	0.002	0.628	G	C
	rs2318543	-0.018	0.003	0.782	G	A
	rs2236519	-0.021	0.002	0.377	A	G
	rs9415106	-0.015	0.003	0.715	A	G
	rs140201358	-0.064	0.010	0.014	G	C
	rs9967367	-0.016	0.003	0.295	T	C
	rs10103997	0.016	0.003	0.225	G	C
	rs62243489	-0.015	0.003	0.257	G	T
	rs11766945	-0.020	0.003	0.201	A	G
	rs6535240	0.014	0.003	0.313	G	A
	rs815335	0.018	0.003	0.371	T	C
	rs4670612	0.016	0.003	0.656	G	A
	rs1618069	0.016	0.003	0.750	G	A
	rs28418580	-0.017	0.002	0.460	T	C
	rs76798800	0.025	0.003	0.267	T	G
	rs756717	-0.016	0.002	0.399	A	G
	rs10269774	0.023	0.003	0.324	A	G
	rs73175572	0.026	0.004	0.112	G	A
	rs73213484	-0.020	0.003	0.139	T	A
	rs143384	0.028	0.002	0.403	G	A
	rs2494196	0.032	0.003	0.288	A	C
	rs7460093	0.015	0.002	0.536	A	G
	rs17770336	0.021	0.003	0.324	T	C
	rs8064502	-0.017	0.002	0.439	A	C

	rs2678204	0.022	0.003	0.342	G	T
	rs6539064	-0.019	0.003	0.253	G	C
	rs9814633	0.015	0.003	0.344	A	G
	rs3943933	0.013	0.002	0.480	A	T
	rs35779991	0.014	0.002	0.531	C	T
	rs11882409	0.018	0.003	0.286	A	C
	rs1477290	0.028	0.004	0.135	C	T
	rs982692	0.014	0.003	0.361	C	T
	rs11245480	0.016	0.003	0.325	G	C
	rs6867299	0.021	0.003	0.373	C	T
	rs13107325	0.043	0.005	0.075	T	C
	rs17024393	0.057	0.008	0.026	C	T
	rs7226064	-0.015	0.002	0.429	G	A
	rs7124681	0.020	0.002	0.408	A	C
	rs6669341	-0.015	0.002	0.582	G	A
	rs12714415	-0.049	0.003	0.167	C	T
	rs2479958	-0.015	0.002	0.515	G	A
	rs12096864	0.022	0.004	0.118	C	T
	rs55886426	-0.034	0.006	0.055	G	C
	rs968379	-0.021	0.003	0.231	T	C
	rs6999725	-0.040	0.007	0.033	T	C
	rs6585201	-0.021	0.002	0.454	A	G
	rs11862944	-0.016	0.003	0.355	C	A
	rs35874463	0.033	0.005	0.058	G	A

	rs11944291	-0.014	0.002	0.474	G	T
	rs6575340	0.020	0.003	0.637	A	G
	rs8133137	0.015	0.003	0.666	G	A
	rs11839227	-0.018	0.003	0.184	C	T
	rs4630170	-0.017	0.003	0.733	G	A
	rs1441264	0.017	0.002	0.592	A	G
	rs75543804	-0.041	0.007	0.033	T	G
	rs3845344	0.015	0.002	0.392	T	C
	rs7426945	0.016	0.002	0.546	G	A
	rs1662185	0.017	0.003	0.709	G	A
	rs75949361	0.044	0.007	0.033	T	C
	rs10883553	0.016	0.002	0.447	A	C
	rs138767	0.015	0.003	0.651	C	T
	rs35099456	-0.041	0.005	0.064	C	G
	rs56288810	0.018	0.003	0.214	G	A
	rs10777859	-0.015	0.002	0.529	G	A
	rs7845090	-0.023	0.003	0.711	A	G
	rs724016	0.026	0.002	0.447	G	A
	rs72801854	0.018	0.003	0.302	A	G
	rs13410783	0.015	0.002	0.368	G	A
	rs12765337	0.016	0.003	0.343	C	G
	rs10153248	-0.017	0.002	0.444	G	A
	rs72959041	-0.070	0.006	0.051	A	G
	rs11915747	-0.019	0.003	0.354	G	C

	rs2715439	0.016	0.002	0.545	T	C
	rs1396513	0.018	0.002	0.514	T	C
	rs77165542	-0.086	0.007	0.036	T	C
	rs17245511	-0.020	0.003	0.149	A	G
	rs7957774	-0.016	0.003	0.244	T	G
	rs10404726	-0.017	0.002	0.467	T	C
	rs56094641	0.061	0.002	0.403	G	A
	rs4794222	-0.017	0.003	0.744	G	A
	rs6142059	0.016	0.002	0.494	C	T
	rs12805742	-0.019	0.003	0.230	T	C
	rs4982753	-0.016	0.003	0.256	T	C
	rs10820852	-0.016	0.003	0.274	A	C
	rs12701265	0.015	0.002	0.395	A	G
	rs2270894	-0.020	0.003	0.205	G	C
	rs1293395	-0.027	0.005	0.076	T	G
	rs1964599	-0.023	0.003	0.344	T	C
	rs2499468	0.014	0.003	0.653	A	C
	rs9843653	0.021	0.002	0.515	C	T
	rs4297095	-0.022	0.004	0.108	A	G
	rs35882248	0.017	0.003	0.315	T	C
	rs7915723	-0.014	0.002	0.559	A	C
	rs1582931	-0.020	0.002	0.472	A	G
	rs879620	0.027	0.002	0.615	T	C
	rs7274811	-0.021	0.003	0.258	T	G

	rs12883788	0.015	0.002	0.459	T	C
	rs12779865	0.023	0.003	0.330	C	T
	rs2439823	0.019	0.002	0.548	G	A
	rs112646560	0.021	0.003	0.218	T	C
	rs8011368	-0.016	0.003	0.724	T	C
	rs7978353	-0.017	0.002	0.405	G	A
	rs10210468	-0.015	0.002	0.467	C	T
	rs8042404	0.017	0.003	0.271	A	G
	rs34049648	0.018	0.003	0.336	A	G
	rs2954021	0.018	0.002	0.506	G	A
	rs11030119	0.031	0.003	0.309	A	G
	rs998584	-0.021	0.002	0.482	A	C
	rs1979440	-0.015	0.002	0.402	C	T
	rs11584359	-0.021	0.003	0.175	T	C
	rs7548408	0.015	0.002	0.517	C	T
	rs7238896	0.021	0.003	0.141	G	A
	rs1320903	0.020	0.003	0.318	A	G
	rs147730268	-0.052	0.004	0.091	T	G
	rs731758	-0.018	0.002	0.616	G	C
	rs60984707	-0.018	0.003	0.183	T	C
	rs58551145	0.024	0.003	0.195	G	A
	rs3218036	0.019	0.003	0.327	A	G
	rs35880697	-0.026	0.003	0.167	T	C
	rs869400	0.020	0.003	0.816	G	T

	rs59738707	-0.023	0.004	0.128	A	G
	rs4402589	0.030	0.002	0.553	G	T
	rs1618725	-0.021	0.002	0.496	T	C
	rs60226453	0.018	0.003	0.175	T	C
	rs3811951	0.015	0.003	0.282	G	A
	rs7930275	0.019	0.003	0.221	T	C
	rs1569497	0.015	0.002	0.441	G	A
	rs9496567	0.022	0.003	0.242	A	G
	rs34629844	0.022	0.004	0.129	G	A
	rs1231281	-0.014	0.002	0.499	A	G
	rs12475388	-0.013	0.002	0.487	A	G
	rs4790292	-0.028	0.003	0.154	A	C
	rs3810291	0.023	0.003	0.677	A	G
	rs273505	0.015	0.002	0.420	C	T
	rs11664106	0.015	0.003	0.374	T	A
	rs4985407	0.015	0.002	0.497	G	A
	rs6501601	-0.019	0.002	0.382	A	G
	rs34013042	0.015	0.003	0.258	T	C
	rs12561919	0.020	0.003	0.148	T	C
	rs11803990	0.025	0.004	0.082	G	C
	rs34748838	0.020	0.002	0.490	T	C
	rs12467963	-0.014	0.002	0.403	T	A
	rs25849	0.020	0.003	0.287	G	C
	rs7893571	0.016	0.003	0.663	T	G

	rs55726687	0.021	0.003	0.211	A	G
	rs882378	0.015	0.003	0.309	C	A
	rs78470967	0.036	0.006	0.043	A	T
	rs12375196	0.017	0.002	0.424	A	C
	rs2802774	0.017	0.002	0.546	A	C
	rs3826408	0.015	0.002	0.457	T	C
	rs667515	-0.014	0.002	0.386	C	G
	rs12972720	0.018	0.002	0.535	C	G
	rs76895963	0.096	0.009	0.019	G	T
	rs33955687	-0.016	0.003	0.292	A	C
	rs13333747	-0.024	0.003	0.182	C	T
	rs62037365	0.032	0.002	0.402	G	C
	rs12920259	-0.016	0.002	0.612	A	G
	rs141622900	0.032	0.006	0.052	A	G
	rs2815753	0.022	0.002	0.599	A	G
	rs7132908	0.025	0.002	0.384	A	G
	rs10953513	-0.019	0.002	0.420	G	A
	rs62473743	-0.019	0.003	0.844	G	A
	rs58584712	0.020	0.003	0.210	A	G
Heart rate (N = 92,355)	rs1015451	-0.741	0.081	0.890	T	C
	rs11153730	0.393	0.051	0.515	T	C
	rs11578508	0.407	0.054	0.665	A	G
	rs13030174	0.337	0.059	0.726	A	C
	rs13245899	-0.520	0.065	0.802	A	G

	rs17287293	0.491	0.072	0.849	A	G
	rs17362588	0.779	0.090	0.116	A	G
	rs174549	0.394	0.055	0.318	A	G
	rs17796783	0.344	0.057	0.722	T	C
	rs2029213	-0.296	0.052	0.362	T	C
	rs365990	-0.606	0.055	0.658	A	G
	rs3729992	0.548	0.099	0.085	A	C
	rs4489968	0.529	0.068	0.833	T	G
	rs6127471	-0.509	0.052	0.457	T	C
	rs7980799	0.405	0.054	0.402	A	C
Maternal history of breast cancer	rs9397437	0.010	0.001	0.070	A	G
	rs4442975	-0.004	0.001	0.512	T	G
	rs7976725	-0.005	0.001	0.224	G	A
	rs78540526	0.012	0.001	0.071	T	C
	rs4784227	0.009	0.001	0.239	T	C
	rs1078806	0.008	0.001	0.404	G	A
	rs58952190	0.013	0.002	0.021	C	T
	rs1269867	-0.004	0.001	0.492	T	C
	rs10941679	0.005	0.001	0.255	G	A
	rs11879798	0.004	0.001	0.377	A	G
	rs11836367	-0.005	0.001	0.352	T	C

Table 4. Two-sample MR analysis results

Outcome	Exposure	Method	Number of SNPs	β	se	p	Lower 95% CI	Upper 95% CI
LDL cholesterol	H.pylori	Wald ratio	1	-0.113	0.029	0.0001	-0.170	-0.056
Hip circumference	H.pylori	Wald ratio	1	-0.032	0.008	0.0001	-0.049	-0.016
Illnesses of mother: Breast cancer	H.pylori	Wald ratio	1	-0.008	0.002	0.0005	-0.013	-0.004
Heart rate	H.pylori	IVW	2	-0.485	0.144	0.0007	-0.767	-0.203
Systolic blood pressure (automated reading)	H.pylori	Wald ratio	1	0.024	0.009	0.0061	0.007	0.040
HOMA-IR	H.pylori	IVW	2	-0.030	0.011	0.0068	-0.052	-0.008
Ischemic stroke	H.pylori	IVW	2	0.123	0.047	0.0084	0.032	0.215
Obesity class 3	H.pylori	IVW	2	-0.251	0.100	0.0125	-0.448	-0.054
HOMA-B	H.pylori	IVW	2	-0.022	0.009	0.0182	-0.041	-0.004
Illnesses of mother	H.pylori	Wald ratio	1	0.008	0.004	0.0270	0.001	0.015
Whole body fat mass	H.pylori	Wald ratio	1	-0.018	0.008	0.0275	-0.035	-0.002
Number of self-reported cancers	H.pylori	Wald ratio	1	-0.006	0.003	0.0380	-0.011	0.000
Obesity class 2	H.pylori	IVW	2	-0.105	0.053	0.0492	-0.209	0.000
Whole body water mass	H.pylori	Wald ratio	1	-0.011	0.005	0.0496	-0.021	0.000
Body fat percentage	H.pylori	Wald ratio	1	-0.013	0.007	0.0544	-0.025	0.000
Body mass index (BMI)	H.pylori	Wald ratio	1	-0.016	0.008	0.0598	-0.032	0.001
Whole body fat-free mass	H.pylori	Wald ratio	1	-0.010	0.005	0.0668	-0.020	0.001
Overweight	H.pylori	IVW	2	-0.040	0.024	0.1011	-0.087	0.008

Cardioembolic stroke	H.pylori	IVW	2	0.156	0.095	0.1018	-0.031	0.343
Illnesses of father: Prostate cancer	H.pylori	Wald ratio	1	0.004	0.002	0.1071	-0.001	0.009
2hr glucose	H.pylori	IVW	2	0.083	0.055	0.1288	-0.024	0.190
Diastolic blood pressure automated reading	H.pylori	Wald ratio	1	0.013	0.009	0.1300	-0.004	0.030
Medication for cholesterol blood pressure diabetes or take exogenous hormones: Cholesterol lowering medication	H.pylori	Wald ratio	1	0.006	0.004	0.1407	-0.002	0.013
Large vessel disease	H.pylori	IVW	2	0.220	0.167	0.1891	-0.108	0.547
Illnesses of father: Lung cancer	H.pylori	Wald ratio	1	-0.003	0.003	0.1932	-0.008	0.002
Fasting glucose	H.pylori	IVW	2	-0.003	0.002	0.2015	-0.007	0.002
Body fat	H.pylori	IVW	2	-0.019	0.015	0.2045	-0.049	0.010
Corrected insulin response	H.pylori	IVW	2	0.079	0.062	0.2049	-0.043	0.200
Extreme waist-to-hip ratio	H.pylori	IVW	2	-0.116	0.094	0.2168	-0.299	0.068
Childhood obesity	H.pylori	Wald ratio	1	-0.162	0.144	0.2583	-0.444	0.119
Illnesses of mother: Diabetes	H.pylori	Wald ratio	1	0.003	0.003	0.2585	-0.002	0.008
Insulin disposition index	H.pylori	IVW	2	0.095	0.085	0.2651	-0.072	0.261
Extreme body mass index	H.pylori	IVW	2	-0.104	0.098	0.2891	-0.297	0.089
Medication for cholesterol blood pressure diabetes or	H.pylori	Wald ratio	1	0.005	0.004	0.2903	-0.004	0.013

take exogenous hormones: Blood pressure medication								
Fasting proinsulin	H.pylori	IVW	2	0.025	0.024	0.3065	-0.023	0.072
Weight	H.pylori	IVW	2	-0.018	0.018	0.3152	-0.053	0.017
Lung adenocarcinoma	H.pylori	Wald ratio	1	-0.089	0.091	0.3283	-0.269	0.090
HDL cholesterol	H.pylori	Wald ratio	1	-0.026	0.026	0.3291	-0.077	0.026
Non-cancer illness code self-reported: high cholesterol	H.pylori	Wald ratio	1	0.003	0.003	0.3483	-0.003	0.008
Illnesses of mother: Chronic bronchitis/emphysema	H.pylori	Wald ratio	1	0.002	0.002	0.3545	-0.002	0.006
Illnesses of siblings: Diabetes	H.pylori	Wald ratio	1	-0.002	0.003	0.3801	-0.007	0.003
Lung cancer	H.pylori	Wald ratio	1	-0.048	0.060	0.4234	-0.166	0.070
Exposure to tobacco smoke at home	H.pylori	Wald ratio	1	-0.002	0.003	0.4785	-0.009	0.004
Medication for cholesterol blood pressure or diabetes: Blood pressure medication	H.pylori	Wald ratio	1	0.004	0.005	0.4895	-0.007	0.014
Illnesses of father: High blood pressure	H.pylori	Wald ratio	1	0.003	0.004	0.4966	-0.005	0.010
Pack years of smoking	H.pylori	Wald ratio	1	-0.010	0.015	0.5001	-0.041	0.020
Illnesses of siblings: Stroke	H.pylori	Wald ratio	1	0.001	0.002	0.5049	-0.002	0.004
Adiponectin	H.pylori	IVW	2	0.014	0.022	0.5141	-0.029	0.058

Waist circumference	H.pylori	IVW	2	-0.008	0.012	0.5146	-0.031	0.016
Insulin sensitivity index	H.pylori	IVW	2	0.042	0.066	0.5281	-0.088	0.172
Tobacco smoking: Ex-smoker	H.pylori	Wald ratio	1	0.005	0.008	0.5298	-0.011	0.021
Illnesses of siblings: High blood pressure	H.pylori	Wald ratio	1	0.002	0.004	0.5379	-0.005	0.010
Number of cigarettes currently smoked daily (current cigarette smokers)	H.pylori	Wald ratio	1	0.013	0.024	0.5840	-0.033	0.059
Apolipoprotein B	H.pylori	IVW	2	-0.018	0.033	0.5887	-0.082	0.047
Blood clot DVT bronchitis emphysema asthma rhinitis eczema allergy diagnosed by doctor: Emphysema/chronic bronchitis	H.pylori	Wald ratio	1	0.001	0.001	0.6022	-0.002	0.003
Medication for cholesterol blood pressure diabetes or take exogenous hormones: Hormone replacement therapy	H.pylori	Wald ratio	1	0.002	0.003	0.6133	-0.004	0.008
Small vessel disease	H.pylori	IVW	2	-0.053	0.106	0.6147	-0.260	0.154
AUCins/AUCglu	H.pylori	IVW	2	0.034	0.068	0.6239	-0.101	0.168

Incremental insulin at 30 minutes	H.pylori	IVW	2	-0.030	0.065	0.6371	-0.157	0.096
Triglycerides	H.pylori	Wald ratio	1	-0.012	0.026	0.6431	-0.063	0.039
Illnesses of father: Diabetes	H.pylori	Wald ratio	1	-0.001	0.003	0.6435	-0.006	0.004
Illnesses of father: Chronic bronchitis/emphysema	H.pylori	Wald ratio	1	-0.001	0.003	0.6452	-0.007	0.004
Illnesses of siblings: Heart disease	H.pylori	Wald ratio	1	-0.001	0.003	0.6643	-0.007	0.004
Current tobacco smoking	H.pylori	Wald ratio	1	-0.002	0.005	0.6759	-0.011	0.007
Waist-to-hip ratio	H.pylori	IVW	2	0.005	0.012	0.7010	-0.019	0.028
Squamous cell lung cancer	H.pylori	Wald ratio	1	0.034	0.091	0.7070	-0.143	0.212
Type 2 diabetes	H.pylori	Wald ratio	1	-0.032	0.089	0.7218	-0.206	0.142
Treatment/medication code: warfarin	H.pylori	Wald ratio	1	0.000	0.001	0.7278	-0.002	0.001
Insulin at 30 minutes	H.pylori	IVW	2	0.025	0.076	0.7364	-0.123	0.174
Illnesses of mother: High blood pressure	H.pylori	Wald ratio	1	0.001	0.004	0.7555	-0.007	0.009
Smoking status: Previous	H.pylori	Wald ratio	1	0.001	0.004	0.7631	-0.007	0.009
Ever smoked	H.pylori	Wald ratio	1	-0.001	0.004	0.7710	-0.009	0.007
Medication for cholesterol blood pressure diabetes or take exogenous hormones: None of the above	H.pylori	Wald ratio	1	-0.002	0.005	0.7769	-0.012	0.009

Vascular/heart problems diagnosed by doctor: High blood pressure	H.pylori	Wald ratio	1	-0.001	0.004	0.8332	-0.008	0.007
Percent emphysema	H.pylori	Wald ratio	1	-0.011	0.054	0.8402	-0.117	0.095
Coronary heart disease	H.pylori	IVW	2	0.009	0.045	0.8427	-0.078	0.096
Illnesses of father: Heart disease	H.pylori	Wald ratio	1	0.001	0.004	0.8597	-0.008	0.009
AUCins	H.pylori	IVW	2	-0.010	0.067	0.8854	-0.141	0.121
Past tobacco smoking	H.pylori	Wald ratio	1	-0.001	0.011	0.8984	-0.023	0.020
Medication for cholesterol, blood pressure or diabetes: Cholesterol lowering medication	H.pylori	Wald ratio	1	-0.001	0.005	0.9005	-0.011	0.010
Medication for cholesterol, blood pressure or diabetes: Insulin	H.pylori	Wald ratio	1	0.000	0.001	0.9021	-0.003	0.003
Non-cancer illness code self-reported: emphysema/chronic bronchitis	H.pylori	Wald ratio	1	0.000	0.001	0.9312	-0.002	0.002
Fasting insulin	H.pylori	IVW	2	0.000	0.002	0.9511	-0.004	0.004
HbA1C	H.pylori	IVW	2	0.001	0.011	0.9525	-0.021	0.022

Medication for cholesterol blood pressure or diabetes: None of the above	H.pylori	Wald ratio	1	0.000	0.006	0.9651	-0.011	0.012
Myocardial infarction	H.pylori	IVW	2	-0.002	0.054	0.9673	-0.108	0.103
Apolipoprotein A-I	H.pylori	IVW	2	-0.003	0.074	0.9704	-0.147	0.142

Table 5. *H.pylori* bidirectional Mendelian randomisation results

LDL cholesterol	Method	SNPs	Odds ratio	Lower 95% CI	Upper 95% CI	se	p
	MR Egger	75	1.069	-0.080	0.214	0.075	0.376
	Weighted median	75	1.077	-0.071	0.219	0.074	0.318
	Inverse variance weighted	75	1.036	-0.064	0.135	0.051	0.489
	Simple mode	75	0.941	-0.329	0.207	0.137	0.657
	Weighted mode	75	1.039	-0.096	0.172	0.069	0.580
Hip circumference	MR Egger	225	0.64	-0.901	0.009	0.232	0.056
	Weighted median	225	0.942	-0.315	0.195	0.130	0.644
	Inverse variance weighted	225	1.054	-0.106	0.212	0.081	0.516
	Simple mode	225	0.974	-0.644	0.591	0.315	0.932
	Weighted mode	225	0.932	-0.504	0.362	0.221	0.748
Heart rate	MR Egger	15	1.065	-0.054	0.179	0.060	0.311
	Weighted median	15	1.011	-0.026	0.049	0.019	0.555
	Inverse variance weighted	15	1.008	-0.020	0.037	0.015	0.564
	Simple mode	15	1.007	-0.054	0.068	0.031	0.823
	Weighted mode	15	1.013	-0.037	0.063	0.026	0.614

Breast cancer	MR Egger	9	29.928	-4.059	10.857	3.805	0.401
	Weighted median	9	0.274	-4.905	2.317	1.842	0.482
	Inverse variance weighted	9	0.47	-3.526	2.017	1.414	0.594
	Simple mode	9	0.114	-7.425	3.078	2.680	0.441
	Weighted mode	9	0.237	-5.770	2.891	2.209	0.533

Table 6. Bidirectional MR: LDL cholesterol single-SNP analysis results. The analysis estimated the causal effect of LDL cholesterol (exposure of interest) on *H.pylori* infection (outcome of interest) for each genetic instrument in a sample size of 173,082 participants.

SNP	β	se	p	Lower 95% CI	Upper 95% CI
rs10195252	-0.207	0.900	0.818	-1.971	1.557
rs10490626	-0.675	0.757	0.373	-2.159	0.809
rs10832962	-0.603	0.758	0.427	-2.088	0.883
rs10893499	-0.263	0.595	0.658	-1.429	0.902
rs10903129	-0.394	0.646	0.543	-1.660	0.873
rs10947332	-0.275	1.473	0.852	-3.162	2.611
rs112201728	0.253	0.657	0.699	-1.033	1.540
rs11563251	-0.462	0.983	0.638	-2.390	1.465
rs11591147	0.193	0.361	0.593	-0.515	0.901
rs1169288	0.928	0.607	0.126	-0.262	2.118
rs12066643	0.571	0.915	0.533	-1.223	2.366
rs1250229	-0.391	0.993	0.694	-2.337	1.555
rs12721109	0.083	0.170	0.624	-0.250	0.417
rs12748152	0.011	0.783	0.989	-1.524	1.546

rs12916	-0.143	0.292	0.626	-0.715	0.430
rs13206249	0.246	0.742	0.740	-1.207	1.700
rs13277801	1.177	0.660	0.075	-0.117	2.470
rs1367117	0.062	0.189	0.743	-0.309	0.433
rs1408272	0.004	0.790	0.996	-1.544	1.552
rs1564348	-0.306	0.585	0.600	-1.453	0.840
rs16831243	0.043	0.916	0.963	-1.753	1.839
rs16891156	0.669	0.797	0.401	-0.892	2.231
rs17404153	-1.074	0.955	0.261	-2.946	0.797
rs174583	-0.325	0.451	0.472	-1.209	0.559
rs1800961	0.755	0.922	0.413	-1.052	2.562
rs1801689	-0.199	0.730	0.785	-1.629	1.231
rs1883025	-0.943	0.818	0.249	-2.546	0.660
rs2000999	0.080	0.407	0.845	-0.719	0.878
rs2030746	-0.399	0.999	0.690	-2.358	1.560
rs2073547	-0.240	0.613	0.695	-1.442	0.962
rs2228603	0.706	0.415	0.089	-0.108	1.520
rs2328223	1.338	0.910	0.142	-0.446	3.122
rs2390536	-3.038	0.984	0.002	-4.967	-1.109
rs2419604	0.588	0.774	0.447	-0.929	2.105
rs2495495	1.439	2.132	0.500	-2.740	5.619
rs2587534	0.513	0.539	0.341	-0.543	1.568
rs2642438	-0.090	1.471	0.951	-2.974	2.794
rs267733	0.628	0.887	0.479	-1.110	2.366

rs2710642	0.395	0.944	0.675	-1.454	2.245
rs2737252	-0.236	0.745	0.751	-1.696	1.224
rs2738459	-0.457	0.410	0.265	-1.260	0.346
rs2886232	0.610	0.741	0.410	-0.842	2.063
rs2965157	-0.186	0.337	0.582	-0.847	0.475
rs314253	1.850	0.912	0.043	0.062	3.638
rs3184504	-0.984	0.787	0.211	-2.526	0.559
rs364585	0.479	0.868	0.581	-1.222	2.180
rs3757354	-0.519	0.682	0.447	-1.855	0.817
rs3780181	-0.539	0.933	0.563	-2.367	1.289
rs4253776	2.111	1.074	0.049	0.006	4.215
rs4530754	-0.521	0.768	0.498	-2.025	0.984
rs4722551	0.345	0.742	0.642	-1.110	1.800
rs4942486	0.876	0.868	0.312	-0.824	2.577
rs4970712	1.554	1.756	0.376	-1.888	4.997
rs5763662	-0.707	0.966	0.464	-2.601	1.186
rs579459	0.025	0.383	0.947	-0.726	0.776
rs6016373	-1.533	0.614	0.013	-2.736	-0.330
rs6065311	-0.636	0.505	0.208	-1.626	0.354
rs646776	0.074	0.160	0.642	-0.239	0.387
rs6504872	1.450	0.768	0.059	-0.055	2.956
rs6511720	-0.112	0.147	0.447	-0.399	0.176
rs6544713	0.089	0.279	0.749	-0.458	0.636
rs6709904	0.254	0.676	0.707	-1.071	1.579

rs676388	0.348	0.795	0.661	-1.210	1.907
rs6818397	0.028	0.964	0.977	-1.862	1.918
rs6882076	-0.249	0.479	0.603	-1.187	0.689
rs6909746	0.236	0.810	0.770	-1.351	1.823
rs7254892	0.103	0.130	0.428	-0.152	0.359
rs72902576	0.292	0.661	0.659	-1.005	1.588
rs7551981	0.201	0.460	0.662	-0.700	1.102
rs75687619	1.239	0.555	0.025	0.153	2.326
rs7640978	-0.136	0.960	0.887	-2.018	1.746
rs7832643	-0.228	0.634	0.719	-1.471	1.014
rs8017377	0.021	0.695	0.976	-1.342	1.383
rs9875338	-0.107	0.834	0.898	-1.742	1.529
rs9987289	-0.007	0.510	0.990	-1.006	0.993

Table 7. Bidirectional MR: Hip circumference single-SNP analysis results. The analysis estimated the causal effect of hip circumference (exposure of interest) on *H.pylori* infection (outcome of interest) for each genetic instrument in a sample size of 336,601 participants.

SNP	β	se	p	Lower 95% CI	Upper 95% CI
rs10100245	0.302	1.073	0.779	-1.802	2.405
rs10118701	-1.050	1.304	0.421	-3.606	1.506
rs1014291	0.802	1.383	0.562	-1.909	3.514
rs10153248	-0.301	1.282	0.814	-2.814	2.211
rs10236214	2.339	1.216	0.054	-0.044	4.722
rs10237317	-4.464	3.418	0.192	-11.164	2.236

rs10269774	-0.058	1.018	0.955	-2.054	1.938
rs10404726	-0.762	1.305	0.559	-3.319	1.795
rs10756798	-2.901	1.516	0.056	-5.873	0.070
rs10777859	-0.763	1.399	0.586	-3.505	1.980
rs10820852	-1.954	1.653	0.237	-5.194	1.286
rs10887571	-0.488	1.616	0.763	-3.656	2.680
rs10938397	-0.136	0.919	0.882	-1.938	1.666
rs11030119	-0.352	0.721	0.625	-1.765	1.060
rs11109097	-0.952	1.624	0.558	-4.136	2.232
rs1118151	-1.165	1.512	0.441	-4.128	1.797
rs112646560	-1.019	1.325	0.442	-3.616	1.578
rs1127100	0.283	1.564	0.856	-2.783	3.349
rs113866544	0.322	1.412	0.820	-2.446	3.090
rs11584359	-1.069	1.461	0.464	-3.932	1.794
rs11766945	-0.026	1.461	0.986	-2.889	2.836
rs11839227	-0.226	1.638	0.890	-3.437	2.985
rs11862944	-0.650	1.477	0.660	-3.544	2.245
rs11882409	-0.105	1.369	0.939	-2.789	2.578
rs11944291	3.351	1.529	0.028	0.353	6.349
rs12096864	1.577	1.702	0.354	-1.758	4.912
rs12140153	-0.959	1.723	0.578	-4.336	2.417
rs12209223	-0.763	1.681	0.650	-4.059	2.532
rs12254441	-0.008	1.534	0.996	-3.013	2.998
rs12300276	-0.706	1.672	0.673	-3.984	2.572

rs1231281	0.269	1.535	0.861	-2.739	3.276
rs12375196	0.630	1.334	0.637	-1.985	3.244
rs12475388	0.189	1.606	0.906	-2.958	3.337
rs12519997	-0.254	1.538	0.869	-3.269	2.761
rs12528644	-1.063	1.092	0.331	-3.203	1.078
rs12561919	4.110	1.653	0.013	0.870	7.350
rs12607512	-0.603	1.603	0.707	-3.744	2.538
rs12619178	-1.011	1.432	0.480	-3.817	1.795
rs12680342	-0.077	1.502	0.959	-3.020	2.866
rs12701265	2.966	1.637	0.070	-0.242	6.174
rs12714415	-0.059	0.632	0.926	-1.297	1.179
rs12779865	0.569	1.042	0.585	-1.473	2.611
rs12805742	0.401	1.458	0.783	-2.457	3.259
rs12877270	0.825	1.461	0.572	-2.038	3.688
rs12883788	1.876	1.400	0.180	-0.868	4.619
rs12920259	0.860	1.393	0.537	-1.871	3.590
rs1293395	-0.538	1.616	0.739	-3.706	2.630
rs1296328	-0.723	1.413	0.609	-3.492	2.046
rs13107325	-0.552	0.907	0.543	-2.330	1.227
rs1320903	-0.501	1.198	0.676	-2.848	1.847
rs13333747	-0.815	1.287	0.527	-3.338	1.708
rs13389219	0.353	0.858	0.681	-1.329	2.034
rs13410783	1.379	1.575	0.381	-1.707	4.465
rs138767	-1.596	1.610	0.322	-4.751	1.560

rs1396513	0.641	1.178	0.586	-1.668	2.949
rs141622900	-3.138	1.976	0.112	-7.010	0.735
rs1428120	-0.988	1.625	0.543	-4.172	2.197
rs143384	-1.888	0.764	0.013	-3.386	-0.391
rs1441264	0.395	1.277	0.757	-2.109	2.898
rs1449630	-1.951	1.692	0.249	-5.267	1.365
rs1458156	1.719	1.618	0.288	-1.451	4.890
rs1477290	-1.671	1.213	0.168	-4.048	0.706
rs147730268	0.310	0.800	0.698	-1.258	1.878
rs1528450	0.610	1.303	0.640	-1.944	3.163
rs1569497	1.306	1.426	0.360	-1.489	4.101
rs1582931	-1.074	1.103	0.330	-3.237	1.088
rs1618069	1.563	1.603	0.330	-1.579	4.705
rs1618725	-0.452	1.017	0.657	-2.445	1.542
rs1662185	-0.561	1.457	0.700	-3.416	2.294
rs17024393	0.064	1.184	0.957	-2.258	2.386
rs17245511	0.091	1.649	0.956	-3.142	3.323
rs1727901	1.277	1.365	0.349	-1.397	3.952
rs17770336	-1.582	1.175	0.178	-3.885	0.721
rs1955695	3.126	1.197	0.009	0.779	5.473
rs1979440	1.085	1.502	0.470	-1.860	4.029
rs2013002	-1.160	1.231	0.346	-3.573	1.253
rs2034768	-0.059	1.304	0.964	-2.614	2.497
rs2102278	1.505	1.553	0.333	-1.539	4.550

rs2187449	2.404	1.693	0.156	-0.914	5.721
rs2236519	-1.309	1.095	0.232	-3.456	0.838
rs2244786	-1.873	1.657	0.258	-5.122	1.376
rs2307111	-0.025	0.763	0.974	-1.520	1.469
rs2318543	0.337	1.530	0.826	-2.662	3.335
rs2439823	0.330	1.161	0.776	-1.946	2.606
rs245775	-2.414	1.668	0.148	-5.683	0.856
rs2494196	-0.179	0.777	0.818	-1.702	1.345
rs2499468	-0.138	1.643	0.933	-3.359	3.083
rs252749	-1.400	1.135	0.217	-3.624	0.824
rs2678204	-0.506	1.088	0.642	-2.637	1.626
rs2715439	1.115	1.382	0.420	-1.593	3.823
rs273505	-0.329	1.507	0.827	-3.283	2.626
rs2737250	0.128	1.092	0.907	-2.012	2.267
rs2814943	-1.073	0.541	0.047	-2.134	-0.012
rs2815753	-0.044	1.025	0.966	-2.053	1.965
rs28377268	-0.534	1.535	0.728	-3.543	2.475
rs28418580	0.214	1.249	0.864	-2.234	2.661
rs28479795	-0.259	1.130	0.819	-2.474	1.956
rs2897968	0.622	1.650	0.706	-2.612	3.856
rs2954021	0.856	1.210	0.480	-1.516	3.228
rs3218036	-1.648	1.161	0.156	-3.924	0.627
rs33955687	2.960	1.545	0.055	-0.068	5.988
rs34013042	-1.377	1.675	0.411	-4.661	1.907

rs34049648	-1.757	1.336	0.189	-4.377	0.862
rs34373881	-0.908	1.690	0.591	-4.221	2.405
rs34517439	0.970	0.808	0.230	-0.613	2.553
rs34629844	-2.039	1.599	0.202	-5.174	1.095
rs34748838	0.346	1.050	0.742	-1.713	2.405
rs34769775	-1.016	1.521	0.504	-3.998	1.966
rs35779991	2.592	1.513	0.087	-0.372	5.557
rs35874463	0.734	1.969	0.709	-3.125	4.594
rs35882248	-1.256	1.422	0.377	-4.043	1.531
rs3746759	-1.809	1.689	0.284	-5.120	1.502
rs3803286	3.333	1.334	0.012	0.718	5.949
rs3807566	1.635	1.383	0.237	-1.076	4.345
rs3810291	-1.540	0.981	0.116	-3.462	0.382
rs3811951	1.726	1.689	0.307	-1.584	5.036
rs3826408	-0.758	1.391	0.586	-3.484	1.968
rs3845344	1.969	1.471	0.181	-0.915	4.853
rs390192	-0.767	1.537	0.618	-3.780	2.246
rs40071	0.500	1.744	0.774	-2.917	3.918
rs41284816	0.983	1.047	0.348	-1.069	3.035
rs4240326	0.162	0.787	0.837	-1.380	1.704
rs4297095	1.662	1.783	0.351	-1.833	5.157
rs4402589	0.010	0.737	0.990	-1.435	1.454
rs4430895	-0.634	0.893	0.478	-2.384	1.116
rs4467770	3.029	1.522	0.047	0.046	6.013

rs4482463	2.565	1.458	0.079	-0.293	5.423
rs4630170	-1.079	1.539	0.483	-4.094	1.937
rs4660586	0.153	1.312	0.907	-2.417	2.724
rs4670612	1.275	1.444	0.377	-1.555	4.105
rs4722398	-1.523	1.539	0.322	-4.539	1.493
rs4741546	0.479	1.281	0.708	-2.032	2.990
rs4777541	2.711	1.412	0.055	-0.056	5.479
rs4790292	0.099	1.166	0.932	-2.187	2.385
rs4794222	-1.241	1.538	0.420	-4.255	1.773
rs4982753	0.159	1.641	0.923	-3.057	3.375
rs4985407	1.861	1.529	0.224	-1.136	4.858
rs543874	-0.124	0.589	0.833	-1.279	1.031
rs55726687	-1.047	1.353	0.439	-3.700	1.605
rs55932154	2.077	1.562	0.184	-0.984	5.138
rs56094641	-0.143	0.367	0.697	-0.863	0.577
rs56288810	2.317	1.555	0.136	-0.731	5.366
rs57636386	0.144	1.125	0.898	-2.061	2.349
rs58551145	0.892	1.199	0.457	-1.458	3.242
rs588660	1.652	1.187	0.164	-0.674	3.978
rs59738707	-0.644	1.530	0.674	-3.642	2.354
rs6080646	1.781	1.602	0.266	-1.358	4.921
rs60984707	1.198	1.669	0.473	-2.073	4.470
rs6142059	0.571	1.340	0.670	-2.056	3.198
rs62243489	0.854	1.711	0.618	-2.500	4.208

rs62246314	-0.045	1.559	0.977	-3.100	3.010
rs62425398	1.297	1.702	0.446	-2.039	4.633
rs6470771	1.554	1.721	0.367	-1.819	4.927
rs6501601	0.180	1.268	0.887	-2.306	2.666
rs6535240	-1.160	1.676	0.489	-4.444	2.125
rs6567160	-0.006	0.501	0.991	-0.987	0.976
rs6575340	-0.735	1.125	0.514	-2.940	1.471
rs6585201	0.009	1.045	0.993	-2.038	2.057
rs6601527	0.565	1.445	0.696	-2.267	3.397
rs66679256	1.228	1.503	0.414	-1.718	4.174
rs6669341	-0.426	1.517	0.779	-3.400	2.548
rs6707036	-1.306	1.629	0.423	-4.499	1.886
rs6739755	0.470	1.487	0.752	-2.444	3.384
rs675162	0.452	1.203	0.707	-1.906	2.811
rs6821305	2.073	1.505	0.169	-0.878	5.023
rs6840236	0.276	1.308	0.833	-2.287	2.839
rs6867299	-1.405	1.181	0.234	-3.720	0.911
rs6907872	0.169	1.670	0.920	-3.104	3.442
rs6973656	-0.044	1.199	0.970	-2.395	2.306
rs6999725	0.247	1.626	0.879	-2.940	3.434
rs7116641	-0.623	1.025	0.543	-2.632	1.386
rs7124681	0.487	1.076	0.651	-1.621	2.596
rs7132908	1.103	0.922	0.232	-0.704	2.910
rs7145337	-1.546	1.559	0.321	-4.601	1.509

rs7226064	-0.121	1.518	0.936	-3.096	2.853
rs7238896	1.197	1.567	0.445	-1.874	4.268
rs724016	2.157	0.798	0.007	0.593	3.721
rs72656010	0.695	1.489	0.640	-2.222	3.613
rs7274811	-0.605	1.141	0.596	-2.841	1.631
rs72801854	1.819	1.345	0.176	-0.818	4.456
rs72892910	-0.968	0.855	0.258	-2.644	0.708
rs72959041	-0.132	0.962	0.891	-2.017	1.754
rs73175572	0.125	1.402	0.929	-2.623	2.873
rs7426945	1.529	1.508	0.311	-1.426	4.484
rs74749286	0.067	1.240	0.957	-2.364	2.498
rs750090	-0.167	1.371	0.903	-2.854	2.520
rs7516554	1.956	1.445	0.176	-0.876	4.788
rs7548408	0.595	1.434	0.678	-2.216	3.407
rs75543804	2.963	1.825	0.104	-0.613	6.539
rs756717	0.267	3.101	0.931	-5.812	6.345
rs75949361	-2.455	1.513	0.105	-5.421	0.512
rs76040172	-3.622	1.526	0.018	-6.612	-0.632
rs76798800	-0.983	1.018	0.334	-2.978	1.012
rs77165542	-0.233	0.911	0.798	-2.019	1.552
rs7845090	-1.065	1.109	0.337	-3.238	1.107
rs7893571	1.182	1.509	0.434	-1.776	4.140
rs7915723	-1.770	1.716	0.302	-5.133	1.592
rs7930275	0.869	1.468	0.554	-2.008	3.745

rs7957774	-0.242	1.626	0.882	-3.430	2.945
rs7978353	2.752	1.317	0.037	0.171	5.333
rs7982447	2.103	1.729	0.224	-1.287	5.493
rs79969674	1.095	1.764	0.535	-2.362	4.553
rs8011368	1.127	1.564	0.471	-1.937	4.192
rs8023263	0.871	1.462	0.551	-1.994	3.737
rs8042404	-0.146	1.490	0.922	-3.066	2.774
rs8064502	-0.974	1.312	0.458	-3.546	1.598
rs8133137	-1.077	1.593	0.499	-4.198	2.045
rs815335	0.678	1.249	0.587	-1.770	3.126
rs8192675	-0.179	1.346	0.894	-2.817	2.459
rs845084	0.447	1.432	0.755	-2.360	3.253
rs869400	3.696	1.550	0.017	0.657	6.734
rs879620	1.214	0.843	0.150	-0.437	2.866
rs882378	-1.400	1.575	0.374	-4.486	1.686
rs894347	-0.452	1.240	0.715	-2.882	1.977
rs9378684	-0.459	1.438	0.750	-3.276	2.359
rs9415106	0.653	1.642	0.691	-2.564	3.871
rs9496567	-1.189	1.288	0.356	-3.714	1.336
rs9512696	-1.193	1.348	0.376	-3.835	1.449
rs962554	-0.404	1.139	0.723	-2.636	1.827
rs968379	1.774	1.324	0.180	-0.821	4.368
rs9808900	-0.226	1.069	0.833	-2.320	1.869
rs9814633	-1.340	1.597	0.401	-4.471	1.791

rs982692	4.433	1.599	0.006	1.300	7.566
rs9843653	-0.457	1.056	0.665	-2.527	1.612
rs9967367	0.967	1.498	0.518	-1.968	3.903