6-4-2019

A MENDELIAN RANDOMIZATION STUDY OF CORONARY ARTERY DISEASE AND THREE AMINO ACIDS: ALANINE, GLYCINE, AND GLUTAMINE

Allan Uribe
CUNY School of Public Health, allanuribe@me.com

Follow this and additional works at: https://academicworks.cuny.edu/sph_etds

Part of the Cardiovascular Diseases Commons, Epidemiology Commons, Genetics Commons, and the Other Public Health Commons

Recommended Citation

This Dissertation is brought to you for free and open access by the CUNY Graduate School of Public Health & Health Policy at CUNY Academic Works. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@cuny.edu.
Preliminary pages

A MENDELIAN RANDOMIZATION STUDY OF CORONARY ARTERY DISEASE AND THREE AMINO ACIDS: ALANINE, GLYCINE, AND GLUTAMINE

A DISSERTATION
by
ALLAN URIBE
Concentration: EPIDEMIOLOGY

Presented to the Faculty at the Graduate School of Public Health and Health Policy in partial fulfillment of the requirements for the degree of Doctor of Public Health

Graduate School of Public Health and Health Policy
City University of New York
New York, New York
MAY, 2019

Dissertation Committee:

C.MARY SCHOOLING, PHD
SHIRO HORIUCHI, PHD
JENNIFER DOWD, PHD
Copyrighted By

ALLAN URIBE

2019

All rights reserved
A MENDELIAN RANDOMIZATION STUDY OF CORONARY ARTERY DISEASE AND THREE AMINO ACIDS: ALANINE, GLYCINE, AND GLUTAMINE

by

Allan Uribe

Advisor: C. Mary Schooling, PhD

ABSTRACT

Cardiovascular disease is the leading cause of death worldwide. Coronary Artery Disease (CAD) accounts for the majority of those deaths. Observational studies have identified risk factors that have been helpful in lowering the death rate, including hypertension, high cholesterol, diabetes, smoking, physical inactivity and poor diet. The effects of these risk factors on CAD remain unclear. To clarify the effect of three amino acids, alanine, glutamine, and glycine on CAD I applied a two sample Mendelian randomization analysis to extensively genotyped observational data. In a sample with up to 184,000 individuals and approximately 60,000 controls, SNPs that reached genome wide significance with each amino acid were identified. Linkage equilibrium was assessed for each SNP. Known pleiotropy was assessed using the phenoscanner database. Unknown pleiotropy was assessed using MR-Egger. Using the IVW approach, the odds of CAD was 0.89 lower per one standard deviation of genetically
determined higher level of alanine and was statistically significant (95% confidence interval (CI) (0.80-0.99). The odds of CAD was 1.09 higher per one standard deviation of genetically determined higher level of glutamine and was not statistically significant (95% confidence interval (CI) (0.99-1.19). The odds of CAD was 0.94 lower per one standard deviation of genetically determined higher level of glycine and was statistically significant (95% confidence interval (CI) (0.91-0.98). After sensitivity analyses suggested pleiotropic effects, the analysis was repeated, removing SNPs with known pleiotropy. Results remained the same for all three amino acids. MI was not statistically significantly associated with any of the amino acids. This study provides genetic validation that alanine could be a new target of intervention to prevent the leading cause of global morbidity and mortality. Glutamine should be investigated further to determine its status as a risk factor for CAD. Glycine may decrease the risk of coronary artery disease however additional research is needed.
# Table of Contents

- Preliminary pages ......................................................... 0
- Table of Contents ....................................................... 4
- Acknowledgements ........................................................ 10
- Disclosure Statement ..................................................... 10
- Introduction (overall) ................................................... 11

## Specific Aims .............................................................. 11

- Aim 1: (Alanine) ........................................................... 11
- Hypothesis 1 .............................................................. 11
- Aim 2: (Glutamine) ....................................................... 11
- Hypothesis 2: A genetically higher level of glutamine is causally associated with a lower risk of CAD and MI .................. 11
- Aim 3: (Glycine) ........................................................... 11
- Hypothesis 3: A genetically higher level of glycine is causally associated with a lower risk of CAD ............................ 11

## Background ................................................................. 12

- Cardiovascular Disease .................................................. 12
- Amino Acids ............................................................... 14

This difficulty justifies a novel methodological approach that assesses the effects of risk factors according to genetically determined serum levels of glycine, glutamine, and alanine. Mendelian randomization (MR) has emerged as a popular epidemiological method to obtain unconfounded causal estimates. [60-62] ................................................. 18

- Methods/Mendelian Randomization .................................. 19
Figure 1. Directed Acyclic Graph for Mendelian Randomization[33, 78]... 23

Paper 1(Alanine) ................................................................. 25

Abstract/Summary ..................................................................... 25

Introduction ............................................................................... 27

Methods .................................................................................... 29

Genetic associations with alanine.............................................. 30
Genetic associations with CAD and MI....................................... 31
Statistical analysis..................................................................... 32
Sensitivity analysis.................................................................... 33

Results ....................................................................................... 34

Genetic determinants of alanine................................................. 34
Association of genetically determined alanine with CAD risk.... 35
Association of genetically determined alanine with MI risk....... 36
Figure 1 – Wald ratio for individual SNPs and MR-Egger and IVW for all SNPs on CAD and MI................................................................. 37
Table 1 MR estimates of the associations of log-transformed standardized residuals of Alanine with IHD and with MI......................... 39

Discussion ................................................................................ 40

Conclusion.................................................................................. 44

Appendix .................................................................................... 45

Table 2 Characteristics of the single nucleotide polymorphisms (SNP) used for genetically determined serum Alanine.............................................. 45
Table 3 Select Pleiotropic effects of Relevant Alanine SNPs.......... 46
Figure 2 – MR Method comparison plot of SNP effect on Alanine and CAD/MI ........................................................................................................... 47
Figure 3 Leave one Alanine SNP out Sensitivity Analysis for CAD and MI. 49
Paper 2 (Glutamine) ........................................ 53

Abstract/Summary ..................................................... 0

Introduction ............................................................. 1

Methods ........................................................................ 4

Genetic associations with glutamine ................................. 5
Genetic associations with CAD ...................................... 6
Statistical analysis ........................................................ 7
Sensitivity analysis ........................................................ 8

Results .......................................................................... 9

Genetic determinants of Glutamine ................................... 9
Association of genetically determined glutamine with CAD risk ... 10
Association of genetically determined glutamine with MI risk .... 11
Figure 1 Wald ratio for individual SNPs and MR-Egger and IVW for all SNPs on CAD and MI .............................................. 13
Table 1 MR estimates of the associations of log-transformed standardized residuals of Glutamine with IHD and with MI .......... 15

Discussion ...................................................................... 16

Conclusion ...................................................................... 19

Appendix ......................................................................... 20

Table 2 Characteristics of the single nucleotide polymorphisms (SNP) used for genetically determined serum glutamine .................. 20
Table 3 Select Pleiotropic effects of Relevant Glutamine SNPs .......... 21
Figure 2 MR Method comparison plot of SNP effect on Alanine and CAD/MI. 22
Abstract/Summary.................................................................0

Introduction .................................................................2

Methods .................................................................................7
   Genetic associations with glycine...........................................8
   Genetic associations with CAD.............................................9
   Statistical analysis.............................................................9
   Sensitivity analysis..........................................................10

Results ....................................................................................11
   Genetic determinants of glycine..........................................11
   Association of genetically determined glycine with CAD risk...12
   Association of genetically determined glycine with MI risk.....13
   Figure 1 Wald ratio for individual SNPs and MR-Egger and IVW for all SNPs on CAD and MI..................................................15
   Table 1 MR estimates of the associations of log-transformed standardized residuals of Glycine with IHD and with MI................17

Discussion ...............................................................................18
   Conclusion...........................................................................23

Appendix .................................................................................24
   Table 2 Characteristics of the single nucleotide polymorphisms (SNP) used for genetically determined serum glycine..............................24
   Table 3 Select Pleiotropic effects of Relevant Glycine SNPs..........25
Figure 2 Leave one Glycine SNP out Sensitivity Analysis for CAD........ 26
Figure 3 Plots of SNPs on glycine against SNPs on CAD/MI.............. 28
Figure 4 Funnel plot of Relevant SNPs to assess heterogeneity of the Glycine on CAD and MI Analysis................................................. 30

Conclusion(overall) ...................................................................... 32

Citations .......................................................................................... 8
Intentionally left blank
Acknowledgements

I would like to thank my dissertation supervisor, Dr. C. Mary Schooling, PhD for all her effort, patience and emotional support. I could not have completed this without your guidance though each stage of this process. I am extremely grateful Dr. Heidi Jones, PhD who always believed in me and rooted for me, even when I couldn’t. I would like to acknowledge Dr. Shiro Horiuchi PhD and Dr. Jennifer Dowd PhD, for their feedback on my research. I would like to acknowledge the CUNY school of Public Health for providing me an environment to develop, personally and professionally. I would like thank my mother, Wilma Valencia, who struggled alongside me for several years as I navigated the ups and downs of this process. Fatima Kadik, I owe you the biggest thanks, for the unwavering support, countless edits, helpful comments, and the often-needed pep talks.

Disclosure Statement

NONE
Introduction (overall)

Specific Aims

Aim 1: (Alanine)
To estimate the causal effect of alanine on coronary artery disease (CAD) and myocardial infarction (MI) using instrumental variable analysis with a genetic instrument, i.e., Mendelian randomization.

Hypothesis 1: A genetically higher level of alanine is causally associated with a lower risk of CAD.

Aim 2: (Glutamine)
To estimate the causal effect of glutamine on CAD/MI using instrumental variable analysis with genetic instruments, i.e., Mendelian randomization.

Hypothesis 2: A genetically higher level of glutamine is causally associated with a lower risk of CAD and MI.

Aim 3: (Glycine)
To estimate the causal effect of glycine on CAD/MI using instrumental variable analysis with genetic instruments, i.e., Mendelian randomization.

Hypothesis 3: A genetically higher level of glycine is causally associated with a lower risk of CAD.
Background

Cardiovascular Disease

Cardiovascular disease (CVD) is the leading cause of death worldwide.[1, 2] The World Health Organization (WHO) estimates that in 2016, 17,900,00 deaths, or 31% of all deaths, were from CVD.[2] In the United States in 2015, about one in four deaths, or approximately 610,000 deaths, were due to CVD.[3-5] Coronary artery disease (CAD) accounted for 42% of all CVD deaths worldwide in 2015, or approximately 7.4 million deaths, while CAD accounted for about 60% of all CVD deaths in the United States, or approximately 375,000 deaths.[3, 4, 6-11]

Observational epidemiological studies have identified numerous risk factors for CAD, including hypertension, hyperlipidemia, smoking, obesity, diabetes, unhealthy diet, and physical inactivity.[6, 12-21] However, a large proportion of CVD events occur in individuals with low predicted-risk based on established risk factors.[22] Furthermore, risk stratification models predict higher CAD for men than women at the same levels of known risk factors.[16-21, 23] Randomized controlled trials (RCTs)[24] have identified that pharmacotherapies capable of mitigating CAD risk factors and CAD have had inconsistent results.[15] How known risk factors and other factors combine to cause CAD remains poorly understood,[6, 25-27] and the elements
of a healthy diet, including fat, protein, carbohydrates, vitamins and minerals have proved difficult to establish, begging the question “does diet matter?”[28] Meta-analyses of observational studies do not always draw the same conclusion about the relationship between protein consumption and CAD/myocardial infarction (MI).[29-32] The building blocks of proteins are amino acids, and their relationship with CAD/MI is not clearly documented in the literature. The mixed results of observational studies provide a justification for exploring the relationship of amino acids, such as alanine, glutamine, and glycine with CAD/MI.

Considerable opportunity to identify new casual factors and treatments for CAD/MI remain. This study will use Mendelian randomization (MR) to estimate the effects of key amino acids on CAD/MI. MR has several advantages when compared to conventional observational studies and is being increasingly used because it overcomes some major limitations of observational study designs, including unmeasured confounding. This MR study will lead to a better understanding of the causes of CAD/MI.

Observational studies have begun investigating the relationship of amino acids to health and in particular, CAD.[33] Some of these studies have identified amino acids that seem to be protective against metabolic diseases such as diabetes or CAD,
and others have identified amino acids that are likely risk factors for metabolic diseases or CAD. In many cases, it remains unclear if biomarkers such as amino acids are merely correlated with CAD/MI, or causally related.[28, 34-44] Moreover, if they are causal, the underlying mechanism that leads to CAD is unclear.

**Amino Acids**

Glutamine, glycine, and alanine have been established in animal studies to be linked with heart function and heart disease.[34-36, 40-42, 45-48] Glutamine is commonly found in nuts, which observational studies and RCTs have found to be cardio-protective.[44, 49-51] Animal model and small RCT studies have found that glutamine improves postoperative outcomes and restores the heart to normal function more quickly than without glutamine supplements.[38, 47, 48] Similarly, glycine studies have shown an inverse relationship with CAD and CAD risk factors such as hypertension. Alanine studies have shown a paradoxical relationship with diabetes, a risk factor for CAD, in light of the beneficial role that alanine plays in glucose homeostasis and the increased alanine observed during cardiac events.[52-54]

Glutamine and glycine are conditionally-essential amino acids.[39, 40] This means under normal circumstances the body
produces the necessary amount of these amino acids for the biosynthesis of proteins. However, in conditions such as injury and stress, the body may supplement these amino acids through dietary intake.[37] Glutamine is found in a variety of foods, including nuts.[49-51] The most relevant glutamine producing human tissue is muscle, accounting for the majority of the glutamine produced by the body.[39] Glycine is most notably found in gelatin and is a major component of collagen, which is abundant in blood vessels and the gut, among other fibrous tissues.[36, 40, 42, 55, 56]

Glutamine is found in many supplements used by athletes because it helps signal cell growth and increases water retention in muscle cells.[34] It is commonly found in walnuts and other nuts.[51] An RCT in 2011 assessed the cardio-protective effects of glutamine on patients with CAD following open heart surgery.[38] Clinical tests after surgery found that glutamine improved heart functioning in the experimental arm.[39] In an animal experiment, researchers reported that post-ischemic reperfusion of a rat heart with glutamine restored normal cardiac output.[41] A meta-analysis of 53 RCT studies found moderate to weak evidence that glutamine supplementation would influence outcomes such as infection and recovery from surgery and other critical illness. However, they found no
effect on the risk of mortality or length of intensive care unit (ICU) stay.[39]

Glycine is an amino acid involved in anti-oxidative reactions and collagen formation.[56] Many observational studies have shown an inverse association of glycine with hypertension and type two diabetes.[42] A 2016 RCT assessed the relationship of plasma glycine and acute myocardial infarction in patients with stable angina. They reported that glycine was inversely associated with the risk of MI.[56] They found a stronger association in patients with apolipoprotein B, low density lipoprotein cholesterol, or apolipoprotein A-1 above the median.[56] Other studies have reported the positive effects of glycine on vascular health.[42] Glycine has been reported to stabilize platelets, reduce the size of the ischemia injury upon reperfusion, correct elevated blood pressure and normalize serum triglycerides.[55, 56] While these studies in rats and humans have reported an inverse relationship with glycine and CAD or CAD related risk factors, they have not compared glycine levels in those with to those without cardiovascular disease.

Alanine is a non-essential amino acid, meaning that the human body can produce the amount required and no diet source is needed to supplement endogenously produced alanine. Alanine is
found in red meat. There have been some studies that reported an association between the levels of alanine in the urine and hypertension, diet, cholesterol, and body mass index (BMI), however, these studies have not measured dietary alanine or plasma alanine levels. Another study found that higher serum levels of alanine were associated with the development of type 2 diabetes.[57]

Alanine and glutamine are considered the most important amino acids in the blood and are the only amino acids that are taken up or released by human skeletal muscle.[36] Once alanine is synthesized, it is converted to pyruvate and is critical in blood sugar management.[36] Animal studies have reported that heart tissue releases alanine only after a CAD event, otherwise the heart only releases alanine under stress.[48, 58] In another study, researchers used the Hoorn cohort study data to assess the relationship between elevated alanine serum levels and actual CAD events at 10-year follow up.[59] The study found an age- and sex-adjusted two-fold increased hazard of experiencing a CAD event at the 10-year follow up point. Furthermore, when adjusted for the most well-known CAD risk factors, the hazard ratio for those with elevated alanine was 1.88, suggesting an independent pathway of alanine from the known risk factors. The effect of amino acids, such as alanine, on glucose metabolism
are well established, however, the net impact on glucose homeostasis remains unclear.[54] The literature contains reports of both positive and negative impacts of amino acids on glycemic regulation and metabolic syndrome.[54] Alanine may be a marker of an underlying etiology relating to the paradoxical association of diabetes with CAD at the population level. However, the discrepancy reported in these studies generates more questions than answers about the relationship of alanine to CAD and in each case, these studies establish an association but do not establish a causal relationship.[54]

The existing literature suggests inconsistent effects of alanine, glycine, and glutamine on CVD and CVD related risk factors. A large RCT with CAD events as the primary outcome would be definitive, but costly. Additionally, randomly assigning research participants to alter amino acid consumption to determine the effect on CAD would be impractical and difficult to implement, as modifying diet on a long-term basis presents a compliance challenge and the outcome of interest takes decades to develop. It would also take several years and be difficult to justify with such little evidence of the benefits of glutamine and glycine.

This difficulty justifies a novel methodological approach that assesses the effects of risk factors according to genetically determined serum levels of glycine, glutamine, and
alanine. Mendelian randomization (MR) has emerged as a popular epidemiological method to obtain unconfounded causal estimates.[60-62]

Methods/Mendelian Randomization

Epidemiological studies fall into two general categories: experimental and observational studies. Experimental studies are characterized by the random assignment of the exposure. Observational studies can be designed in a variety of ways; however, none lend themselves to estimating causal relationships unequivocally, because of the risk of selection bias and confounding.

In health-related research, the randomized control trial (RCT) is the defining experimental study design. In the RCT, study participants are randomly assigned to a study arm, receiving one of several interventions or no intervention entirely by chance. This approach ensures that any measured, and, more importantly, unmeasured confounding is equally distributed across the study arms, minimizing the risk of spurious or erroneous findings. Blinding of all those involved with an RCT further strengthens the findings. However, RCTs are time-consuming and expensive. Moreover, not all exposures can feasibly be randomized. Observational studies may complement some of the limitations of RCTs but are not without limitations of their own. Determining
a causal relationship between a risk factor and health outcome in observational studies often relies on making some untestable and unlikely assumptions, such as no unmeasured confounding and no selection bias.

Mendelian randomization (MR) uses genetic variants as instrumental variables for modifiable risk factors that affect population health.[63] MR utilizes the econometrics method of instrumental variable (IV) analysis with genetic variants as the IV and offers a novel approach to estimating causal effects from observational studies. The principle underlying MR originates from the “The Law of Independent Assortment,” also known as Mendel’s second law, which states that inheritance of one trait is independent of other traits, thus allowing for the use of observational data with genetic variants as a natural experiment.[63, 64] MR is the use of data from nonexperimental studies to determine the causal effect of a phenotype on a disease outcome by leveraging the random assignment of genetic material at birth. Genetic variants, being fixed at conception, support causal inference by minimizing confounding.[63] As the knowledge of the human genome has grown, and access to genetic data increased, researchers have identified an increasing number of genetic determinants of common exposures.
As of 2015, genome-wide association studies had identified 46 loci directly linked to CAD, yet a majority of these loci have no apparent relation to cholesterol or other classical risk factors.[65] Evans and Smith (2015), note that since the formal presentation of the MR approach, an increasing number of mentions appear in reviews of scientific abstracts of peer-reviewed publications, with a steeply increasing trajectory.[66]

The first use of the term Mendelian randomization was in 1991 to describe a means of obtaining unbiased causal estimates when comparing bone marrow transplant with chemotherapy outcomes in the treatment of leukemia.[62] The first generation of genome-wide association Studies (GWAS) were used to show that several genetic variants were significantly associated with CAD in 2007.[67] The MR approach has been used to elucidate causal relationships in several areas of observational epidemiology, including the causal role of body mass index on blood pressure,[68, 69] and in evidence against the causal role of C-reactive protein in CAD or atherosclerosis.[33, 70-74] MR has been used extensively in CAD research, for example, to demonstrate that C-reactive protein levels are not causally associated with CAD, whereas lipoprotein(a) (Lp[a]) levels are causally associated with CAD, and to further our understanding of cardiovascular therapies, such as statins.[75, 76]
With the increased availability of molecular profiling techniques, particular interest has been shown in identifying novel biomarkers for cardiovascular or metabolic risk. [33] These methods have increased the availability of observational data sources that include genetic information to examine different exposures using MR to assess causal effects. [26, 62, 66, 76-80] This dissertation project will apply MR to publicly available data to evaluate the causal effect of three amino acids, specifically alanine, glutamine, and glycine, on the risk of CAD in adults. [26, 79, 80]

MR is a very powerful technique, but it does have strong assumptions that must be met to obtain a valid estimate. The assumptions of MR can be most clearly conveyed using a directed acyclic graph (DAG). The DAG for MR shows the causal relationships between the genetic instrument (Z), the modifiable risk factor (X), the outcome (Y) and the (known or unknown, measurable or non-measurable) confounders (U), as shown in Figure 1. [60-62, 66, 77, 78]
In an RCT, generally a treatment or other preventive measure is randomly assigned and so is not confounded. MR is instrumental variable analysis with genetic instruments.[60-62, 66, 77, 78] MR is based on the premise of random allocation of alleles at time of gamete formation, and hence is thought to be independent of confounding factors[60-62, 66, 77, 78, 81-84]. MR approximates RCT study design in this respect. However, it is important to note that the DAG shown in Figure 1 requires three key assumptions to be valid. These assumptions are implicit in the DAG, and normally expressed as
1) Relevance: The instrumental variables strongly predict the exposure
2) Independence: The instrumental variable is not confounded by the confounders of the exposure outcome relation
3) Exclusion-restriction: The instrumental variable only affects the outcome through the exposure, i.e., no pleiotropy

The purpose of this study is to better understand the causal relationship of three amino acids: glycine, glutamine and alanine with CAD. However, the casual mechanism of the RCT may more closely resemble the real-world relationship of the exposure with disease than the mechanism evaluated in an MR. First, the exposure that is assessed in the MR is a lifelong difference between genetic groups, while most RCTs and observational studies of CAD/MI are conducted in mature individuals. In this case, canalization, or an individual developing a compensatory mechanism in response to prolonged higher or lower levels of the exposure, is a possibility. Also, the prolonged exposure time assessed in the MR analysis could represent a cumulative effect of the exposure over time, rather than the acute response to a larger dose in a shorter period of time.
Second, some diseases are irreversible at a certain point of progression. No RCT or observational study could be designed that would be able to replicate the protective effects of reducing a harmful exposure that would be a result of genetic effects.

The efficacy of acute responses to sudden large increases in a harmful exposure cannot be assessed by a MR analysis. However, interest is usually in the lifelong effects of usual levels of exposure on disease. MR is an advantageous approach in this regard, as it is not possible to design an RCT that would assess lifelong average usual levels of exposure.

Alanine, glutamine and glycine and their relationship to CAD have yet to be evaluated using MR, instrumental variable analysis with genetic instruments.

**Paper 1 (Alanine)**

**Abstract/Summary**

Cardiovascular disease (CVD) is the leading cause of death worldwide. CVD is poorly understood. Little attention has been
given to the potential cardioprotective effects of alanine in humans. The relationship has primarily been explored in vitro and in animal models. Few observational studies have explored the effect of alanine on heart disease. To clarify we assessed whether people with genetically higher levels of alanine had lower coronary artery disease (CAD) and myocardial infarction (MI) risk. We used a two sample Mendelian randomization study. Six SNPs strongly associated (genome wide significant) with alanine were applied to large extensively genotyped CAD and MI studies to obtain unconfounded estimates. A one standard deviation increased level of alanine was found to protect against CAD (odds ratio (OR) 0.89, 95% confidence interval (CI) 0.80-0.99) but not MI (OR 0.90, 95% CI 0.80-1.02). Sensitivity analysis supports these findings. Our findings are consistent with the limited observational evidence, in vitro and animal studies. These findings genetically validate alanine as a potential target of intervention in CAD. Future clinical research is needed to better understand the mechanism through which alanine influences CAD and to develop effective treatment interventions that include alanine.
Introduction

Cardiovascular disease (CVD) is the leading cause of death worldwide.[1, 2] The World Health Organization (WHO) estimates that in 2016, 17,900,00 deaths, or 31% of all deaths, were from CVD.[2] In the United States in 2015, about one in four deaths, or approximately 610,000 deaths, were due to CVD.[3-5] Coronary artery disease (CAD) accounted for 42% of all CVD deaths worldwide in 2015, or approximately 7.4 million deaths, while CAD accounted for about 60% of all CVD deaths in the United States, or approximately 375,000 deaths.[3, 4, 6-11]

Observational epidemiological studies have identified numerous risk factors for CAD, including hypertension, hyperlipidemia, smoking, obesity, diabetes, unhealthy diet, and physical inactivity.[12-14] Yet, how diet, including protein, causes CAD remains poorly understood.[6, 25-27] Meta-analyses of observational studies do not always draw the same conclusion about the relationship between protein consumption and CAD/myocardial infarction (MI).[29-32]

Alanine is a non-essential amino acid, meaning that the human body can produce the amount required and that no diet source is needed to supplement endogenously produced alanine.[36] Alanine can also be obtained from diet, and is found in red meat[85].
Alanine levels have been shown to deviate from normal in response to cardiac insult, but studies of alanine and CAD have not established a clear association.[86] There have been some studies that reported associations between levels of alanine in the urine and hypertension, diet, cholesterol and body mass index (BMI); however, these studies did not measure dietary alanine or plasma alanine levels.[45, 87] Some studies have found that higher serum levels of alanine may protect against CAD.[52-54] Other studies have shown that increased levels of alanine are associated with CAD and other CVD.[88-91] However, higher levels of alanine appear to mitigate some type 2 diabetes symptoms, in particular by activating an adenosine monophosphate (AMP)-activated protein kinase nutrient sensor.[88] Alanine may be a marker of an underlying etiology explaining the paradoxical association of diabetes with CAD at the population level.[54] However, the inconsistencies reported in these studies have generated more questions than answers about the relationship of alanine to CAD, and in each case, these studies established an association but did not establish a causal relationship.[54]

There remains a considerable opportunity to identify new casual factors and treatments for CAD/MI. This study will use Mendelian randomization (MR) to estimate the effects of alanine on CAD/MI. MR has several advantages when compared to conventional
observational studies and is being increasingly used because it overcomes some major limitations of observational study designs, including unmeasured confounding. MR uses genetic variants as instrumental variables for modifiable risk factors that affect population health.[63] The principle underlying MR originates from the “The Law of Independent Assortment,” also known as Mendel’s second law, which states that inheritance of one trait is independent of other traits, thus allowing use of observational data with genetic variants as natural experiments.[63, 64] Genetic variants, being fixed at conception, support causal inference by minimizing confounding.[63]

**Methods**

MR depends on three assumptions. The relevance, independence and exclusion restriction assumptions are key in selecting valid instrumental variables. The relevance assumption means that the genetic instruments predict the exposure reliably. The relevance assumption is assured by selecting single nucleotide polymorphisms (SNPs) strongly associated with the exposure of interest, ideally at genome wide significance, from genome-wide association studies (GWAS).[63] The independence assumption means that the genetic instruments are not associated with
confounders of the exposure outcome relation. The independence assumption requires that potential confounders of the exposure outcome relation do not confound the genetic variant outcome relation, to minimize the possibility of confounding.[63, 92] The exclusion-restriction assumption means that the genetic instruments are only associated with the outcome via the exposure. The exclusion restriction assumption is assessed from biological knowledge of the effects of the genetic instruments, known associations of the genetic instruments with other factors and statistical tests for pleiotropy.

Contributed data was corrected for population stratification, which has been shown to occur even in single ethnicity populations. Ancestry and the non-random distribution of SNPs was corrected for using genomic control by adjusting for principal components.

Genetic associations with alanine

Genetically determined levels of alanine were obtained from a GWAS.[93, 94] The Computational Medicine extended GWAS combined data from 11 genetic studies and 14 data sources for a total sample size of 24,925 individuals of European descent (Finland, Estonian, Netherlands and Germany).[93, 94] The mean age of the study participants was 44.6 years. The individual study average ages ranged from 31.2 to 61.3 years. Women were 55% of the total
sample, and individual study samples ranged from 37% to 64% female.[93, 94] SNPs strongly and independently associated with alanine (p<5x10^-8) were selected.[93, 94] The MR-Base “clumping” function was used to identify correlations between the chosen SNPs,[95] and only uncorrelated SNPs were retained (r^2<0.001). SNP effects on CAD/MI via physiologic pathways other than alanine (pleiotropy) were assessed from a comprehensive genotype to phenotype cross-reference, PhenoScanner.[96, 97]

Genetic associations with CAD and MI
Data on coronary artery disease / myocardial infarction have been contributed by the CARDIoGRAMplusC4D. Data have been downloaded from www.CARDIOGRAMPLUSC4D.ORG. CARDIoGRAMplusC4D 1000 Genomes is a CAD and MI case (n=60,801)-control (n=123,504) study, extensively genotyped using the 1000 Genomes catalog. The majority (77%) of the participants were of European ancestry; 13% and 6% were of South Asian (India and Pakistan) and East Asian (China and Korea) ancestry, respectively, with smaller samples of Hispanic and African Americans. The mean age was about 56.9 years, with genetic associations adjusted for study specific covariates. Phenotyping of CAD, MI or both, were based on medical records, clinical diagnosis, procedures that indicate
CAD, medications or symptoms that indicate angina or from self-reported CAD.

**Statistical analysis**

Associations of alanine with CAD/MI were obtained from separate sample instrumental variable analysis with genetic instruments. SNP-specific Wald estimates (ratio of SNP on CAD/MI to SNP on alanine) of the effect of alanine on CAD/MI were obtained. The confidence interval for the Wald estimates were obtained using an approximation to Fieller’s theorem[66, 98-101].[102]

The SNP specific Wald-estimates were combined using inverse-variance weighting (IVW) with multiplicative random effects.[103] The IVW approach is asymptotically equivalent to the two-stage least squares (2SLS) estimate commonly used with individual-level data and yields comparable causal estimates.[104, 105]

SNPs were aligned on effect allele, and direction of effect amended accordingly using allele harmonization. There were no palindromic SNPs. All six SNPs for alanine were available for CAD and MI.
Sensitivity analysis

For sensitivity analysis, I used weighted median and MR-Egger regression. [104] The weighted-median is robust even in the event of some violations to the IV assumptions. This method provides a consistent estimate if at least 50% of the weight comes from valid IVs. [104] In simulation studies, the weighted-median estimation method produces similar point estimates, narrower confidence intervals and better finite-sample Type 1 error rates than IVW and the 2SLS method. [104]

The MR-Egger method uses a different, still untestable set of assumptions. [104] MR-Egger regression performs well even when 100% of the IVs are invalid. [104] The slope coefficient from the Egger regression method provides an estimate of the causal effect that is consistent asymptotically even if all the genetic variants have pleiotropic effects on the outcome; but assumes the inSiDe assumption— (instrument strength independent of direct effect). A stronger genetic variant should have more reliable estimates of the causal effect than a weaker variant. Once the average pleiotropic effect of variants is accounted for through the intercept term in Egger regression, any residual dose-response relationship in the genetic associations provides evidence of a causal effect. A non-null intercept from MR-Egger indicates an invalid IVW estimate.
Analysis was conducted using the TwoSample MR package version 0.4.14 for R software version 3.5.1, available at https://github.com/MRCIEU/TwoSampleMR. Analysis was verified using the MR-base platform version 1.2.1 e646be (05 December 2018), available at http://app.mrbase.org/. This analysis of publicly available data did not require IRB approval.

Results

Genetic determinants of alanine

One hundred and twenty-six SNPs were identified from GWAS with p<1x10^{-5}. Of these, 120 SNPs were excluded because they did not reach genome wide significance (p<5x10^{-8}). Of the remaining six SNPs, none were in linkage disequilibrium (R^2<0.001). The F-statistic for SNPs included in the CAD and MI analyses was 56.92 and SNP specific F-statistic ranged from 36.79 to 110.70. No palindromic SNPs were included in this analysis, and all SNPs in the SNP exposure dataset were available in the outcome dataset, for additional information about the included SNPs see table 2 in the appendix.

In MR studies obtaining adequate statistical power is frequently a concern due to the small amount of variation in a phenotypic trait that is typically explained by genetic variants, which is addressed by using large samples for the outcome. To achieve the
widely accepted power of 80% for epidemiological studies a
sample of about 255,000 participants would be needed for CAD and
about 326,553 participants would be needed for MI.[106-108] The
power calculated for these analyses was .66 for CAD with 184,000
participants and .52 for MI with 171,000 participants.[106-108]

Association of genetically determined alanine with CAD risk
Table 1 shows that the unconfounded estimate of one standard
deviation (SD) higher genetically determined alanine was
associated with lower risk for CAD. Using the IVW approach, the
odds of CAD was 0.89 lower among those with genetically
determined higher levels of alanine than those with lower
genetically determined alanine and was statistically significant
(95% confidence interval (CI) (0.80-0.99). Sensitivity analysis
results obtained using the weighted median and MR-Egger
regression method were not statistically significant, but were
similar in direction, as shown in table 1 and figure 1. Figure
1 shows the Wald estimates for included SNPs individually, as
well as the MR-Egger and IVW regression results for all SNPs
combined. All but one SNP (rs12578760) was found to be
associated with a lower risk of CAD. The funnel plot in figure
4 does not indicate heterogeneity among the SNPs, however, the
lack of symmetry could be an indication that an invalid SNP was
included.
Association of genetically determined alanine with MI risk

Table 1 shows that the causal effect of one standard deviation (SD) higher genetically determined alanine was not associated with MI. Using the IVW approach, the odds of MI was 0.90 lower among per SD of alanine (95% CI (0.800-1.015). Sensitivity analysis results obtained using the weighted median and MR-Egger regression method were not statistically significant, but were directionally similar, as shown in table 1 and figure 1. Figure 1 shows the Wald estimate for included SNPs individually, as well as the MR-Egger and IVW regression results for all SNPs combined. All but one SNP were found to be associated with lower MI. The funnel plot in figure 4 does not indicate heterogeneity among the SNPs, however, the lack of symmetry could be an indication that an invalid SNP was included.
Figure 1 - Wald ratio for individual SNPs and MR-Egger and IVW for all SNPs on CAD and MI
Table 1 MR estimates of the associations of log-transformed standardized residuals of Alanine with IHD and with MI

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Outcome</th>
<th># of SNPs</th>
<th>Mendelian Randomization Method</th>
<th>OR</th>
<th>95% CI</th>
<th>Cochran’s Q statistic P-Value</th>
<th>MR-Egger Intercept P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>CAD</td>
<td>6</td>
<td>IVW</td>
<td>0.889</td>
<td>0.795-0.993</td>
<td>0.369</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Weighted Median</td>
<td>0.928</td>
<td>0.807-1.066</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MR-Egger</td>
<td>0.875</td>
<td>0.486-1.578</td>
<td>0.249</td>
<td>0.961</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>6</td>
<td>IVW</td>
<td>0.901</td>
<td>0.800-1.015</td>
<td>0.763</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Weighted Median</td>
<td>0.929</td>
<td>0.801-1.076</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MR-Egger</td>
<td>0.932</td>
<td>0.530-1.637</td>
<td>0.631</td>
<td>0.911</td>
</tr>
</tbody>
</table>

One SNP, rs1260326 was associated with cholesterol and triglyceride levels, fasting glucose and blood pressure. The remaining SNPs had either no known pleiotropic effects or were associated with factors not associated with heart disease, such as glaucoma. Table 3 in the appendix shows these relations. Estimates were similar after excluding rs1260326. The IVW estimate remained statistically significant (odds ratio (OR) - 0.848, 95% confidence interval (CI) 0.742-0.969 and the MR-Egger results were (OR-0.641, 95% CI 0.321-1.281. The intercept P-value was 00.479 and the Q-statistic P-value was 0.3468.
Discussion

This novel study provides unconfounded estimates that alanine is associated with a lower risk of CAD. Clinical studies have found that perfusing the heart with alanine after surgery has improved recovery, and decreased risk of post-operative infection.\[89, 90, 109, 110\] Furthermore, small studies conducted in athletes have shown alanine supplementation has led to improved sports and exercise performance.\[52, 53\]. This study provides genetic validation for alanine as a new target of intervention in CAD.

This study has the advantage of using a novel technique which enables estimates to be made cost-effectively even when no study including both the exposure and outcome is available. It also has the advantage of providing unconfounded estimates. However, MR has stringent assumptions. First, the relevance assumption requires that SNPs be strongly associated with the exposure of interest.\[63\] This study used SNPs that were associated with the exposure at a P-value $<5 \times 10^{-8}$. Second, the independence assumption requires that SNPs be independent of exposure outcome confounders. This would be expected, as genetic makeup is determined a conception and independently from each other.\[64\] Third, the exclusion-restriction assumption requires that the SNP affects the outcome only via its effect on the risk factor
of interest. Known pleiotropy was assessed using phennoscanner and few associations with known CAD/MI risk factors were found for the SNPs included in this analysis. However, estimates were similar after pleiotropic SNPs were excluded. Unknown pleiotropy was evaluated using MR-Egger intercept, and found no indication of pleiotropy. Given the use of summarized data in two samples, serum alanine levels were not measured in the sample with the outcome. However, two-sample instrumental variable analysis is more robust to chance associations than analysis of a single sample. Despite the advantages of MR analysis, several limitations warrant mention.

Making inferences from population based genetic studies, one would be concerned with population stratification; when a population subgroup experiences different disease and SNP frequencies resulting in confounded associations in a whole population study. The exposure and outcome data included people largely of European descent. While this study can no assess the effect of alanine on CAD/MI by subgroup, there is no reason to think that mechanism by which alanine affects CAD/MI is group specific. Canalization, or developmental compensation occurs when molecular mechanisms compensate for the effect of the genotype of interest on the outcome of interest.
Weak instrument bias should not influence results of this study. All included SNPs reach genome wide significance. However, a weak instrument would bias results toward the null. Non-linear effects have successfully been identified using MR, but depend on additional assumptions. I had to assume a linear relationship. Participant overlap may have occurred in this two sample MR analysis. However, given the very large sample size of the exposure and outcome datasets, the bias if any should not be of major concern. The CAD/MI case-control studies included are not wholly composed of incident cases. Therefore associations may represent survival with CAD and of a MI rather than prevention of these outcomes. Finally, MR is better at identifying the direction of effect rather than its magnitude, though, the entire population is exposed to alanine, making any effect relevant to population health.

Alanine might protect against CAD. In rats, introducing alanine in their diets consistently activated the adenosine monophosphate-activated protein kinase (AMPK) enzyme.[88] AMPK is an enzyme found in the body that is activated when nutrients are low and during exercise.[88] This study also found that in rat, mouse and human cells introduction of alanine reduced glucose levels, not via increases of insulin secretion rather via increased glucose uptake in the liver.[88]
Chronic beta-alanine supplementation increases muscle carnosine concentration, subsequently leading to an improved performance in high-intensity exercise.[52] The mechanisms by which enhanced muscle carnosine concentrations result in improved exercise performance need to be refined in detail, with particular focus on a complex interaction between pH buffering and improved Ca2+ handling.[52] Observational studies and MR studies have clearly established the positive relationship between calcium and CAD.[26, 52, 111, 112]

Since the 1980’s the peer-reviewed literature has supported the potential for alanine to influence some mechanisms related to CAD and CAD risk factors.[113] The majority of studies exploring this relationship have been explored in vitro and in animal studies.[113] The paucity of population-based peer-reviewed literature addressing the relationship of alanine with human CAD suggests little consideration has been given to the relationship in epidemiological literature. Alanine is found abundantly in the musculoskeletal and heart cells.[85, 88, 90, 91, 113-116] Furthermore, beta-alanine is known to increase muscle carnosine levels.[35, 36, 52, 53, 85, 113, 116] Observational studies among athletes have shown that beta-alanine supplementation improves exercise performance.[52, 53, 113] Carnosine is widely considered an important anti-glycating agent, potentially
improving energy metabolism and protein homeostasis.[113]
Improved handling of glycation end products has had promising
results with diabetes complications, a CAD risk factor.[113]
Additionally, carnosine has antioxidative properties, which are
thought to improve heart health.[90, 113] However, little data
exists regarding alanine promoting antioxidative processes via
carnosine. Therefore, future research evaluating its
antioxidative potential in humans is warranted. Observational
studies in the early 2000’s energized interest in antioxidants
processes and heart disease.[89, 110, 113] Subsequent RCTs
concluded that there was little or no benefit of antioxidant
supplementation in reducing cardiovascular risk. It is possible
that the wrong antioxidants or combination of antioxidants, such
as α-tocopherol and β-carotene have been researched.[113, 116]
Other reasons for the failure of these RCTs to confirm the
observational findings are incorrect dosage, synthetic
antioxidants rather than dietary source and selection bias.[113, 116]

Conclusion
This study provides genetic validation that alanine could be a
new target of intervention to prevent the leading cause of
global morbidity and mortality. Alanine can be modified by diet
changes as it is found in red meat and poultry or via dietary
supplements. The effects and mechanisms by which beta alanine impacts health should be further evaluated in experimental and clinical studies, beyond the existing literature of in vitro and animal studies.

Appendix

Table 2 Characteristics of the single nucleotide polymorphisms (SNP) used for genetically determined serum Alanine.

<table>
<thead>
<tr>
<th>SNPs</th>
<th>Nearest Gene</th>
<th>Effect Allele</th>
<th>Other Allele</th>
<th>effect</th>
<th>Standard error</th>
<th>MAF</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12578760</td>
<td>RP11-96H19.1</td>
<td>C</td>
<td>T</td>
<td>0.07967</td>
<td>0.01314</td>
<td>0.8539</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>rs1260326</td>
<td>GCKR</td>
<td>C</td>
<td>T</td>
<td>-0.10458</td>
<td>0.00994</td>
<td>0.5895</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>rs149191093</td>
<td>DDX19B</td>
<td>C</td>
<td>T</td>
<td>-0.161931</td>
<td>0.024508</td>
<td>0.9632</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>rs2160387</td>
<td>-</td>
<td>T</td>
<td>C</td>
<td>-0.071001</td>
<td>0.009603</td>
<td>0.5567</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>rs2694917</td>
<td>RBMS2</td>
<td>C</td>
<td>T</td>
<td>0.086007</td>
<td>0.012989</td>
<td>0.2266</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>rs4554975</td>
<td>SLC38A4</td>
<td>A</td>
<td>G</td>
<td>-0.069135</td>
<td>0.009598</td>
<td>0.4294</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Table 3 Select Pleiotropic effects of Relevant Alanine SNPs

<table>
<thead>
<tr>
<th>SNPs</th>
<th>Nearest Gene</th>
<th>Associated Disease</th>
<th>Associated Metabolites</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12578760</td>
<td>RP11-96H19.1</td>
<td>Plateletcrit* Birth weight of first child* Self-reported glaucoma*</td>
<td>Mannose Serum total triglycerides VLDL Isoleucine</td>
</tr>
<tr>
<td>rs1260326</td>
<td>GCKR</td>
<td>Triglycerides Plasma lactate Urate Total cholesterol Whole body water mass C reactive protein Fasting glucose Self-reported gout Platelet count Sitting height Plateletcrit Neutrophil Reticulocyte albumin</td>
<td>Mannose Serum total triglycerides VLDL Isoleucine</td>
</tr>
<tr>
<td>rs149191093</td>
<td>DDX19B</td>
<td></td>
<td>Pyruvate</td>
</tr>
<tr>
<td>rs2160387</td>
<td>-</td>
<td></td>
<td>2-aminobutyrate Leucine Valine Phenylalanine</td>
</tr>
<tr>
<td>rs2694917</td>
<td>RBMS2</td>
<td></td>
<td>Glutamine Histidine Threonine</td>
</tr>
<tr>
<td>rs4554975</td>
<td>SLC38A4</td>
<td>Height</td>
<td>Alanine Glutamine ApoA1</td>
</tr>
</tbody>
</table>

Note: * $1 \times 10^{-5}$
Figure 2 - MR Method comparison plot of SNP effect on Alanine and CAD/MI
Figure 3 Leave one Alanine SNP out Sensitivity Analysis for CAD and MI
Figure 4 - Funnel plot of Relevant SNPs to assess heterogeneity of the Alanine on CAD and MI Analysis
MR Method
- Inverse variance weighted
- MR Egger

\[
\begin{align*}
\beta_{IV} & \quad \frac{1}{SE_v} \\
-0.25 & \quad 6 \\
-0.20 & \quad 6 \\
-0.15 & \quad 5 \\
-0.10 & \quad 5 \\
-0.05 & \quad 5 \\
0.00 & \quad 5
\end{align*}
\]
References

54 Schooling CM, Kelvin EA, Jones HE. Alanine transaminase has opposite associations with death from diabetes and ischemic


Paper 2 (Glutamine)

Abstract/Summary

Cardiovascular disease (CVD) is the leading cause of death worldwide. The strong relationship between CVD and lipids has been well documented in the peer reviewed literature. Little attention has been given to other, smaller molecular level dietary components. Glutamine, a non-essential amino acid and its effect on Coronary Artery Disease (CAD) have been studied in animal, observational, in vitro and RCT studies. These studies have had mixed findings raising doubt as to the protective effect of glutamine on CAD/MI. To clarify we assessed whether people with genetically higher levels of glutamine had lower coronary artery disease (CAD) and myocardial infarction (MI) risk. We used a two sample Mendelian randomization study. five SNPs strongly associated (genome wide significant) with glutamine were applied to large extensively genotyped CAD and MI studies to obtain unconfounded estimates. A one standard deviation increased level of glutamine was not statistically significantly associated with CAD (OR 1.09, 95% CI 0.99-1.19) or MI (OR 1.05, 95% CI 0.95-1.16). The sensitivity analysis also indicated that the IVW estimate might be invalid because the p-value for the MR-Egger intercept was significant. After removing pleiotropic SNPs the MR-Egger intercept was no longer significant. Our findings are inconsistent with the
observational, in vitro and animal studies findings in the older existing literature. However, more recently, observational studies have reported glutamine has atherogenic properties and two randomized studies found no association between glutamine and critical illness. These findings raise new questions of the causal effect of glutamine on CAD/MI. A harmful effect of glutamine on CAD cannot be ruled out, future research is needed to better understand whether glutamine causes CAD.

Introduction

Cardiovascular disease (CVD) is the leading cause of death worldwide.[1, 2] The World Health Organization (WHO) estimates that in 2016, 17,900,000 deaths, or 31% of all deaths, were from CVD.[2] In the United States, about one in four deaths, or approximately 610,000 deaths, were due to CVDs.[3-5] Coronary artery disease (CAD) accounted for 42% of all CVD deaths worldwide, or approximately 7.4 million deaths, while CAD accounted for about 60% of all CVD deaths in the United States, or approximately 375,000 deaths.[3, 4, 6-11]

Observational epidemiological studies have identified numerous risk factors for CAD, including hypertension, hyperlipidemia, smoking, obesity, diabetes, unhealthy diet, and physical
inactivity.[12-14] Yet, how these and other factors combine in complex causal mechanisms to cause CAD remains poorly understood.[6, 25-27] Meta-analyses of observational studies do not always draw the same conclusion about the relationship between protein consumption and CAD/myocardial infarction (MI).[29-32] Furthermore, randomized controlled trials (RCTs)[24] that have identified pharmacotherapies capable of mitigating CAD risk factors and CAD have had inconsistent results.[15]

Glutamine is a conditionally-essential amino acid.[39, 40] This means that under normal circumstances, the body produces the necessary amount of this amino acid for the biosynthesis of proteins. However, in conditions such as injury and stress, the body may require supplementation of the amino acid through dietary intake.[37] Glutamine is found in a variety of foods, notably nuts.[49-51] It is commonly found in walnuts and other nuts.[51] The most prominent glutamine-producing human tissue is muscle, accounting for the majority of the glutamine produced by the body.[39]

Glutamine is found in many supplements used by athletes because it helps signal cell growth and increases water retention in muscle cells.[34] A 2011 RCT assessed the cardio-protective
effects of glutamine on patients with CAD following open-heart surgery.[38] Clinical tests after surgery found that glutamine improved heart function in the experimental arm.[39] In an animal trial, researchers reported that post-ischemic reperfusion of a rat heart with glutamine restored normal cardiac output.[41] A meta-analysis of 53 RCT studies found moderate to weak evidence that glutamine supplementation influences outcomes such as infection and recovery from surgery and other critical illness. However, it found no effect on the risk of mortality or length of ICU stay.[39]

The existing literature suggests that glutamine has inconsistent effects on CVD and CVD-related risk factors. A large RCT with CAD events as the primary outcome would be definitive, but costly. Additionally, randomly assigning research participants to alter amino acid consumption to determine the effect on CAD would be impractical and difficult to implement, as modifying diet on a long-term basis presents a compliance challenge, and the outcome of interest takes decades to develop. It would also take several years and be difficult to justify with such mixed evidence of the benefits of glutamine. The potential harmful effect of glutamine would also make an RCT difficult to justify.
There remains a considerable opportunity to identify new casual factors and treatments for CAD/MI. This study will use Mendelian randomization (MR) to estimate the effects of glutamine on CAD/MI. MR has several advantages when compared to conventional observational studies and is being increasingly used because it overcomes some major limitations of observational study designs, including unmeasured confounding. MR uses genetic variants as instrumental variables for modifiable risk factors that affect population health.[63] The principle underlying MR originates from the “The Law of Independent Assortment,” also known as Mendel’s second law, which states that inheritance of one trait is independent of other traits, thus allowing for the use of observational data with genetic variants as natural experiments.[63, 64] Genetic variants, being fixed at conception, support causal inference by minimizing confounding.[63] MR studies are not influenced by confounding due to Mendel’s Law of independent assortment.[60]

**Methods**

As is the case in all analytic approaches, MR depends on three assumptions. The relevance, independence and exclusion restriction assumptions are key in selecting valid instrumental variables. The relevance assumption means that the genetic
instruments predict the exposure reliably. The relevance assumption is assured by selecting single nucleotide polymorphisms (SNPs) strongly associated with the exposure of interest from genome-wide association studies (GWAS).[63] The independence assumption means that the genetic instruments are not associated with confounders of the exposure outcome relation. The independence assumption requires the assessment of SNP effects, to minimize the possibility of confounding.[63, 92] The exclusion-restriction assumption means that the genetic instruments are only associated with the outcome via the exposure. The exclusion restriction assumption is assessed from biological knowledge of the effects of the genetic instruments, known associations of the genetic instruments with other factors and statistical tests for pleiotropy.

Contributed data was corrected for population stratification, which has been shown to occur even in single ethnicity populations. Ancestry and the non-random distribution of SNPs was corrected for using genomic control by adjusting for principal components.

*Genetic associations with glutamine*

Genetically determined levels of glutamine were obtained from a genome-wide association study (GWAS).[93, 94] The Computational Medicine extended GWAS combined data from 11 genetic studies and
14 data sources for a total sample size of 24,925 individuals of European descent (Finland, Estonian, Netherlands and Germany).[93, 94] The mean age of the study participants was 44.6 years. The individual study average ages ranged from 31.2 to 61.3 years. Women were 55% of the total sample, and individual study samples ranged from 37% to 64% female.[93, 94] Single nucleotide polymorphisms (SNPs) strongly and independently associated with glutamine (p<5x10^{-8}) were selected.[93, 94] The MR-Base “clumping” function was used to identify correlations between the chosen SNPs.[95] SNP effects on CAD/MI via physiologic pathways other than glutamine (pleiotropy) were assessed from a comprehensive genotype and phenotype cross-references, PhenoScanner.[96, 97]

Genetic associations with CAD

Data on CAD/MI were contributed by the CARDIoGRAMplusC4D. Data were downloaded from www.CARDIOGRAMPLUSC4D.ORG. CARDIoGRAMplusC4D 1000 Genomes is a CAD and MI case (n=60,801)-control (n=123,504) study extensively genotyped using the 1000 Genomes catalog that partially overlaps with CARDIoGRAMplusC4D. The majority (77%) of the participants were of European ancestry; 13% and 6% were of South Asian (India and Pakistan) and East Asian (China and Korea) ancestry, respectively, with smaller samples of Hispanic and African Americans. The mean age
was about 56.9 years, with genetic associations adjusted for study specific covariates. Phenotyping of CAD, MI or both was based on medical records, clinical diagnosis, procedures that indicate CAD, medications or symptoms that indicate angina, or self-reporting of CAD.

Statistical analysis
Associations of glutamine with CAD/MI were obtained from separate sample instrumental variable analysis with genetic instruments. SNP-specific Wald estimates (ratio of SNP on CAD/MI to SNP on glutamine) of the effect of glutamine on CAD/MI were obtained. The confidence intervals for the Wald estimates were obtained using an approximation to Fieller’s theorem.[66, 98-102]

The SNP specific Wald-estimates were combined using inverse-variance weighting (IVW) with multiplicative random effects.[103] The IVW approach is asymptotically equivalent to the two-stage least squares (2SLS) estimate commonly used with individual-level data and yields comparable causal estimates.[104, 105]
Sensitivity analysis

For sensitivity analysis, I used weighted median and MR-Egger regression.\[104\] The weighted-median is a robust alternative even in the event of some violations to the IV assumptions. Assuming that no single SNP contributes more than 50% of the weight to the estimate, this method provides a consistent estimate if at least 50% of the weight comes from valid IVs.\[104\] In simulation studies, the weighted-median estimation method produces similar point estimates, narrower confidence intervals and better finite-sample Type 1 error rates than IVW and the 2SLS method.\[104\]

The MR-Egger method uses a different, still untestable set of assumptions.\[104\] MR-Egger regression performs well even when 100% of the IVs are invalid.\[104\] The slope coefficient from the Egger regression method provides an estimate of the causal effect that is consistent asymptotically even if all the genetic variants have pleiotropic effects on the outcome; this is known as the inSiDe assumption—instrument strength independent of direct effect. A stronger genetic variant should have more reliable estimates of the causal effect than a weaker variant. Once the average pleiotropic effect of variants is accounted for through the intercept term in Egger regression, any residual dose-response relationship in the genetic associations provides
evidence of a causal effect. A non-null intercept from MR-Egger indicates an invalid IVW estimate.

Analysis was conducted using the TwoSample MR package version 0.4.14 for R software version 3.5.1, available at https://github.com/MRCIEU/TwoSampleMR. Analysis was verified using the MR-base platform version 1.2.1 e646be (05 December 2018), available at http://app.mrbase.org/. This analysis of publicly available data did not require IRB approval.

Results

Genetic determinants of Glutamine

Forty-four SNPs were identified from GWAS with \( p < 1 \times 10^{-5} \). Of these, 38 SNPs were excluded because they did not reach genome wide significance \( (p < 5 \times 10^{-8}) \). We attempted to align one palindromic SNP, rs7952320 using a minor allele frequency (MAF) of 0.3 but the results were ambiguous. No proxy SNP was available and rs7952320 was excluded for both the CAD and MI analysis. Of the remaining five SNPs, all were available in the outcome dataset. The final analysis was limited to the 5 SNPs available in both datasets. No SNP was in linkage disequilibrium \( (R^2 < 0.001) \). The F-statistic for SNPs included in the CAD and MI analyses was 91.48 and SNP specific F-statistic ranged from
31.33 to 314.03. [106-108] For additional information about the included SNPs see table 2 in the appendix.

In MR studies obtaining adequate statistical power is frequently a concern due to the small amount of variation in a phenotypic trait that is typically explained by genetic variants, which is addressed by using large samples for the outcome. To achieve the widely accepted power of 80% for epidemiological studies a sample of about 194,406 participants would be needed for CAD and about 879,248 participants would be needed for MI. [106-108] The power calculated for these analyses was .78 for CAD with 184,000 participants and .23 for MI with 171,000 participants. [106-108]

Association of genetically determined glutamine with CAD risk
Table 1 shows the unconfounded estimate of one standard deviation (SD) higher genetically determined glutamine was not clearly associated with increased risk for CAD or MI. Using the IVW approach, the odds of CAD was 1.09 higher per 1 SD increase in glutamine but was not statistically significant 95% confidence interval (CI) (0.99-1.19). Sensitivity analysis results obtained using the weighted median and MR-Egger regression method were statistically significant, with a similar odds ratio estimate as the IVW method, as shown in table 1 and figure 1. The sensitivity analysis also indicated that the IVW estimate might be invalid because the p-value for the MR-Egger
intercept was significant. Figure 1 shows the Wald ratio for included SNPs individually, as well as the MR-Egger and IVW regression results for all SNPs combined. The funnel plot in figure 4 does not indicate heterogeneity among the SNPs, however, because so few SNPs are included symmetry is difficult to assess and an invalid SNP may have been included. The MR-Egger method is valid even when invalid SNPs are included. The findings are unclear and do not unequivocally suggest that genetically determined glutamine is associated with increased risk of CAD/MI.

Association of genetically determined glutamine with MI risk
Table 1 shows the unconfounded estimate of one standard deviation (SD) higher genetically determined glutamine was not associated with increased risk of MI. Using the IVW approach, the odds of MI were 1.05 higher among those with genetically determined higher levels of glutamine than those with lower genetically determined glutamine, and was not statistically significant 95% CI (0.95-1.16); however, sensitivity analysis results obtained using the MR-Egger regression method was statistically significant, and showed increased risk of MI for genetically determined higher glutamine levels, as shown in table 1 and figure 1. MR-Egger also suggested that the IVW estimate was not valid. Figure 1 shows the Wald ratio for
included SNPs individually, as well as the MR-Egger and IVW regression results for all SNPs combined. The funnel plot in figure 4 does not indicate heterogeneity among the SNPs, however, the lack of symmetry could be an indication that an invalid SNP was included. The MR-Egger method is valid even when invalid SNPs are included and the findings among all three MR methods suggest that the possibility that genetically higher glutamine levels are associated with MI, but are not definitive.
Figure 1 Wald ratio for individual SNPs and MR-Egger and IVW for all SNPs on CAD and MI
Table 1 MR estimates of the associations of log-transformed standardized residuals of Glutamine with IHD and with MI

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Outcome</th>
<th># of SNPs</th>
<th>Mendelian Randomization Method</th>
<th>OR</th>
<th>95% CI</th>
<th>Cochran’s Q statistic P-Value</th>
<th>MR-Egger Intercept P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glutamine</td>
<td>CAD</td>
<td>5</td>
<td>IVW</td>
<td>1.09</td>
<td>0.99-1.19</td>
<td>0.5682</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Weighted Median</td>
<td>1.10</td>
<td>1.00-1.22</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MR-Egger</td>
<td>1.21</td>
<td>1.01-1.44</td>
<td>0.7765</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>5</td>
<td>IVW</td>
<td>1.05</td>
<td>0.95-1.16</td>
<td>0.4202</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Weighted Median</td>
<td>1.08</td>
<td>0.96-1.20</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MR-Egger</td>
<td>1.21</td>
<td>1.00-1.48</td>
<td>0.7809</td>
<td>0.019</td>
</tr>
</tbody>
</table>

Two SNPs, rs1260326 and rs6729711 were associated with cholesterol and triglyceride levels, fasting glucose and blood pressure. The remaining SNPs had either no known pleiotropic effects or were associated with factors not associated with heart disease, such as glaucoma. Table 3 in the appendix shows these relations. Estimates were similar after excluding rs1260326 and rs6729711. The IVW estimate CAD was (OR – 1.103,
95% CI 0.989-1.231) and the MR-Egger results were (OR- 1.254, 95% CI 1.019-1.543. The intercept P-value was 0.401 and the Q-statistic P-value was 0.4602.

**Discussion**

This novel study does not exclude the possibility that genetically determined glutamine is associated with an increased risk of CAD and MI. The existing literature remains divided on the direction of effect that glutamine has on CAD. Research has suggested supplementing glutamine in critically ill patients and after surgery reduces infections and mortality rates.[39] Clinical studies have found that perfusing the heart with glutamine after surgery has improved recovery, and decreased risk of post-operative infection.[89, 90, 109, 110] Animal models have shown that glutamine, but not glutamate or aspartate improved the cardiac output of rat hearts.[41] The same study reported that in a small sample, a single oral dose of glutamine improved cardiac function on patients with stable angina.[41] Another study linked glutamine supplements to improved immune and gastrointestinal function.[37] RCTs have demonstrated that glutamine supplements improve outcomes in critically ill patients, including one study that found glutamine was cardioprotective following cardiopulmonary bypass.[38] Some RCT studies have not found that glutamine has any beneficial effects.
in critically ill patients.[39, 117, 118] Our study does not clearly suggest that glutamine is a risk factor for CAD and MI.

This study has the advantage of using a novel technique which enables estimates to be made cost-effectively even when no study including both the exposure and outcome is available. It also has the advantage of providing unconfounded estimates. However, MR has stringent assumptions. First, the relevance assumption requires that SNPs be strongly associated with the exposure of interest.[63] This study used SNPs that were associated with the exposure at a P-value <5 \times 10^{-8}. Second, the independence assumption requires that SNPs be independent of exposure outcome confounders. This would be expected, as genetic makeup is determined a conception and independently from each other.[64] Third, the exclusion-restriction assumption requires that the SNP affects the outcome only via its effect on the risk factor of interest.[63] Known pleiotropy was assessed using phennoscanner and few associations with known CAD/MI risk factors were found for the SNPs included in this analysis.[96, 97] two SNPs, rs1260326 and rs6729711 are known to also be associated with cholesterol and triglyceride levels, fasting glucose and blood pressure. The remaining SNPs had either no known pleiotropic effects or were associated with factors not associated with heart disease, such as hypothyroidism. For
additional information please see the table 3 in the appendix.

Given the use of summarized data in two samples, serum glutamine levels were not measured in the sample with the outcome. However, two-sample instrumental variable analysis is more robust to chance associations than analysis of a single sample.[26]

Making inferences from population based genetic studies, one would be concerned with population stratification[83]; when a population subgroup experiences different disease and SNP frequencies resulting in confounded associations in a whole population study. The exposure and outcome data included people largely of European descent. While this study cannot assess the effect of glutamine on CAD/MI by subgroup, there is no reason to think that mechanism by which glutamine affects CAD/MI are group specific. Canalization, or developmental compensation occurs when molecular mechanisms compensate for the effect of the genotype of interest on the outcome of interest.[83]

Weak instrument bias should not influence results of this study.[83] All included SNPs reach genome wide significance. However, a weak instrument would bias results toward the null.[119] Non-linear effects have successfully been identified using MR, but depend on additional assumptions. I had to assume
a linear relationship in this analysis. Participant overlap may have occurred in this two sample MR analysis. However, given the very large sample size of the exposure and outcome datasets, the bias if any should not be of major concern.[26, 120, 121] Finally, MR is better at identifying the direction of effect rather than its magnitude, though, the entire population is exposed to glutamine, making any effect relevant to population health.

The role of glutamine in CAD and MI remains unclear. Studies have found glutamine a reliable discriminator of CAD and non-CAD among cardiac catheterization patients.[114, 122] Glutamine has also been found to be associated with plaque development and intima-media thickness.[114, 115, 123] Some in vitro and animal studies have found that glutamine levels have been associated with atherosclerosis and CAD risk. [114] Glutamine stimulates the accumulation of triglycerides in macrophages by affecting the SREBP-1, a key regulator in cellular triglyceride biosynthesis.[114] Furthermore, glutamine stimulates macrophage oxidative stress, affecting lipid metabolism.[114]

Conclusion
This study found glutamine had no statistically significant relationship with CAD or MI. It is possible that a study with higher power would have detected a small increased risk of CAD
and MI in individuals with genetically determined higher glutamine. Glutamine can be modified by diet changes as it is found in nuts and red meat or via dietary supplements. The literature has hypothesized several biological mechanisms by which glutamine effects CAD and MI. The effect on CAD should be conclusively evaluated in a larger MR study to assess glutamine as a risk factor for CAD and MI.

**Appendix**

*Table 2 Characteristics of the single nucleotide polymorphisms (SNP) used for genetically determined serum glutamine.*

<table>
<thead>
<tr>
<th>SNPs</th>
<th>Nearest Gene</th>
<th>Effect Allele</th>
<th>Other Allele</th>
<th>effect</th>
<th>Standard error</th>
<th>MAF</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12306007</td>
<td>RP11-96H19.1</td>
<td>T</td>
<td>C</td>
<td>0.079358</td>
<td>0.013439</td>
<td>0.860</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>rs1260326</td>
<td>GCKR</td>
<td>C</td>
<td>T</td>
<td>0.061666</td>
<td>0.010082</td>
<td>0.361</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>rs2657879</td>
<td>DDX19B</td>
<td>G</td>
<td>A</td>
<td>-</td>
<td>0.221406</td>
<td>0.824</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>rs6729711</td>
<td>-</td>
<td>A</td>
<td>G</td>
<td>-</td>
<td>0.069575</td>
<td>0.819</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>rs7078003</td>
<td>RBMS2</td>
<td>T</td>
<td>C</td>
<td>0.073854</td>
<td>0.011712</td>
<td>0.804</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>rs7952320</td>
<td>SLC38A4</td>
<td>C</td>
<td>G</td>
<td>0.053765</td>
<td>0.009369</td>
<td>0.466</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
### Table 3 Select Pleiotropic effects of Relevant Glutamine SNPs

<table>
<thead>
<tr>
<th>SNPs</th>
<th>Nearest Gene</th>
<th>Associated Disease</th>
<th>Associated Metabolites</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12306007</td>
<td>chr12:46790259</td>
<td>Birth weight of first child Height</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Triglycerides, Plasma lactate, Urate, Total cholesterol, Whole body water mass, C reactive protein, Fasting glucose, Self-reported gout, Platelet count, Sitting height, Plateletcrit, Neutrophil, Reticulocyte, albumin</td>
<td>Mannose, Serum total triglycerides, VLDL, Isoleucine</td>
</tr>
<tr>
<td>rs1260326</td>
<td>GCKR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2657879</td>
<td>GLS2</td>
<td>Fasting glucose</td>
<td></td>
</tr>
<tr>
<td>rs6729711</td>
<td>AC005540.3</td>
<td>Mean platelet volume, Height*, reticulocyte count*, hypothyroidism*, myxedema*, Pulse rate*, Diastolic blood pressure*</td>
<td></td>
</tr>
<tr>
<td>rs7078003</td>
<td>RP11-548K23.11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: * $1 \times 10^{-5}$
Figure 2 MR Method comparison plot of SNP effect on Alanine and CAD/MI
Figure 3 Leave one Glutamine SNP out Sensitivity Analysis for CAD and MI
Figure 4 Funnel plot of Relevant SNPs to assess heterogeneity of the Glutamine on CAD and MI Analysis
References


Cardiovascular disease (CVD) is the leading cause of death worldwide. CVD is poorly understood. Some attention has been given in the literature to the cardioprotective effects glycine has in humans. The relationship has primarily been explored in clinical and animal models. Few large observational studies have explored the effect of glycine on heart disease. To clarify we assessed whether people with genetically higher levels of glycine had lower coronary artery disease (CAD) and myocardial infarction (MI) risk. We used a two sample Mendelian randomization study. Six SNPs strongly associated (genome wide significantly) with glycine were applied to large extensively genotyped CAD and MI studies to obtain unconfounded estimates. A one standard deviation increase in genetically determined glycine was associated with lower risk of CAD (odds ratio (OR) 0.94, 95% confidence interval (CI) 0.91-0.98) but not MI (OR 0.97, 95% CI 0.93-1.02). Sensitivity analysis indicated that the IVW estimate might be invalid because of pleiotropy (significant p-value for the MR-Egger intercept). After removing pleiotropic SNPs the MR-Egger intercept was no longer significant. These findings cannot rule out the possibility that glycine is cardioprotective and raises new questions of the causal effect.
of glycine on CAD/MI. Future research is needed to better understand whether glycine influences CAD.
Introduction

Cardiovascular disease (CVD) are the leading cause of death worldwide.[1, 2] The World Health Organization (WHO) estimates that in 2016, 17,900,000 deaths, or 31% of all deaths, were from CVD.[2] In the United States, about one in four deaths, or approximately 610,000 deaths, were due to CVDs.[3-5] Coronary artery disease (CAD) accounted for 42% of all CVD deaths worldwide, or approximately 7.4 million deaths, while CAD accounted for about 60% of all CVD deaths in the United States, or approximately 375,000 deaths.[3, 4, 6-11]

Observational epidemiological studies have identified numerous risk factors for CAD, including hypertension, hyperlipidemia, smoking, obesity, diabetes, unhealthy diet, and physical inactivity.[12-14] Yet, how these and other factors combine in complex causal mechanisms to cause CAD remains poorly understood.[6, 25-27] Meta-analyses of observational studies do not always draw the same conclusion about the relationship between protein consumption and CAD/myocardial infarction (MI).[29-32] Furthermore, randomized controlled trials (RCTs)[24] that have tested pharmacotherapies capable of mitigating CAD risk factors have had inconsistent effects on CAD.[15]
Glycine is a conditionally-essential amino acid.\cite{39, 40} This means that under normal circumstances, the body produces the necessary amount of this amino acid for the biosynthesis of proteins. However, in conditions such as injury and stress, the body may require supplementation of the amino acid through dietary intake.\cite{37} Glycine is most notably found in gelatin and is a major component of collagen, which is abundant in blood vessels and the gut, among other fibrous tissues.\cite{36, 40, 42, 55, 56}

Glycine is an amino acid involved in anti-oxidative reactions and collagen formation.\cite{56} Many studies have shown an inverse association of glycine with hypertension and type two diabetes.\cite{42} A 2016 clinical study assessed the relationship of plasma glycine and acute MI. The authors reported that glycine was inversely associated with the risk of MI.\cite{56} They found a stronger association in patients with apolipoprotein B, low density lipoprotein cholesterol, or apolipoprotein A-1 above the median.\cite{56} Other studies have reported the positive effects of glycine on vascular health.\cite{42} Glycine has been reported to stabilize platelets, reduce the sizes of areas of ischemic injury upon reperfusion, correct elevated blood pressure and normalize serum triglycerides.\cite{55, 56} While these studies in rats and humans reported an inverse relationship with glycine
and CAD or CAD-related risk factors, none estimated a causal relationship or the underlying pathway by which glycine reduces CAD or its related risk factors.

The existing literature is consistent but inconclusive about the effects glycine has on CVD and CVD-related risk factors. A large RCT with CAD events as the primary outcome would be definitive, but costly. Additionally, randomly assigning research participants to alter amino acid consumption to determine the effect on CAD would be impractical and difficult to implement, as modifying diet on a long-term basis presents a compliance challenge, and the outcome of interest takes decades to develop. It would also take several years and be difficult to justify with limited evidence of the benefits of glycine.

A recent study published by Wittemans, Laura B. L. et.al. in Nature Communications used MR to estimate the causal relationship of glycine with CAD, CAD risk factors and diabetes.[124] The researchers make a compelling case for the cardioprotective effect of glycine on CAD.[124] They also make a case for systolic and diastolic blood pressure as a mediator between glycine and CAD.[124] Several differences warrant mention. The study sample used in the recently published study expanded on the exposure sample used in our study but may have
been affected by heterogeneity between studies due to differences metabolomics platforms and in analytical decisions between the studies, such as transformation of the glycine measures and inclusion of covariates.\[124-126]\] The SNP exposure estimates in the previously published study were adjusted for age, sex and the first 10 principal components, while my study only adjusted for the first 4 principal components.\[79, 93, 94, 124]\] Thus, raising a question about overfitting the SNP exposure model. The recently published paper developed sex specific models adjusting for several covariates, including age, sex, genotyping array, testing center and the first 4 principal components. Sex combined associations made no such adjustments. Our study also makes no such adjustments in the SNP outcome estimates.\[79, 93, 94, 124]\] However, the recently published paper used a weighted median regression MR analysis, while our study used IVW MR analysis as the main analysis. Weighted median regression analysis was used only as a sensitivity analysis in our study. Weighted median MR is an overidentification method that assume the instrumental variable assumptions hold for some of the genetic variants, but not necessarily for all genetic variants.\[105, 125-129]\] Median-based methods may suggest a positive causal effect despite evidence for directional pleiotropy as a result of the relaxed assumptions.\[105, 125-129]\] One issue with down-weighting is
that giving less weight to Wald estimates that deviate significantly from the majority and standard error of the causal effect estimate will be reduced after down weighting those SNPs that appear to deviate from the majority.[105, 125-129] This can be particularly problematic when no sound biologically plausible reason justifies such down weighting as the down-weighted outlier may be the most biologically reasonable.[105, 125-129]

There remains a considerable opportunity to identify new casual factors and treatments for CAD/MI. This study will use Mendelian randomization (MR) to estimate the effects of glycine on CAD/MI. MR has several advantages when compared to conventional observational studies and is being increasingly used because it overcomes some major limitations of observational study designs, including unmeasured confounding. MR uses genetic variants as instrumental variables for modifiable risk factors that affect population health.[63] The principle underlying MR originates from the “The Law of Independent Assortment,” also known as Mendel’s second law, which states that inheritance of one trait is independent of other traits, thus allowing for the use of observational data with genetic variants as natural experiments.[63, 64] Genetic variants, being fixed at conception, support causal inference by
minimizing confounding.[63] MR studies are not influenced by confounding due to Mendel’s Law of independent assortment.[60]

**Methods**

As is the case in all analytic approaches, MR depends on three assumptions. The relevance, independence and exclusion restriction assumptions are key in selecting valid instrumental variables. The relevance assumption means that the genetic instruments predict the exposure reliably. The relevance assumption is assured by selecting single nucleotide polymorphisms (SNPs) strongly associated with the exposure of interest from genome-wide association studies (GWAS).[63] The independence assumption means that the genetic instruments are not associated with confounders of the exposure outcome relation. The independence assumption requires the assessment of SNP effects, to minimize the possibility of confounding.[63, 92] The exclusion-restriction assumption means that the genetic instruments are only associated with the outcome via the exposure. The exclusion restriction assumption is assessed from biological knowledge of the effects of the genetic instruments, known associations of the genetic instruments with other factors and statistical tests for pleiotropy.
Genetic associations with glycine

Genetically determined levels of glycine were obtained from a genome-wide association study (GWAS).[93, 94] The Computational Medicine extended GWAS combined data from 11 genetic studies and 14 data sources for a total sample size of 24,925 individuals of European descent (Finland, Estonian, Netherlands and Germany).[93, 94] The mean age of the study participants was 44.6 years. The individual study average ages ranged from 31.2 to 61.3 years. Women were 55% of the total sample, and individual study samples ranged from 37% to 64% female.[93, 94] Single nucleotide polymorphisms (SNPs) strongly and independently associated with glycine (p<5x10^{-8}) were selected.[93, 94] The MR-Base “clumping” function was used to identify correlations between the chosen SNPs.[95] SNP effects on CAD/MI via physiologic pathways other than glycine (pleiotropy) were assessed from a comprehensive genotype and phenotype cross-references, PhenoScanner.[96, 97] Contributed data was corrected for population stratification, which has been shown to occur even in single ethnicity populations. Ancestry and the non-random distribution of SNPs was corrected for using genomic control by adjusting for principal components.
Genetic associations with CAD
Data on CAD/MI were contributed by the CARDIoGRAMplusC4D. Data were downloaded from www.CARDIOGRAMPLUSC4D.ORG.
CARDIoGRAMplusC4D 1000 Genomes is a CAD and MI case (n=60,801)-control (n=123,504) study extensively genotyped using the 1000 Genomes catalog that partially overlaps with CARDIoGRAMplusC4D. The majority (77%) of the participants were of European ancestry; 13% and 6% were of South Asian (India and Pakistan) and East Asian (China and Korea) ancestry, respectively, with smaller samples of Hispanic and African Americans. The mean age was about 56.9 years, with genetic associations adjusted for study specific covariates. Phenotyping of CAD, MI or both was based on medical records, clinical diagnosis, procedures that indicate CAD, medications or symptoms that indicate angina, or self-reporting of CAD.

Statistical analysis
Associations of glycine with CAD/MI were obtained from separate sample instrumental variable analysis with genetic instruments. SNP-specific Wald estimates (ratio of SNP on CAD/MI to SNP on glycine) of the effect of glycine on CAD/MI were obtained. The confidence interval for the Wald estimates were obtained using an approximation to Fieller’s theorem.[66, 98-102]
The SNP specific Wald-estimates were combined using inverse-variance weighting (IVW) with multiplicative random effects.[103] The IVW approach is asymptotically equivalent to the two-stage least squares (2SLS) estimate commonly used with individual-level data and yields comparable causal estimates.[104, 105]

Sensitivity analysis

For sensitivity analysis, I used weighted median and MR-Egger regression.[104] The weighted-median is a robust alternative even in the event of some violations to the IV assumptions. Assuming that no single SNP contributes more than 50% of the weight to the estimate, this method provides a consistent estimate if at least 50% of the weight comes from valid IVs.[104] In simulation studies, the weighted-median estimation method produces similar point estimates, narrower confidence intervals and better finite-sample Type 1 error rates than the IVW and the 2SLS method.[104]

The MR-Egger method uses a different, still untestable set of assumptions.[104] MR-Egger regression performs well even when 100% of the IVs are invalid.[104] The slope coefficient from the Egger regression method provides an estimate of the causal
effect that is consistent asymptotically even if all the genetic
variants have pleiotropic effects on the outcome; this is known
as the inSiDe assumption—instrument strength independent of
direct effect. A stronger genetic variant should have more
reliable estimates of the causal effect than a weaker variant.
Once the average pleiotropic effect of variants is accounted for
through the intercept term in Egger regression, any residual
dose-response relationship in the genetic associations provides
evidence of a causal effect. A non-null intercept from MR-Egger
indicates an invalid IVW estimate.

Analysis was conducted using the TwoSample MR package version
0.4.14 for R software version 3.5.1, available at
https://github.com/MRCIEU/TwoSampleMR. Analysis was verified
using the MR-base platform version 1.2.1 e646be (05 December
2018), available at http://app.mrbase.org/. This analysis of
publicly available data did not require IRB approval.

Results

Genetic determinants of glycine

Six SNPs identified reached genome wide significance (p<5x10⁻⁸).
All six SNPs found for the exposure were available for the
outcome. One palindromic SNP, rs147007805, was aligned using
the minor allele frequency (MAF). The MAF threshold was set to
0.3. No SNP was in linkage disequilibrium ($R^2<0.001$). The F-statistic for SNPs included in the CAD and MI analyses was 376.75 and SNP specific F-statistic ranged from 33.31 to 1927.22. [106-108]

In MR studies obtaining adequate statistical power is frequently a concern due to the small amount of variation in a phenotypic trait that is typically explained by genetic variants, which is addressed by using large samples for the outcome. In a sample size of 184,000, assuming the genetic variants explain 2% of the variance in glycine, then this study had 80% power to detect an OR of 0.907 [106-108]

**Association of genetically determined glycine with CAD risk**

Table 1 shows the unconfounded estimate of one standard deviation (SD) higher genetically determined glycine was associated with decreased risk of CAD. Using the IVW approach, the odds of CAD was 0.94 per 1 SD increase in glycine 95% CI (0.91-0.98). Sensitivity analysis results obtained using the weighted median and MR-Egger produced similar estimates as the IVW method, as shown in table 1 and figure 1. Figure 1 shows the Wald estimate for each included SNP individually, as well as the MR-Egger and IVW regression results for all SNPs combined. However, the p-value for the MR-Egger intercept was significant suggesting that the IVW estimate is invalid.
The funnel plot in figure 4 does not indicate heterogeneity among the SNPs, however, because so few SNPs are included symmetry is difficult to assess and an invalid SNP may have been included.

Association of genetically determined glycine with MI risk
Table 1 shows that the effect of one standard deviation (SD) higher genetically determined glycine was not clearly associated with decreased risk of MI. Using the IVW approach, the odds of MI were 0.97 lower among those with genetically determined higher levels of glycine than those with lower genetically determined glycine, and was not statistically significant 95% CI (0.93-1.02). Furthermore, sensitivity analysis results obtained using the MR-Egger regression method were not statistically significant, and produced similar estimates of no association between MI and genetically determined levels of glycine, as shown in table 1 and figure 1. However, the p-value for the MR-Egger intercept was significant suggested that the IVW estimate is invalid.

Figure 1 shows the Wald estimate for included SNPs individually, as well as the MR-Egger and IVW regression results for all SNPs combined. The funnel plot in figure 4 does not indicate
heterogeneity among the SNPs, however, the lack of symmetry could be an indication that an invalid SNP was included.
Figure 1 Wald ratio for individual SNPs and MR-Egger and IVW for all SNPs on CAD and MI

MR effect size for 'Gly' on 'Coronary heart disease || id:7'
rs2169387
rs1047891
rs13298772
rs10083777
rs1992855
rs147007805

All − IVW
All − Egger
rs147007805
rs1992855
rs10083777
rs13298772
rs1047891
rs2169387
−0.6 −0.4 −0.2 0.0 0.2
MR effect size for 'Gly' on 'Myocardial infarction || id:798'
Table 1 MR estimates of the associations of log-transformed standardized residuals of Glycine with IHD and with MI

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Outcome</th>
<th># of SNPs</th>
<th>Mendelian Randomization Method</th>
<th>OR</th>
<th>95% CI</th>
<th>Cochran’s Q statistic P-Value</th>
<th>MR-Egger Intercept P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycine</td>
<td>CAD</td>
<td>IVW</td>
<td>0.94</td>
<td>0.905–0.980</td>
<td>0.8743</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighted Median</td>
<td>0.95</td>
<td>0.91–0.99</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MR-Egger</td>
<td>0.97</td>
<td>0.91–1.03</td>
<td>0.9685</td>
<td>-0.0096</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>IVW</td>
<td>0.97</td>
<td>0.93–1.02</td>
<td>0.5967</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighted Median</td>
<td>0.97</td>
<td>0.94–1.07</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MR-Egger</td>
<td>1.00</td>
<td>0.94–1.07</td>
<td>0.7456</td>
<td>-0.013</td>
<td></td>
</tr>
</tbody>
</table>

Two SNPs, rs1047891 and rs2169387 were associated with cholesterol and triglyceride levels, fasting glucose and blood pressure. The remaining SNPs had either no known pleiotropic effects or were associated with factors not associated with heart disease, such as glaucoma. Table 3 in the appendix shows these relations. Estimates were similar after excluding rs1047891 and rs2169387. The IVW estimate CAD was (OR - 0.8861, 95% CI 0.788–0.996) and the MR-Egger results were (OR - 0.8772, 95% CI 0.771–0.998). The intercept P-value was 0.569 and the Q-statistic P-value was 0.8893.
Discussion

This novel study does not exclude the possibility that genetically determined glycine is associated with a lower risk of CAD. Glycine is most notably found in gelatin and is a major component of collagen, which is abundant in blood vessels and the gut, among other fibrous tissues.[36, 40, 42, 55, 56] Glycine is part of the human diet, with the highest concentrations found in shellfish, white fish, pork, turkey and chicken skin as well as soybeans.[130]

Studies, including in vitro, animal and human observational studies have shown an inverse association of glycine with CAD and CAD risk factors.[42] A 2016 RCT found that glycine was inversely associated with the risk of MI.[56] A stronger protective effect was found in patients with a level of apolipoprotein B, low density lipoprotein cholesterol, or apolipoprotein A-I above the median.[56] Other studies have reported positive effects of glycine on vascular health.[42] Glycine has been reported to stabilize platelets, reduce the sizes of areas of ischemic injury upon reperfusion, correct elevated blood pressure and normalize serum triglycerides.[55, 56]

This study has the advantage of using a novel technique which enables estimates to be made cost-effectively even when no study
including both the exposure and outcome is available. It also has the advantage of providing unconfounded estimates. However, MR has stringent assumptions. First, the relevance assumption requires that SNPs be strongly associated with the exposure of interest.[63] This study used SNPs that were associated with the exposure at a P-value <5 x10^{-8}. Second, the independence assumption requires that SNPs be independent of exposure outcome confounders. This would be expected, as genetic makeup is determined a conception and independently from each other.[64] Third, the exclusion-restriction assumption requires that the SNP affects the outcome only via its effect on the risk factor of interest.[63] Known pleiotropy was assessed using Phenoscanner and few associations with known CAD/MI risk factors were found for the SNPs included in this analysis.[96, 97] Two SNPs, rs1047891 and rs2169387 are known to also be associated with cholesterol and triglyceride levels, fasting glucose and blood pressure. The remaining SNPs had either no known pleiotropic effects or were associated with factors not associated with heart disease, such as height. For additional information please see the table 3 in the appendix. Given the use of summarized data in two samples, serum glycine levels were not measured in the sample with the outcome. However, two-sample instrumental variable analysis is more robust to chance associations than analysis of a single sample.[26]
Making inferences from population based genetic studies, one would be concerned with population stratification[83]; when a population subgroup experiences different disease and SNP frequencies resulting in confounded associations in a whole population study. The exposure and outcome data included people largely of European descent. While this study cannot assess the effect of glycine on CAD/MI by subgroup, there is no reason to think that mechanism by which glycine affects CAD/MI are group specific. Canalization, or developmental compensation occurs when molecular mechanisms compensate for the effect of the genotype of interest on the outcome of interest.[83] Weak instrument bias should not influence results of this study.[83] All included SNPs reach genome wide significance. However, a weak instrument would bias results toward the null.[119] Non-linear effects have successfully been identified using MR, but depend on additional assumptions. I had to assume a linear relationship in this analysis. Participant overlap may have occurred in this two sample MR analysis. However, given the very large sample size of the exposure and outcome datasets, the bias if any should not be of major concern.[26, 120, 121] Finally, MR is better at identifying the direction of effect rather than its magnitude, though, the entire population is
exposed to glycine, making any effect relevant to population health.
Glycine may be associated with a decreased risk of CAD. There has been evidence since as far back as the 1970’s that that the levels of myocardial glycine metabolism change due to myocardial ischemia.[34] Glycine is known to stimulate the release of insulin from the pancreas, and affects glucose homeostasis.[43] Glycine is also been shown to modulate concentrations of nitric oxide, a vasodilator; and in at least one study it was associated with improved blood pressure.[115, 131] However, observational studies have mixed findings. The INTERMAP study showed an inverse relationship of glycine and blood pressure, and the Rotterdam study found no significant associations between the two.[115, 132, 133] A recent study evaluated the effect of glycine on atherogenesis using murine macrophages.[114] Glycine showed clear anti-atherogenic effects, by significantly reducing macrophage triglyceride content, related to decreased VLDL uptake.[114] Glycine was the only amino acid that attenuated both the uptake of the triglyceride-rich VLDL and the triglyceride biosynthesis rate in macrophages.[114] there is some evidence that the effect glycine has on CVD differs by sex. A 2016 study found that there was an inverse relationship between glycine and acute MI and was more favorable among women than men.[134] Several biological
pathways have been hypothesized explaining the potential positive effect that glycine has on CAD, including increasing intracellular antioxidants, thus having a positive effect on vascular health. Glycine intake counteracts many of the adverse effects of a high-sucrose diet on the liver, adipose mass and vascular function in rats.[55] Glycine decreased the fatty acid content of the liver of sucrose-fed rats, corrected an elevation of blood pressure, normalized the serum triglycerides and insulin, and, in the vasculature, boosted glutathione, decreased oxidative stress and normalized endothelium-dependent vasodilation.[55] Glycine may stimulate and glucagon release and decrease membrane polarization in cells.[55] This may be why when glucose was fed to patients with type 2 diabetes in conjunction with gelatin it strengthened the insulin response.[55] Clinically, glp-1 agonist drugs promote modest weight loss in patients, and exert favorable effects on systolic blood pressure, serum lipids, inflammatory markers and endothelial function.[55]
Conclusion

This study does not exclude the possibility that genetically determined glycine is associated with a lower odd of CAD.[46] We replicate a similar study conducted by Wittemans, Laura B. L. et.al. in Nature Communications that used MR to estimate the causal relationship of glycine with CAD, CAD risk factors and diabetes.[124]. Although both studies conducted a MR of glycine on cardiovascular disease Several differences exist between the two, including the use of different MR methods, population sizes, outcomes and homogeneity. The recently published MR study found glycine to be cardioprotective.[124] That same study found that the protective effect of glycine on CAD was mediated via improved blood pressure.[124]

Glycine is generally found in the skin, bones and organs of some animals. Because most people do not eat these parts of an animal, glycine can be modified via dietary supplements along with what is part of a normal diet. Additional research is needed to conclusively establish glycine is statistically significantly associated with lower risk of CAD, and via what mechanisms glycine affects CAD.
Appendix

Table 2 Characteristics of the single nucleotide polymorphisms (SNP) used for genetically determined serum glycine.

<table>
<thead>
<tr>
<th>SNPs</th>
<th>Nearest Gene</th>
<th>Effect Allele</th>
<th>Other Allele</th>
<th>effect</th>
<th>Standard error</th>
<th>MAF</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10083777</td>
<td>CENPN</td>
<td>T</td>
<td>C</td>
<td>-0.10606</td>
<td>0.01</td>
<td>0.827109</td>
<td></td>
</tr>
<tr>
<td>rs1047891</td>
<td>CPS1</td>
<td>A</td>
<td>C</td>
<td>0.487116</td>
<td>0.01</td>
<td>0.666106</td>
<td></td>
</tr>
<tr>
<td>rs13298772</td>
<td>KIAA2026</td>
<td>C</td>
<td>T</td>
<td>0.273079</td>
<td>0.02</td>
<td>0.946912</td>
<td></td>
</tr>
<tr>
<td>rs147007805</td>
<td>MAP2</td>
<td>A</td>
<td>T</td>
<td>-0.14046</td>
<td>0.02</td>
<td>0.925411</td>
<td></td>
</tr>
<tr>
<td>rs1992855</td>
<td>ALDH1L1</td>
<td>C</td>
<td>T</td>
<td>0.061971</td>
<td>0.01</td>
<td>0.590467</td>
<td></td>
</tr>
<tr>
<td>rs2169387</td>
<td>RP11-115J16.1</td>
<td>G</td>
<td>A</td>
<td>-0.13006</td>
<td>0.02</td>
<td>0.13332</td>
<td></td>
</tr>
</tbody>
</table>
Table 3 Select Pleiotropic effects of Relevant Glycine SNPs

<table>
<thead>
<tr>
<th>SNPs</th>
<th>Nearest Gene</th>
<th>Associated Disease</th>
<th>Associated Metabolites</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10083777</td>
<td>CENPN</td>
<td>Homocysteine</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Whole body water mass</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Trunk fat-free mass</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Creatinine</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Platelet</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chronic kidney disease</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weight</td>
<td>Homoarginine</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Illnesses of mother: heart disease</td>
<td>Betaine</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hemoglobin</td>
<td>Sarcosine</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Plateletcrit</td>
<td>Guanidinosuccinate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Homocysteine</td>
<td>Serine</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fibrinogen</td>
<td>Histidine</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hip circumference</td>
<td>Glutamylalanine</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Self-reported hypertension</td>
<td>Arginine</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Self-reported migraine</td>
<td>Creatinine</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chronic kidney disease</td>
<td>Threonine</td>
</tr>
<tr>
<td>rs1047891</td>
<td>CPS1</td>
<td>Impedance of arm right</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Impedance of whole body</td>
<td></td>
</tr>
<tr>
<td>rs13298772</td>
<td>KIAA2026</td>
<td>Impedance of arm right</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Impedance of whole body</td>
<td></td>
</tr>
<tr>
<td>rs147007805</td>
<td>MAP2</td>
<td>Heel bone mineral density</td>
<td>Free cholesterol in medium HDL</td>
</tr>
<tr>
<td>rs1992855</td>
<td>ALDH1L1</td>
<td>LDL cholesterol</td>
<td>Acetoacetate</td>
</tr>
<tr>
<td>rs2169387</td>
<td>RP11-115J16.1</td>
<td>Neutrophil count</td>
<td>3Lhydroxybutyrate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Granulocyte count</td>
<td>ApoA1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Neuroticism</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Illnesses of father: heart disease</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Irritability*</td>
<td></td>
</tr>
</tbody>
</table>

Note : * 1x10^-5
Figure 2 Leave one Glycine SNP out Sensitivity Analysis for CAD

MR leave-one-out sensitivity analysis for 'Gly' on 'Coronary heart disease || id:7'
MR leave-one-out sensitivity analysis for 'Gly' on 'Myocardial infarction || id:798'
Figure 3 Plots of SNPs on glycine against SNPs on CAD/MI
SNP effect on Myocardial infarction || id: 798

MR Test

- Inverse variance weighted
- Weighted median
- MR Egger

SNP effect on Gly
Figure 4 Funnel plot of Relevant SNPs to assess heterogeneity of the Glycine on CAD and MI Analysis
References


100 Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using


114 Grajeda-Iglesias C, Aviram M. Specific Amino Acids Affect Cardiovascular Diseases and Atherogenesis via Protection against


Observational epidemiologic studies are susceptible to confounding and selection bias.\cite{103, 135} The strength of a Mendelian randomization (MR) analysis is that it leverages the natural randomization that is inherent in the assignment of genetic information at birth increasing the robustness of causal inference.\cite{103, 135} Observational studies and RCTs have contributed to the reduction of CHD and MI that has occurred in western settings in recent decades. Despite these improvements, cardiovascular diseases, including CHD and MI remain the leading causes of death.\cite{60} Opportunity remains to identify new casual factors and treatments for CAD/MI. This study used MR to estimate the effects of three amino acids on CAD/MI.

This study is the first MR study to assess the relationship of CAD/MI with alanine and glutamine. This study also replicates an analysis of CAD/MI with glycine, strictly adhering to the MR assumptions. After an extensive literature review, it was hypothesized that alanine, glutamine and glycine would all have a cardioprotective effect. Glutamine, glycine, and alanine have been established in animal studies to be linked with heart function and heart disease.\cite{34-36, 40-42, 45-48} Glutamine is commonly found in nuts, which observational studies and RCTs
have found to be cardio-protective.[44, 49-51] Animal model and small RCT studies have found that glutamine improves postoperative outcomes and restores the heart to normal function more quickly than without glutamine supplements.[38, 47, 48] Similarly, glycine studies have shown an inverse relationship with CAD and CAD risk factors such as hypertension. Alanine studies have shown an inverse relationship with diabetes, a risk factor for CAD, in light of the beneficial role that alanine plays in glucose homeostasis and the increased alanine observed during cardiac events.[52-54]

Alanine

Analysis in this dissertation found that alanine might lower the risk of CAD. The unconfounded estimate of one standard deviation (SD) higher genetically determined alanine was associated with lower risk. Using the IVW approach, the odds of CAD was 0.89 lower among those with genetically determined higher levels of alanine than those with lower genetically determined alanine and was statistically significant (95% confidence interval (CI) (0.80-0.99) and the odds of MI was 0.90 lower per SD increase of alanine (95% CI (0.800-1.015). Sensitivity analysis results obtained using the weighted median and MR-Egger regression method were not statistically
significant, though the effect was in the same protective
direction as was found in the IVW method for both CAD and MI.
Beta-alanine is known to increase muscle carnosine levels.[35, 36, 52, 53, 85, 113, 116] Carnosine is widely considered an
important anti-glycating agent, potentially improving energy
metabolism and protein homeostasis.[113] Improved handling of
glycation end products has had promising results with diabetes
complications, a CAD risk factor.[113] Additionally, carnosine
has antioxidative properties, which are thought to improve heart
health.[90, 113]

Glutamine
Analysis in this dissertation found that one standard deviation
(SD) higher genetically determined glutamine was not clearly
associated with CAD or MI. Using the IVW approach, the odds of
CAD was 1.09 higher per 1 SD increase in glutamine but was not
statistically significant 95% confidence interval (CI) (0.99–
1.19) and the odds of MI was 1.05 higher per one SD increase in
glutamine but was not statistically significant 95% confidence
interval (CI) (0.99-1.19). Sensitivity analysis results obtained
using the weighted median and MR-Egger regression method were
statistically significant for both CAD and MI, with a similar
odds ratio estimate as the IVW method. The sensitivity analysis
also indicated that the IVW estimate for CAD and MI might be
invalid because the p-value for the MR-Egger intercept was significant.

The role of glutamine in CAD and MI is unclear. A study found glutamine a reliable discriminator of CAD and non-CAD among cardiac catheterization patients. [114, 122] Glutamine has also been associated with plaque development and intima-media thickness. [114, 115, 123] Some in vitro and animal studies have found that glutamine levels have been associated with atherosclerosis and CAD risk. [114] Glutamine stimulates the accumulation of triglycerides in microphages by affecting the SREBP-1, a key regulator in cellular triglyceride biosynthesis. [114] Furthermore, glutamine stimulates microphage oxidative stress, affecting lipid metabolism. [114] The direction of effect that glutamine has on CAD and MI remains unclear and should be conclusively evaluated in a large RCT study to confirm that glutamine is a risk factor for CAD and MI.

Glycine

Analysis in this dissertation found that one standard deviation (SD) higher genetically determined glycine was not clearly associated with increased risk for CAD or MI. Using the IVW approach, the odds of CAD was 0.94 per 1 SD increase in glycine 95% CI (0.91–0.98) and odds of MI was 0.97 lower among those with genetically determined higher levels of
glycine and was not statistically significant 95% CI (0.93-1.02). Sensitivity analysis results produced similar estimates however, the p-value for the MR-Egger intercept was significant suggesting that the IVW estimate is invalid.

Glycine may protect against CAD. Several biological pathways have been hypothesized explaining the potential positive effect that glycine has on CAD, including increasing intracellular antioxidants. Glycine intake counteracts many of the adverse effects of a high-sucrose diet on the liver, adipose mass and vascular function in rats and decreased the fatty acid content of the liver of sucrose-fed rats, corrected an elevation of blood pressure, normalized the serum triglycerides and insulin, and, in the vasculature, boosted glutathione, decreased oxidative stress and normalized endothelium-dependent vasodilation.[55] Glycine may stimulate and glucagon release and decrease membrane polarization in cells.[55] This may be why when glucose was fed to patients with type 2 diabetes in conjunction with gelatin it strengthened the insulin response.[55] This analysis replicates the study conducted by Wittemans, Laura B. L. et.al., but uses IVW rather than a weighted median approach to obtain unconfounded estimates.[124] The IVW method strictly adheres to the assumptions required of a valid MR analysis, possibly explaining the difference in the
significance of the effect observed in the two analyses. The recently published MR study found glycine to be cardioprotective.\[124\] My analysis could not rule out the possibility of a cardioprotective effect, however additional research is needed to conclusively establish glycine as cardioprotective, and via what mechanisms glycine affects heart disease.

Limitations
This study has several limitations. MR analysis makes a number of assumptions, which if incorrect can affect the results and subsequent conclusions. There is an unlikely possibility for confounding by population stratification. However, the sample used for this study is mostly homogeneous, made up overwhelmingly of people of European descent. Pleiotropic effects, or violation of the exclusion-restriction assumption, occur when the chosen genetic variant has a direct effect on the outcome of interest via a pathway other than the exposure of interest. The two-sample analytic approach means I used summary data about the relationship of the genetic variant to the exposure of interest from one sample and summary data about the relationship between the genetic variant and the outcome of interest to determine the relationship between the exposure and outcome. By design, the exposure of interest and the outcome are
not measured in the same sample. This limits my ability to perform a rigorous subgroup analysis. However, the two-sample method is more robust to chance associations than one sample analyses. Additionally, use of summary data from two samples may result in sample overlap, which could bias the findings. The sample size of both data sources is very large, and bias due to sample overlap is not of major concern. The use of summary data also makes it difficult to test whether associations vary by sub-group. Although causal effects should be consistent they may not be relevant in all populations.[136] Similarly, Mendelian randomizations estimates for the effects of these amino acids may vary according to their levels in different populations. Finally, the NO Measurement Error (NOME) assumption, specifically that there is no measurement error in the genetic variant and exposure association, cannot be tested directly. An adaptation of the I² statistic from the field of meta-analysis is proposed to quantify the strength of NOME violation for MR-Egger.[127]

One additional limitation of this study, and MR in general, is that in most instances it specifies the endogenous effect on the outcome when most times we are interested in the exogenous effect.
Contributions

The dissertation contributes to knowledge of CAD and MI by using a novel analytic method to obtain unconfounded estimates for three amino acids where the epidemiological, clinical, and in vitro studies have been inconclusive. It has the advantage of using a novel technique which enables estimates to be made cost-effectively even when no study including both the exposure and outcome is available. MR is not well suited for assessing the magnitude of the effect, but rather the direction of the effect. The statistically significant lower risk of CAD with higher alanine requires replication to more clearly understand the magnitude and pathway by which alanine affects CAD in order to develop targeted treatments and pharmacotherapies.[76] The analyses for glutamine and glycine did not have statistically significant findings, and is a call for additional studies, possibly with larger samples, or stronger genetic instruments.
Citations


10 Centers for Disease C, Prevention. Heart Disease Fact SheetCDC.  


54 Schooling CM, Kelvin EA, Jones HE. Alanine transaminase has opposite associations with death from diabetes and ischemic


120 Kodali HP, Pavilonis BT, Schooling CM. Effects of copper and zinc on ischemic heart disease and myocardial infarction: a


