

CURRICULUM VITAE

Kevin TK Nguyen

Education

University of Maryland, Baltimore 2024
Ph.D. Molecular Medicine: Genome Track

University of Maryland, Baltimore County 2016
BS Bioinformatics

Awards And Scholarships

Pre-doctoral Research Fellowship, Epidemiology of Aging Training Program,
Department of Epidemiology and Preventive Medicine, School of Medicine, University of
Maryland Baltimore; funded by the National Institute on Aging, National Institutes of
Health (T32 AG00262) 2023-2024

NIDA Scholar
University of Maryland, Baltimore County 2014-2016

Meyerhoff Scholars Program
University of Maryland, Baltimore County 2012-2016

Professional Experience

Graduate Research Assistant. University of Maryland, Baltimore, Baltimore, MD
2019 - 2024

Pre-Doctoral student researching the causal effects of conventional and novel risk
factors in age of stroke onset

Bioinformatics Engineer, J. Craig Venter Institute, Rockville, MD 2016 - 2019

Assist in the GCID Bacterial Projects by assembling genomes and providing
statistical analysis

Developing and implementing of new informatics tools as well as pipeline

Research Intern, UMBC, Baltimore, MD 2014-2015

Involved in a collaborative project between two schools to develop a prognosis
tool for Neonatal care

Used Perl to develop an application that sorts and searches various ontologies and
database based on given parameters

Publications

- 1) **Nguyen, K.** and Mitchell, B.D. (2024), A Guide to Understanding Mendelian
Randomization Studies. Arthritis Care Res. <https://doi.org/10.1002/acr.25400>

- 2) Alhassan, E., **Nguyen, K.**, Hochberg, M.C. and Mitchell, B.D. (2024), Causal Factors for Osteoarthritis: A Scoping Review of Mendelian Randomization Studies. *Arthritis Care Res*, 76: 366-375. <https://doi.org/10.1002/acr.25252>
- 3) Xu H, **Nguyen K**, Gaynor BJ, et al. Exome Array Analysis of 9721 Ischemic Stroke Cases from the SiGN Consortium. *Genes (Basel)*. 2022;14(1):61. Published 2022 Dec 24. doi:10.3390/genes14010061
- 4) Nguyen LP, Park CS, Pinto NA, Lee H, Seo HS, Vu TN, Mai H, Pham AHT, Jang E, Cho YL, Goglin K, **Nguyen K**, White R, D'Souza R, Fouts DE, Yong D. In Vitro Activity of a Novel Siderophore-Cephalosporin LCB10-0200 (GT-1), and LCB10-0200/Avibactam, against Carbapenem-Resistant *Escherichia coli*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, and *Pseudomonas aeruginosa* Strains at a Tertiary Hospital in Korea. *Pharmaceuticals (Basel)*. 2021 Apr 16;14(4):370. doi: 10.3390/ph14040370. PMID: 33923801; PMCID: PMC8072773.
- 5) Boonyasiri A, Jauneikaite E, Brinkac LM, Greco C, Lerdlamyong K, Tangkoskul T, **Nguyen K**, Thamlikitkul V, Fouts DE. Genomic and clinical characterisation of multidrug-resistant carbapenemase-producing ST231 and ST16 *Klebsiella pneumoniae* isolates colonising patients at Siriraj hospital, Bangkok, Thailand from 2015 to 2017. *BMC Infect Dis*. 2021 Feb 4;21(1):142. doi: 10.1186/s12879-021-05790-9. PMID: 33541274; PMCID: PMC7859894.
- 6) Duan Y, Llorente C, Lang S, Brandl K, Chu H, Jiang L, White RC, Clarke TH, **Nguyen K**, Torralba M, Shao Y, Liu J, Hernandez-Morales A, Lessor L, Rahman IR, Miyamoto Y, Ly M, Gao B, Sun W, Kiesel R, Hutmacher F, Lee S, Ventura-Cots M, Bosques-Padilla F, Verna EC, Abraldes JG, Brown RS Jr, Vargas V, Altamirano J, Caballería J, Shawcross DL, Ho SB, Louvet A, Lucey MR, Mathurin P, Garcia-Tsao G, Bataller R, Tu XM, Eckmann L, van der Donk WA, Young R, Lawley TD, Stärkel P, Pride D, Fouts DE, Schnabl B. Bacteriophage targeting of gut bacterium attenuates alcoholic liver disease. *Nature*. 2019 Nov;575(7783):505-511. doi: 10.1038/s41586-019-1742-x. Epub 2019 Nov 13. PMID: 31723265; PMCID: PMC6872939.
- 7) Brinkac LM, White R, D'Souza R, **Nguyen K**, Obaro SK, Fouts DE. Emergence of New Delhi Metallo- β -Lactamase (NDM-5) in *Klebsiella quasipneumoniae* from Neonates in a Nigerian Hospital. *mSphere*. 2019 Mar 13;4(2):e00685-18. doi: 10.1128/mSphere.00685-18. PMID: 30867330; PMCID: PMC6416368.
- 8) White RC, Torralba M, Colt K, Harrison F, Goglin K, **Nguyen K**, D'Souza R, Bristow CC, Ellis O, Soge OO, Klausner JD, Fouts DE. Closed Genome Sequences of Clinical *Neisseria gonorrhoeae* Strains Obtained from Combined Oxford Nanopore and Illumina Sequencing. *Microbiol Resour Announc*. 2019 Feb 28;8(9):e00072-19. doi: 10.1128/MRA.00072-19. PMID: 30834380; PMCID: PMC6395865.
- 9) Becka SA, Zeiser ET, Barnes MD, Taracila MA, **Nguyen K**, Singh I, Sutton GG, LiPuma JJ, Fouts DE, Papp-Wallace KM. Characterization of the AmpC β -Lactamase from *Burkholderia multivorans*. *Antimicrob Agents Chemother*. 2018 Sep 24;62(10):e01140-18. doi: 10.1128/AAC.01140-18. PMID: 30012762; PMCID: PMC6153817.
- 10) Becka SA, Zeiser ET, Marshall SH, Gatta JA, **Nguyen K**, Singh I, Greco C, Sutton GG, Fouts DE, LiPuma JJ, Papp-Wallace KM. Sequence heterogeneity of the PenA

- carbapenemase in clinical isolates of *Burkholderia multivorans*. *Diagn Microbiol Infect Dis*. 2018 Nov;92(3):253-258. doi: 10.1016/j.diagmicrobio.2018.06.005. Epub 2018 Jun 18. PMID: 29983287; PMCID: PMC6173980.
- 11) McDermott, J., **Kevin T. Nguyen**, Mead, J. Sequence of *Wolffia australiana* thioredoxin M-type EST clone 35KN4.10, 2012

Presentations

ISGC 30th Workshop, Munich, Germany. “Genetically Predicted CD40 Levels Have Differential Effects on Early and Late Onset Ischemic Stroke Subtypes” April 2024

45th Annual Graduate Research Conference, Baltimore, MD. “Genetic Risk of BMI, Blood Pressure, and Smoking is More Strongly Associated with Early Than Late Onset Ischemic Stroke” March 2023

ISGC 27th Workshop, Bordeaux, France. “Genetic Risk of BMI, Blood Pressure, and Smoking is More Strongly Associated with Early Than Late Onset Ischemic Stroke” September 2022

ABSTRACT

Dissertation Title: Differential Impact of Conventional and Novel Risk Factors on Early and Late Onset Stroke

Kevin Nguyen, Doctor of Philosophy, 2024

Dissertation Directed by:

Braxton Mitchell, PhD, MPH

Professor

Department of Medicine and Department of Epidemiology
and Public Health

Background: Ischemic stroke (IS) is the leading cause of disability and the 4th leading cause of death in the US. It is a multifactorial disease with genetic and environmental components that influences an individual's risk of stroke. Thanks to increased awareness of stroke risk factors as well as advancement in intervention and treatment, there has been a substantial fall in stroke incidence and improvement in patient outcome in high-income countries. However, numerous studies have highlighted a divergent temporal trend among young adults; stroke incidence rate among young adults has steadily risen thus far in the 21st century. In tandem to this concerning trend, young adults are experiencing a higher prevalence of conventional stroke risk factors due to sedentary lifestyle and other behavioral changes.

Objective: The goals of this study were to compare and contrast the impact of conventional and novel risk factors on early (EOS, onset 18-59 years) and late (LOS, onset > 60 years) onset stroke. This study has two specific aims:

- 1) To demonstrate and compare causal associations of five modifiable stroke risk factors (blood pressure, body mass index, type 2 diabetes, hyperlipidemia, and smoking) within EOS and LOS.
- 2) To assess causality between higher genetic predisposition of inflammation to risk of stroke and identify potential differential effects in EOS and LOS.

Methods: Two-sample MR design was employed to assess the causal relationships between various exposures (previously observed stroke risk factors) and outcome (stroke onset). From publicly available genome-wide association studies' (GWAS) summary results, genetic variants were identified to be used as instrumental variables to proxy levels of the stroke risk factors. Variant-Stroke Onset associations were derived from GWAS performed on the Early Onset Stroke Consortium (n = 40492) and the Stroke Genetics Network (n = 34396). Causal estimates were calculated respectively in the two groups and then odds ratios between EOS and LOS, as well as stratified by TOAST subtypes.

Results: Results from this study suggest that some genetically determined levels of risk factors play a causal role in the increased risk of EOS and LOS. While conventional risk factors do not impact stroke subtypes uniquely, there may be a differential effect attributed to inflammatory biomarkers in EOS and LOS subtypes.

Differential Impact of Conventional and Novel Risk Factors on
Early and Late Onset Stroke

By
Kevin Nguyen

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, Baltimore in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2024

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION AND BACKGROUND.....	1
1.1 Specific Aims.....	2
1.2 Importance and Significance.....	3
1.3 Stroke, Subtype, and Prevalence.....	4
1.4 Conventional Stroke Risk Factors.....	6
1.5 Inflammation and Stroke.....	6
1.6 Mendelian Randomization	8
1.7 Review of Mendelian Randomization Study of Conventional Risk Factors and Inflammatory biomarkers in Stroke.....	11
 CHAPTER 2: THE IMPACT OF CONVENTIONAL STROKE RISK FACTORS ON EARLY AND LATE ONSET ISCHEMIC STROKE: A MENDELIAN RANDOMIZATION STUDY.....	14
2.1 Abstract.....	14
2.2 Introduction.....	16
2.3 Methods.....	17
2.4 Results.....	22
2.5 Discussion.....	24
 CHAPTER 3: DIFFERENTIAL EFFECTS OF INFLAMMATORY BIOMARKERS ON ISCHEMIC STROKE ONSET AND SUBTYPES: A MENDELIAN RANDOMIZATION STUDY	33
3.1 Abstract.....	33
3.2 Introduction.....	35
3.3 Methods.....	36
3.4 Results.....	40
3.5 Discussion.....	46
 CHAPTER 4: SUMMARY AND IMPLICATIONS OF THE RESEARCH.....	51
4.1 Main findings.....	51
4.2 Strength and limitations.....	52
4.3 Research, policy implications, and follow-up.....	54
4.4 Future work and Considerations.....	57
 REFERENCES.....	59

Tables

Table 1: Ischemic Stroke Subtype Definition and Global Percentage Distribution in 2015.....	5
Table 2: Summary of Stroke MR studies using modifiable risk factors.....	11
Table 3: Summary of causal findings in Stroke MR studies using inflammatory biomarkers.....	12
Table 4: Genetic Variants that passed filtering criteria per risk factor and were used in the All-Stroke Analysis.....	28
Table 5: Inflammatory biomarkers implicated with risk of ischemic stroke.....	38
Table 6: IVW Causal Estimate of 17 Inflammatory Biomarkers with EOS and LOS.....	41
Table 7: Comparison of inflammatory biomarker IVW estimate using primary genetic instrument versus cis gene instrument.....	46

Figures

Figure 1: The Three MR Assumptions.....	29
Figure 2: Odds ratio and 95% confidence interval for association of nine stroke risk factors with EOS and LOS for all ischemic strokes.....	30
Figure 3: MR Scatterplots and Causal Estimators of BMI, DBP, SBP, and HDL Association with All Stroke EOS and LOS.....	31
Figure 4A-F: Odds ratio and 95% confidence interval for association of Adiponectin, CD40, CCL2, CRP, Factor VIII, and VWF with EOS and LOS for all ischemic strokes and TOAST subtype.....	42

List of Abbreviations

IS	Ischemic Stroke
SNPs	Single Nucleotide Polymorphism
GWAS	Genome-wide Association Study
MR	Mendelian Randomization
IV	Instrumental variable
EOSC	Early Onset Stroke Consortium
SiGN	Stroke Genetic Network
EOS	Early Onset Stroke
LOS	Late Onset Stroke
CE	Cardioembolic
LAA	Large Artery Atherosclerosis
SAO	Small Artery Occlusion

CHAPTER 1: INTRODUCTION AND BACKGROUND

Ischemic stroke is due to a sudden interruption in the blood supply to the brain, resulting in neurological damage, leading to lasting impairments in mobility and cognitive function or death. Stroke remains a critical public health challenge in the United States, ranking as one of the leading causes of mortality and long-term disability. Every year, roughly 800,000 people in the US have a stroke,¹ with a mortality rate of 24.4 per 100,000,² and 76.0% of survivors have at least 1 disability.³

Fortunately, the US and other high-income countries have seen a steady decline in stroke incidence over the last four decades. From 1987 to 2011, a prospective study of the Atherosclerosis Risk in Communities (n = 15,792) found a 24% overall decline in first time stroke.⁴ Mirroring this, data from the National Center for Health Statistics (n = 4,332,220) indicated that stroke related deaths had also declined between 1975 to 2019; the Age-standardized mortality rate (95% CI) per 100,000 population decreased from 87.5 to 30.9 (women) and 112.1 to 38.7 (men).⁵ However, while these declines in overall stroke incidence are encouraging, these trends do not occur across all ages. Specifically, younger adults (age 18-59) have instead seen an increased incidence of stroke since the 2000s. For example, data from the Nationwide Inpatient Sample (n ≈ 8 million) showed that stroke hospitalization had increased for individuals aged 25 to 44 years (from 16 to 23 per 100,000) and 45 to 64 years (from 149 to 156 per 100,000).⁶

While these diverging trends in stroke incidence can be at least partially attributable to diverging patterns of stroke risk factors between younger and older adults, it is also possible that the impact of the same stroke risk factors differs between younger and older individuals. The goal of this study is to understand how the underlying genetic component

of stroke risk factors impact stroke and whether or not they affect age of stroke onset and stroke subtypes differently.

1.1 Specific Aims

This study has two specific aims that explore the differential impact of stroke risk factors and their relationship with age of stroke onset and subtypes.

- 1) Assess causal relationship of conventional stroke risk factors and age of ischemic stroke onset
 - a. Hypothesis 1.1: The contributions of five modifiable ischemic stroke risk factors (blood pressure, body mass index, type 2 diabetes, hyperlipidemia, and smoking) differ between early (stroke before age 60) and late onset (stroke at age 60 or older) ischemic stroke (EOS and LOS)
 - b. Hypothesis 1.2: The differential effects of stroke risk factors on EOS and LOS cannot be explained by the differential distribution of stroke subtypes
- 2) Identify inflammatory biomarkers with age-dependent and subtype-specific causal effects on ischemic stroke
 - a. Hypothesis 2.1: Inflammatory biomarkers have differential effects on ischemic stroke stratified by age of onset, increasing susceptibility to earlier onset of stroke and specific stroke subtypes.

1.2 Importance and Significance

Stroke has fallen from the third to the fourth most common form of death in the United States over the past 60 years.⁷ This can be attributed to increased education and awareness of stroke risk factors and improvements in vascular risk factor control.⁸⁻¹⁰ However, the decline of stroke incidence has stagnated in the last decade.¹¹ Furthermore, there has been a rise in the incidence rate of stroke affecting adults 18-55 years old^{12,13} associated with a higher prevalence of stroke risk factors in this younger age group.^{14,15} This can be in part due to shifts towards in sedentary behaviors and poorer control and management of cardiometabolic risk factors seen in younger adults.^{16,17}

This adverse trend in stroke incidence among younger adults could exacerbate the burden on the healthcare system and have repercussions in economic productivity. Not only would this younger population experience an impact on quality of life within their prime years, but years of productivity would also be lost along with a heavy toll in terms of healthcare costs. Addressing this rising public health concern of early onset stroke requires a multifaceted approach, including preventive measures, early detection, effective management, and ongoing support for patients and their families. Thus, research into the underlying causes of early onset stroke can contribute to better prevention and intervention strategies.

This study is innovative because it explores the causal effects of stroke risk factors in younger adults. Because the analysis utilizes genetic variants as proxies for the stroke risk factors, the risk factor – stroke outcome relationships can be quantified in the absence of confounding factors such as lifestyle behavior and healthcare. The results observed in this study may allow policymakers and healthcare providers to better target public health and

clinical interventions focused on primordial and primary stroke prevention for young adults.

1.3 Stroke, Subtype, and Prevalence

A stroke occurs when the blood supply to the brain is restricted, usually by a disruption in the blood vessels. There are two main types of strokes: 1) ischemic, which accounts for approximately 87% of cases and occurs when a blood clot obstructs vessels that supply blood to the brain, and 2) hemorrhagic, which arises from the rupture of a weakened blood vessel, causing bleeding in or around the brain. Some early signs of a stroke include sensory deficits, weakness and mobility dysfunction, and impaired or slurred speech. If left untreated, symptoms may worsen to permanent neurological damage and complications may lead to death. The etiology of stroke is multifactorial and incompletely understood. Stroke risk can be reduced through preventative strategies, such as diet and exercise. However, without fully understanding the mechanism of stroke including the genetic risk factors that lead into it, ischemic stroke will continue to be one of the leading causes of disability and death in the US.

Ischemic stroke and Subtypes

IS can be classified based on the etiology and age of onset. The TOAST¹⁸ classification system is a commonly used system for categorizing ischemic strokes into different subtypes based on their etiology and clinical feature. The TOAST classification includes the following subtypes:

Table 1: Ischemic Stroke Subtype Definition and Global Percentage Distribution in 2015

Subtype	Definition	Global Percentage of IS Subtypes¹⁹
Cardioembolic (CE)	occurs when a blood clot or debris originating in the heart travels to the brain and blocks a blood vessel.	22%
Large-Artery Atherosclerosis (LAA)	characterized by the presence of atherosclerosis, or the buildup of fatty deposits (plaque), causing a blockage in major arteries leading to the brain.	23%
Small Artery Occlusion (SAO)	involves the blockage of small arteries within the brain.	22%
Other Determined Etiology (OTHER)	This category includes strokes with determined causes that do not fit into the above subtypes. It could include uncommon causes, such as arterial dissection or vasculitis.	3%
Undetermined Etiology (UNDETER)	Some strokes may not have a clearly identified cause despite thorough investigation. These are classified as undetermined etiology.	26%

IS can be further classified into early and late onset. While most strokes occur in individuals over the age of 50 years, approximately 10 to 15% of strokes occur in younger adults aged 18-50 years.²⁰ Subtype distribution varies between the two ages of onset. The most common types of IS in LOS are toastLAA and toastSAO while these two subtypes account for roughly 10% to 20% of stroke in EOS. Studies of stroke subtypes have reported a broad

range and a somewhat higher prevalence (20%–47%) of toastCE among younger stroke patients.²⁰

1.4 Conventional Stroke Risk Factors

Conventional, also referred to as cardiometabolic, risk factors of stroke include obesity, hypertension, hyperlipidemia, smoking, and type 2 diabetes. These risk factors can be attenuated through adjustment in diet, increased exercise, and clinical intervention through medication. Even so, stroke risk factors are on the rise among young adults due to lifestyle changes (i.e., unhealthy diet, sedentary behavior, high stress levels). Data from the National Health and Examination Survey indicate that prevalence in hypertension among US adults aged 20-44 years between 2009 and 2020 rose from 9.3% to 11.5%, prevalence of diabetes rose from 3.0% to 4.1%, and prevalence of obesity rose from 32.7% to 40.9%.¹⁴ Even though the older adults have also seen similar rises in stroke risk factors, this population often have greater success rates in controlling them.¹⁵ In order to counteract the increasing prevalence of risk factors among young adults, earlier detection and re-evaluation of risk factor treatment threshold must be made.

1.5 Inflammation and Stroke

Inflammation plays a significant role in the pathophysiology of ischemic stroke through various mechanisms. Data from experimental studies indicates that inflammation is central in all stages of the initiation, progression, erosion, and rupture of atherosclerotic plaque leading to thrombo-embolic events.²¹ The main pathogenesis of atherosclerosis starts with an imbalance of lipid metabolism and excess within arteries, which in turn exacerbates

immune response, and results in the chronic inflammation of the arterial walls. Hypercholesterolemia changes endothelial permeability, allowing for increased migration of LDL into the intima. The trapped LDL is oxidized, inducing endothelial cells to release pro-inflammatory mediators and express adhesion molecules. Immune cells such as monocytes sequester to the site, absorb the LDL, become foam cells, and form fatty streaks along the walls. As this process repeats, plaque is formed, which thickens the walls and decreases the flow of blood. Further build up or damage to the plaque makes it unstable and prone to rupture, leading to pieces of plaque breaking off and potentially cause thrombotic events down the line. Greater monocyte/macrophage and T-cell infiltration was associated with ruptured plaques and stroke in multiple studies.²²⁻²⁴ While atherosclerosis is primarily associated with LAA, evidence of cervical, intracranial, or aortic atherosclerosis was found within 79% of patients of all TOAST subtypes.²⁵

Inflammation may contribute to other stroke subtypes through mechanisms alternative to atherosclerotic plaque rupture. For example, atrial fibrillation (AF) is a well-documented risk factor for thrombo-embolic stroke, and both may be mediated through atrial myocardial inflammation. Markers of inflammation, such as C-reactive protein, fibrinogen, intercellular adhesion molecule 1, and interleukin 6, have been associated with AF prevalence, future AF risk, and stroke.²⁶⁻²⁹ An increase in tissue factor expression caused by inflammation can also shift the hemostatic balance leading to a prothrombotic state. Inflammatory cytokines such as IL-1 α , TNF- α , and IL-6 upregulate of tissue factor from monocyte-macrophages, causing increased platelet activation/aggregation and coagulation.^{30,31} Significantly elevated levels of these same cytokines have been found in

patients with cardioembolic stroke subtypes, suggesting a relationship between thromboinflammation and stroke.³²

While there are a multitude of studies that have associated increased inflammation with IS, it is difficult to determine whether or not the relationship is causal or exclude the possibility that inflammation is a consequence of the IS pathology.³³ This distinction is further complicated by the link between traditional stroke risk factors and inflammation.³⁴⁻³⁶

1.6 Mendelian Randomization

Thousands of genome-wide association studies have been performed to identify genetic variations associated with a particular trait, phenotype, or disease. The plethora of estimated effect size can be leveraged to generate predictive risk scores in individuals to assess their genetic predisposition to disease or identify potential drug targets. In recent decades, a new technique called Mendelian Randomization has been utilized to extend these findings into demonstrating causality between exposure risk factors and disease outcome. Assuming that a risk factor has a direct effect on the risk of disease onset or outcome, genetic variants that modulate said risk factors would also have a direct effect on the disease (Gene-Environment Equivalence).³⁷ By leveraging the principles of Mendel's laws of segregation and independent assortment, MR can be equivalent to a natural experiment that examines the relationship of exposure to outcome.³⁷

The MR framework overcomes two of the main challenges that plague traditional observational studies: bias due to reverse causality and unmeasured confounding. Reverse causality bias stems from the difficulty in distinguishing the directionality of the observed association between the intended exposure and the outcome of interest. Unmeasured

confounders are variables that influence both the exposure and the outcome in a study. These variables are often difficult to account for or quantify accurately, and thus lead to bias when estimating the association between traits of interest. MR focuses on the genetic component of the exposure and outcome, and as such, these issues can be circumvented. The flow of cause-and-effect is strictly analyzed through genetics, to the exposure, and onto the outcome, avoiding the bias through reverse causality. In addition, because genetics are established at conception, and are typically unchanged throughout a lifetime, the proxied exposure is minimally unaffected by confounding factors.

MR Causal Estimate

In mendelian randomization, genetic variants are used to estimate the effects of the exposure upon an outcome. This estimate is mathematically equivalent to the ratio of the effect size of the variant-outcome divided by the effect size of the variant-exposure. Using this model, the per variant estimate shows the proportional relationship between the two; i.e., variants with larger effect sizes on the exposure should have larger effect sizes on the outcome. When multiple variants are used in calculating the causal estimate, the set of effect size ratios can be summarized using linear regressions such as the inverse-variance weighted methods.

MR Assumption Testing and Sensitivity Analyses

For the MR analysis to work, the selected genetic variants, also known as instrument variable (IV) must adhere to three assumptions: 1) the IV is strongly associated with the exposure, 2) the IV is not associated with the outcome through confounding pathways, and

3) the IV does not affect the outcome except through the exposure.³⁸ If not properly scrutinized, estimates may be biased or lack the power to detect causal effects.

Variants should be chosen through biological or statistical criteria, reducing the likelihood of using a weak, non-robust IV that explains little variation in the exposure phenotype. The second assumption assumes that there are no unmeasured confounders between the IV and the outcome. Unmeasured confounders could result from population sampling bias and imbalance in allele frequency from different ethnic groups. This can be addressed with careful experimental design that stratifies the discovery GWAS sample by ancestry or social-economic factors as well as adjusting for covariates such as principal components. The third assumption of MR assumes the IV affects the outcome only through their effect on the risk factor of interest. For example, aldehyde dehydrogenase 2 (ALDH2) is a gene that is biologically involved in alcohol metabolism. However, it is also known to influence other behavior such as smoking. Using ALDH2 variants as an IV for an outcome such as cardiovascular health would invalidate the third assumption and bias the causal estimate because both alcohol consumption and smoking usage affect the outcome.

The second and third assumption can be tested by looking at the IVs p-values; a strongly associated genetic variant as indicated by its p-value is less likely to have alternative pathways aside from the intended exposure. Additionally, there are more robust methods, such as MR-Egger, which can provide valid causal inferences under weaker assumptions than the standard IVW method. Although robust methods typically use the term ‘pleiotropy,’ any source of instrument invalidity can be expressed as algebraically equivalent to bias from pleiotropy, and so these methods can help assess sensitivity of findings to instrument invalidity more generally, and not simply invalidity that arises from

horizontal pleiotropy. Although these assessments cannot prove that the independence and exclusion restriction assumptions hold, they can provide evidence regarding the lack of invalidity.

1.7 Review of Mendelian Randomization Study of Conventional Risk Factors and inflammatory biomarkers in Stroke

In previous years, large stroke-focused consortia, such as the Stroke Genetic Network (SiGN)³⁹ and MEGASTROKE,⁴⁰ have been published. The availability of this data and their accompanying GWAS summary statistics have enabled the use of MR in dissecting the risk factor profile of stroke. Table 2 and 3 provide brief overview of previously conducted stroke MR analyses for conventional stroke risk factors and inflammatory biomarkers, respectively.

Table 2: Summary of Stroke MR studies using modifiable risk factors

Modifiable Risk factor	Result Summary	Estimates OR [L95, U95]
Body Mass Index	BMI was not statistically significantly associated with any ischemic stroke subtype (Larsson, Neurology 2017) ⁴¹	BMI-AllStroke 1.11 [0.98-1.27] p = 0.11
Blood Pressure (BP)	BP was a significant risk factor in all ischemic strokes, LAA, CE, and SAO (Georgakis, Neurology 2020) ⁴²	SBP-AllStroke 1.39 [1.33, 1.44] p = 1.9E-60 (10 mmHg) DBP-Allstroke 1.27 [1.23, 1.32] p = 1.2E-42 (5 mmHg)
Type 2 Diabetes (T2D)	T2D was associated with all ischemic strokes, LAA, and SAO. (Georgakis, Neurology 2021) ⁴³	T2D-AllStroke: 1.11 [1.08-1.13] p = 5.5×10-24

Table 2 continued

Smoking Initiation (SmkInit)	Genetic predisposition to smoking initiation was significantly associated with all ischemic stroke, LAA, and SAO, but not CE (Larsson, Ann Neurol. 2019) ⁴⁴	SmkInit-AllStroke 1.24 [1.17-1.33] p = 1.3x10 ⁻¹⁰
HDL	HDL was associated with decreased risk of all ischemic stroke (Yuan S, Ann Neuro 2020) ⁴⁵	HDL-AllStroke 0.89 [0.85, 0.94] p = 4.0x10 ⁻⁴
LDL	LDL was associated only with a LAA (Hindy, Stroke 2018) ⁴⁶	LDL-LAA 1.12 [1.01; 1.24] p = 0.024
Triglycerides (TG)	Genetically predicted increased levels of triglycerides were significantly associated with higher risk of all ischemic stroke, LAA(Yuan S, Ann Neuro 2020) ⁴⁶	TG-AllStroke 1.08 [1.03, 1.14] p = 0.003

Table 3: Summary of causal findings in Stroke MR studies using inflammatory biomarkers

Inflammatory Biomarker	Result Summary	Estimates OR [L95, U95]
C-reactive protein (CRP)	CRP was not associated with risk of IS or subtypes (Lin, Neurological Research 2020) ⁴⁷	CRP-AllStroke 1.01 [0.94,1.09] p = 0.78
Interleukin-6	Genetically downregulated interleukin-6 signaling was associated with lower risk of ischemic stroke (Georgakis, Circulation 2019) ⁴⁸	IL6-AllStroke 0.89 [0.82, 0.97] p = 3.3x10 ⁻³
Interleukin-6 receptor	IL6-R was not associated with risk of IS or subtypes (Lin, Neurological Research 2020) ⁴⁷	IL6R-AllStroke 0.93 [0.87,0.99] p = 0.03
Monocyte chemoattractant protein-1 / C-C motif ligand 2 (MCP-1/CCL2)	Genetic predisposition to higher MCP-1 levels was associated with higher risk of ischemic stroke, LAA, and CE (Georgakis, Circulation 2019) ⁴⁹	MCP-1-AllStroke 1.06 [1.02, 1.10] p = 0.002

Table 3 Continued

Tumor Necrosis Factor	Genetically predicted TNF levels was associated with ischemic stroke (Yuan, eBioMedicine 2020) ⁵⁰	TNF-AllStroke 2.27 [1.50, 3.43] p = 1.0x10 ⁻⁴
Cluster of Differentiation 40 (CD40)	CD40 was associated with decreased risk of LAA (Chong, Circulation 2019) ⁵¹	CD40-LAA 0.73 [0.66, 0.80] p = 1.90x10 ⁻¹⁰
Apolipoprotein(a) (LPA)	LPA increased the risk of large artery atherosclerosis (Chong, Circulation 2019) ⁵¹	LPA-LAA 1.22[1.14, 1.30] p = 3.19x10 ⁻⁹

In summary, ischemic stroke is a major debilitating disease but is preventable through risk factor management and early intervention. While incidence of stroke overall has been declining in the US, incidence among younger adults has increased. Noting that risk factor prevalence has also risen among younger adults, it is possible that the impact of these risk factors is stronger relative to older adults. MR is a methodology that can be used to assess the causal relationships of risk factors on outcomes, minimizing the biases that can affect traditional epidemiological studies. In this study, the MR framework is used to calculate the causal estimates between conventional stroke risk factors and chronic inflammation with the risk of EOS and LOS.

CHAPTER 2: THE IMPACT OF CONVENTIONAL STROKE RISK FACTORS ON EARLY AND LATE ONSET ISCHEMIC STROKE: A MENDELIAN RANDOMIZATION STUDY

2.1 Abstract

Objective: Although stroke incidence is decreasing in older ages, it is increasing in young adults. While these divergent trends in stroke incidence are at least partially attributable to diverging prevalence trends in stroke risk factors, age-dependent differences in the impact of stroke risk factors on stroke may also contribute. To address this issue, we utilized Mendelian Randomization (MR) to assess differences in the association of stroke risk factors between early onset ischemic stroke (EOS, onset 18-59 years) and late onset ischemic stroke (LOS, onset ≥ 60 years).

Methods: We employed a two-sample MR design with inverse variance weighting as the primary method of analysis. From publicly available genome-wide association summary results, we first identified genetic variants to be used as instrumental variables to proxy levels of the stroke risk factors (body mass index (BMI), total, HDL-and LDL-cholesterol, triglycerides, type 2 diabetes (T2D), systolic (SBP) and diastolic blood pressure (DBP), and smoking). We then derived age-specific variant-risk factor association effect sizes for each instrumental variable by performing age-stratified association analyses in the UK Biobank. From these summary results we calculated estimates for stroke risk factors in EOS cases (n = 6,728) and controls from the Early Onset Stroke Consortium and in LOS cases (n = 9,272) and controls from the Stroke Genetics Network. Lastly, we compared odds ratios between EOS and LOS, stratified by TOAST subtypes, to determine if any differences observed between causal estimates could be attributed to differences in the distribution of stroke subtypes.

Results: EOS was associated with higher levels of BMI, DBP, SBP, T2D, and lower levels of HDL (all $p \leq 0.002$) while LOS was associated with higher levels of SBP ($p = 0.0001$).

For all risk factors except LDL- and total cholesterol, unfavorable levels of the risk factor were more strongly associated with EOS than LOS. The causal effect of BMI on stroke was significantly stronger for EOS than for LOS (OR = 1.26 vs 1.03; $p=0.008$). After the subtype-stratified analysis, the difference in causal effect sizes between EOS and LOS for BMI diminished and was no longer significant, suggesting that differences in subtype distribution could account at least partially for these differing causal estimates.

Conclusion: These results support a causal relationship between BMI, blood pressure, T2D, and HDL levels with early onset ischemic stroke and blood pressure levels in late onset stroke. Interventions that target these traits may reduce stroke risk.

Keywords: Ischemic Stroke, Mendelian Randomization, Risk Factors, Young Adults.

2.2 Introduction

Stroke is one of the leading causes of death and disability, with more than 795,000 new and recurrent cases annually in the United States as of 2023.⁵² Despite its high incidence, stroke is largely a preventable disorder. The Global Burden of Disease Study estimated that 91% of stroke burden, measured as disability-adjusted life years, can be attributable to modifiable risk factors and 72% of stroke burden is attributable to clusters of metabolic risk factors, namely hypertension, obesity, hyperglycemia, hyperlipidemia, and renal dysfunction.^{53,54} While most strokes occur in individuals over the age of 50 years, approximately 10% of strokes occur in younger adults aged 18-50 years.⁵⁵ Since the year 2000, incidence rates of ischemic stroke in high-income countries have been declining among older individuals but have been increasing in individuals younger than age 55.¹² While these diverging trends in stroke incidence can be at least partially attributable to diverging patterns of stroke risk factors between younger and older adults, it is also possible that the impact of stroke risk factors differs between younger and older individuals.

To evaluate the impact of modifiable risk factors on ischemic stroke at different ages, we employed Mendelian randomization⁵⁶ (MR) to compare causal associations by age of stroke onset and across stroke subtypes. MR uses genetically predicted levels of traits to serve as proxies. Since alleles are randomly assigned at conception, these genetic proxies are generally independent of the risk factor-outcome relationship and thus not easily subject to reverse causality or confounding factors as seen in observational studies. We hypothesize that the contributions of five modifiable ischemic stroke risk factors (blood pressure, body mass index, type 2 diabetes, hyperlipidemia, and smoking) differ between early and late onset ischemic stroke (EOS and LOS) and that these differences cannot be explained by

the differential distribution of stroke subtypes. Using Mendelian randomization, we estimated the causal effects of these risk factors and compared these estimates between early (age of stroke onset < age 60 yrs.) and late (age of stroke onset \geq 60 yrs.).

2.3 Methods

Study design. This study was performed according to the Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization (STROBE-MR) and the recommended guidelines.⁵⁷ We employed a 2-sample MR design using IVW as our primary method of analysis to assess the causal estimates of conventional stroke risk factors on early and late onset ischemic stroke.

Study sample for Primary Outcome. This study utilizes stroke cases and controls assembled from two large GWAS consortiums: The Early Onset Stroke Consortium⁵⁸ (EOSC) and the Stroke Genetics Network³⁹ (SiGN). Stroke cases in these Consortia underwent brain imaging at each site to exclude diagnoses other than ischemic stroke and to assist with subtype classification. Additional screening was performed in some, but not all, studies to exclude cases believed to be due to a known monogenic cause (e.g., sickle cell disease) or to a known non-genetic cause (e.g., drug use, complications of procedures). Ischemic stroke subtyping was performed using the TOAST criteria¹⁸ in most, but not all, sites.

Consistent with criteria used in the EOSC, we defined early onset stroke for these analyses as cases with stroke onset 18-59 years, and late-onset stroke as those with age at first stroke 60 years or older. Subjects included in this report are restricted to a subset of 6,728 early-onset cases (and 33,764 controls) and 9,272 late-onset stroke cases (and 25,124 controls)

who are of European ancestry and for whom individual-level genotypes were available (**Table S1**).

The genotype data from stroke cases and controls were based on hg38 and imputed using the TOPMed reference panel on the University of Michigan Imputation Server.⁵⁹

Exposure Genetic instrument selection. An important consideration for MR analysis is that the population in which the genetic instrument is developed should be as comparable as possible to the population in which the outcome is measured. To address this issue, we developed two sets of genetic instruments, one in a population < 60 years of age, and the second in a population ≥ 60 years of age. Our strategy was to use a common set of risk factor-associated variants, but then weighing them differently according to their population-specific effect sizes. First, we obtained summary genetic association results from large publicly GWAS available from the GWAS catalog⁶⁰ (<https://www.ebi.ac.uk/gwas/>) for 9 stroke risk factors: body mass index (BMI), systolic blood pressure (SBP), diastolic blood pressure (DBP), total cholesterol (TCHOL), LDL cholesterol (LDL), HDL cholesterol (HDL), triglycerides (TG), type 2 diabetes (T2D), and smoking initiation (SmkInit). Sample sizes for the genome-wide association analyses of each trait ranged from 339,224 to 1,232,091 (**Table S2**). We identified variants from GWAS that were associated with each risk factor at genome-wide significance ($p < 5 \times 10^{-8}$) and selected the most significant variant at each associated locus by removing SNPs in linkage disequilibrium with the lead SNP using the clumping procedure in PLINK with parameters $\text{clump-kb} = 10,000$ and $\text{clump-r}^2 > 0.001$. We assessed the strength of each SNP by calculating its F-statistic, which is a function of the proportion of the variance explained by the genetic instrument, and the sample size.⁶¹ The total number of SNPs obtained from

each GWAS, the number of SNPs pruned at each filtering step, and the corresponding F-statistic are provided in **Table S3**.

Having identified the risk factor variants to include in our instruments, we then estimated age-specific effect sizes (age < 60 yrs and age \geq 60 yrs) for each selected variant by performing an age-stratified genetic association analysis of these variants for each risk factor in the UK Biobank and extracting their effect sizes. This enabled us to obtain MR estimates of the casual effect of the risk factor on outcome in populations of approximately the same age as those in which the genetic instrument-exposure effect sizes were calculated. We utilized data from the UK Biobank, a large ongoing prospective cohort study involving 506,682 participants between 2006 and 2010. Participants were aged 40–69 at the time of recruitment. Study populations were identified using BMI, DBP, SBP, HDL, LDL, TG, TCHOL, Smoking Status, and T2D as phenotypes based on UDI and phenocodes. See **Table S4** for study population size and definition of each phenotype. Details of the UK Biobank genetic association analyses, including analysis models and genetic association results, are provided in the Supplement (Supplementary Info, Methods; **Table S5-13**). We compared effect sizes of the risk factor PRS on the risk factor between individuals < age 60 and individuals \geq age 60 using a paired t-test. We performed a formal comparison of effect sizes across all variants using Cochran's Q and I².

Mendelian Randomization Assumptions To minimize the potential for bias in making causal inferences, SNPs were selected to adhere to three assumptions of valid instrumental variables (**Figure 1**):⁶¹ (1) the SNPs used must be strongly associated with the exposure; (2) the SNPs must not be associated with measured and unmeasured confounders; and (3) the SNPs affect the outcome only through the effects of the exposure (e.g., no horizontal

pleiotropy.) We used the F-statistic as a measure of the strength of the SNP-exposure association (Assumption 1) and performed sensitivity analyses (see below) to assess violations of Assumptions 2 and 3.

Statistical Analyses Following clumping and F-statistic filtering, SNP-exposure (risk factors) association statistics were pulled for selected SNPs to create nine age-specific exposure genetic risk scores instruments. The SNP-outcome (stroke) associations were obtained from genetic association analyses performed on the early and late onset stroke datasets from the EOSC and SiGN. The SNP-outcome associations were calculated in PLINK2⁶² (PLINK v2.00a3.3LM) software for all ischemic stroke and for the TOAST subtypes using logistic regression, controlling for genetic ancestry with principal components 1 through 10 and sex.

We used the random-effects inverse-variance weighted³⁸ (IVW) as the primary method for computing causal estimates of the association of each exposure (risk factor) with stroke. This approach entails calculating a Wald ratio for each SNP by dividing the SNP-outcome association by the SNP-exposure association and then estimating the mean of these Wald ratios, weighing each by the inverse of their variances. We used the random-effects model to adjust for heterogeneity among Wald ratios by accounting for over-dispersion in the regression model. Odds ratios were calculated for both EOS and LOS using the final IVW estimate for the causal association of each risk factor with all stroke and TOAST subtype. The odds ratios for continuous risk factors correspond to the change in odds of stroke corresponding to a one standard deviation change in risk factor. To account for multiple testing, we considered a P-value < 0.0055 ($P < 0.05/9$) to be statistically significant, accounting for the nine different risk factors.

To evaluate whether the IVW estimates comply with the independence and exclusion assumption, we performed several sensitivity analyses. As a measure of pleiotropy, we assessed heterogeneity among the individual Wald ratios from the initial estimates using I^2 and Cochran's Q as well as the intercept taken from the MR-Egger method. We also performed the MR analysis using other methods (e.g., Simple median⁶³, weighted median⁶³, and MR-Egger⁶⁴) that are more robust than the IVW approach against deviation from the MR assumptions. Although these methods have less power, estimates from these analyses that are directional discordant from the IVW estimates could be an indication of the presence of pleiotropy. Cook's distance and the MR Pleiotropy Residual Sum and Outlier⁶⁵ (MR-PRESSO) method were used to identify and remove pleiotropic outliers. SNPs with Cook's distance $> (4/\text{number of SNPs})$ were tagged as outliers and filtered out due to their disproportionate level of influence on the MR models. MR-PRESSO uses a leave-one-out methodology to detect pleiotropic SNPs and quantifies their distortion in the causal estimate. To ensure the instrumental variables used were the same in our EOS and LOS estimates, we removed non-overlapping SNPs before recalculating the IVW estimate. All statistical analyses were done in R (4.0.3; The R Foundation for Statistical Computing) using the MendelianRandomization (0.9.0) and MR-PRESSO packages (1.0).

Homogeneity Test of Causal Estimates in EOS and LOS To compare the difference of the causal estimates between early and late onset stroke groups, we performed a t-test, calculated as the difference between the betas divided by the variance of the difference. For this hypothesis we accounted for multiple testing by adjusting for the number of significantly associated risk factors we tested.

We also evaluated whether any differences in effect sizes of risk factors between early and late onset stroke could be driven by differences in the distribution of stroke subtypes between the two groups. For this analysis, we estimated the causal effects for each subtype and then computed the mean of the differences in effect sizes between early and late onset across the five different subtypes. We then computed the variance of the mean difference and performed a t-test to evaluate the significance of the difference in causal effects while accounting for subtype differences between EOS and LOS.

Data Availability

Individual level data from SiGN, where permitted by participant consent and institutional certification, has been deposited into dbGaP. Summary level GWAS statistics from SiGN and EOSC are available on the Cerebrovascular Disease Knowledge Portal.

Ethical approval The genetic association analyses, which involved deidentified data obtained from the UK Biobank Resource under Application Number 49852, underwent ethical oversight, including the determination by the University of Maryland, Baltimore Institutional Review Board that the study is not human research (IRB #: HF-00088022).

2.4 Results

Causal effects of stroke risk factors on EOS and LOS

From published GWAS and our filtering steps (see **Table S2-3**), we identified genetic instruments comprising 20 - 803 variants for the nine risk factors we analyzed (see **Table 4**). The IVW causal estimates between the stroke risk factors and ischemic stroke, and their corresponding odds ratios (scaled to a 1-standard deviation unit change for continuous risk

factors), are shown in **Figure 2** and **Table S14** for EOS and LOS and stroke subtypes. We identified causal effects on EOS for BMI (OR = 1.26, 95% CI: 1.13-1.40), DBP (OR = 1.39, 95% CI: 1.21-1.60), SBP (OR = 1.47, 95% CI: 1.28-1.69), HDL (OR = 0.82, 95% CI: 0.73-0.93), and T2D (OR = 1.17, 95% CI: 1.06-1.29), all $p \leq 0.002$. The causal effects of LDL, CHOL, and smoking on EOS did not meet criteria for statistical significance. In contrast, only the causal effect for SBP on LOS met criteria for statistical significance (OR = 1.24, 95% CI: 1.11-1.38, $p = 0.0001$).

Comparison of effect sizes between EOS and LOS

For all risk factors except LDL and total cholesterol, unfavorable levels of the risk factor were more strongly associated with EOS than LOS. Heterogeneity testing indicated that the causal effect of BMI on stroke was significantly stronger for EOS than for LOS (OR = 1.26 vs 1.03; $p=0.008$). The causal effects of DBP, SBP, and HDL were also stronger for EOS than for LOS (DBP: OR = 1.39 vs 1.11, $p = 0.016$; SBP: OR = 1.47 vs 1.24, $p = 0.051$; HDL: OR = 0.82 vs 0.90, $p = 0.022$), although none achieved statistical significance at our threshold of $P\text{-value} < 0.01$ for 5 risk factors tested (**Table S14**).

Assessment of the MR assumptions

Weighted median, simple median, and MR-Egger were used as alternative causal estimators, and their estimates remained stable relative to the IVW estimate (**Figure 3**). The MR-Egger intercept indicated no evidence for pleiotropy ($p > 0.05$; **Table S15-16**). There was no strong evidence of heterogeneity among the Wald ratios using I^2 and the Cochran Q test ($I^2 > 50\%$ and $p < 0.05$; **Table S17**).

Subtype-adjusted MR analyses

To evaluate whether the stronger associations of BMI, DBP, SBP, and HDL with EOS could be attributed to differences in stroke subtypes between EOS and LOS, we performed TOAST subtype-adjusted MR analyses of these risk factors. We reasoned that if the EOS vs. LOS differences in risk factor associations were attributable wholly to differences in the distribution of stroke subtypes between EOS and LOS, then there would be no difference in risk factor associations within stroke subtypes. These analyses indicated that the stronger association of BMI with EOS was diminished and no longer statistically significant after accounting for differing subtypes ($p = 0.33$). However, one caveat with these analyses is that stroke subtypes were available on only 81% of the stroke cases, thus diminishing power to detect differences. **(Table S14)**

2.5 Discussion

The contribution of conventional stroke risk factors to the development of ischemic stroke has been established previously through prospective epidemiologic studies^{66,67} and causal effects of these estimated through prior MR analyses.⁶⁸ The novel contribution of our study is our use of age-specific weights to evaluate the impact of these risk factors on EOS and LOS separately, where we show that BMI, blood pressure, T2D, and HDL, but not total cholesterol, LDL and smoking, are significantly associated with EOS, while, in contrast, only SBP is significantly associated with LOS. Our analysis of effect size differences further revealed that that the causal effect of BMI on stroke was significantly stronger for EOS than for LOS. While the differences in effect sizes were not statistically significant, the causal effects of blood pressure and smoking were stronger for EOS than for LOS, while the causal effects for lipids were stronger for LOS than for EOS.

Our results are consistent with prior epidemiologic studies reporting relatively large associations of some conventional stroke risk factors on early onset stroke. For example, in a case-control study of young ischemic stroke (15-49 years old), Mitchell et al.⁶⁹ found obesity to be significantly associated with an increased risk of ischemic stroke in young adults with an odds ratio of 1.57 (1.28 – 1.94). In other epidemiologic studies of BMI, dominated by older onset strokes, odds ratios in the range of 1.02 - 1.30 have been reported.⁷⁰ Similarly, observational studies have shown relatively stronger associations with smoking,^{71,72} and hypertension in younger compared to older adults.⁷³ Lower levels of HDL-cholesterol have also been reported in at least one study to be stronger in older compared to younger individuals (age <65: OR = 0.76 (0.44-1.32), age 65-74: OR = 0.38 (0.22-0.65), and age ≥75: OR = 0.51 (0.27-0.94)).⁷⁴

Our MR estimates are generally in concordance with the prior Mendelian randomization studies, reviewed by Georgakis et al.⁶⁸, finding causal associations of elevated levels of blood pressure and LDL, lower levels of HDL, and smoking and type 2 diabetes with ischemic stroke. The absence of associations of some of these risk factors with LOS in our study may be related to our estimation of age-specific effects or to weak instrument bias given that our instruments were based on genetic association analyses restricted to UK Biobank participants aged 60 and older. One notable exception is that a large MR analysis of BMI did not find evidence for a causal association with stroke.⁷⁵ This partially contrasts with our own findings, where BMI was causal for EOS but was null for LOS. However, this prior study used MEGASTROKE (mean age of stroke = 67.4 yrs), an age range more closely matching our LOS group.

The prevalence of many of the conventional stroke risk factors has steadily risen over the past decades. In a review of the National Health and Examination Survey, Aggarwal et al. reported that between 2009 and 2020 the prevalence in hypertension among US adults aged 20-44 years rose from 9.3% to 11.5%, prevalence of diabetes rose from 3.0% to 4.1%, and prevalence of obesity rose from 32.7% to 40.9%.¹⁴ Concurrent with this rise in stroke risk factors among the young, stroke incidence has increased among younger adults. For example, from 1995 to 2012, US ischemic stroke hospitalization rate increased by 41.5% and 30% for males and females, respectively, aged 35-44 years old.⁷⁶ Among those hospitalized in this age group, the prevalence of traditional risk factors nearly doubled during this time period.⁷⁶ Thus, the increased prevalence of stroke risk factors among the young, combined with the greater impact they have on younger adults, may partly explain the rising incidence of ischemic stroke in this age group.

Like many studies, a major limitation of our study is its restriction to individuals of European ancestry, due primarily to the relatively small contribution of non-European samples to existing genome-wide association studies of stroke risk factors and stroke. Future studies involving non-European samples are urgently needed.⁷⁷ Additionally, the number of stroke cases in each subtype classification was relatively small, thus limiting the power to detect subtype-specific associations. Finally, these results should be interpreted cautiously due to the risk of survival biases. Subjects with high genetic susceptibility to elevated BP, BMI, smoking, etc. may have already died and therefore may not have been included in the outcome cohorts.⁷⁸

In summary, to our knowledge, ours is the first study to assess the causal effects of conventional stroke risk factors separately on early and late onset stroke. We found BMI,

blood pressure T2D, and HDL, but not total and LDL-C and smoking to be causally associated with EOS, while, in contrast, only SBP was significantly associated with LOS. With the exception of total and LDL cholesterol, the causal estimates were generally stronger in EOS than LOS, although only for BMI did the difference in effect sizes achieve statistical significance. Larger studies of this issue, including non-European populations, are needed.

Study Funding

Partial funding was provided by NIH grants R01 NS100178, R01 NS105150, and P30 AG028747, and U01HG011717. Kevin Nguyen was supported by a T32 AG000262 Epidemiology of Aging grant and R01 NS114045 CNV And Stroke (CaNVAS). Dr. Xu was supported by the AHA (Grant 19CDA34760258). Please see Supplemental Materials for the funding and acknowledgements for each contributing study.

Table 4: Genetic Variants that passed filtering criteria per risk factor and were used in the All-Stroke Analysis

Risk Factor	Onset	F-Statistics Average [Min, Max]	# Variants
BMI	EOS	60.89 [28.45, 1426.17]	791
BMI	LOS	60.89 [28.45, 1426.17]	791
DBP	EOS	68.88 [29.58, 850.69]	772
DBP	LOS	68.88 [29.58, 850.69]	768
HDL	EOS	127.4 [29, 4451.54]	312
HDL	LOS	127.4 [29, 4451.54]	312
LDL	EOS	146.96 [29.03, 2967.58]	294
LDL	LOS	146.96 [29.03, 2967.58]	294
SBP	EOS	68.54 [29.69,700.65]	800
SBP	LOS	68.54 [29.69,700.65]	795
SmkInit	EOS	32.43 [17.15, 105.64]	252
SmkInit	LOS	32.43 [17.15, 105.64]	252
T2D	EOS	69.59 [24.52, 833.23]	52
T2D	LOS	69.59 [24.52, 833.23]	52
TCHOL	EOS	147.77 [29, 3099.41]	300
TCHOL	LOS	147.77 [29, 3099.41]	300
TG	EOS	120.97 [28.59, 2173.62]	281
TG	LOS	120.97 [28.59, 2173.62]	281

Figure 1: The Three MR Assumptions, where Z is the IV associated with the exposure, X is the exposure, Y is the outcome, and U is confounder. 1) Relevance Assumption: The IV is strongly associated with the exposure of interest. 2) Independence assumption: there are no confounders of the association between the IVs and the outcome, and 3) exclusion restriction assumption: the IV is not related to the outcome other than via the exposure

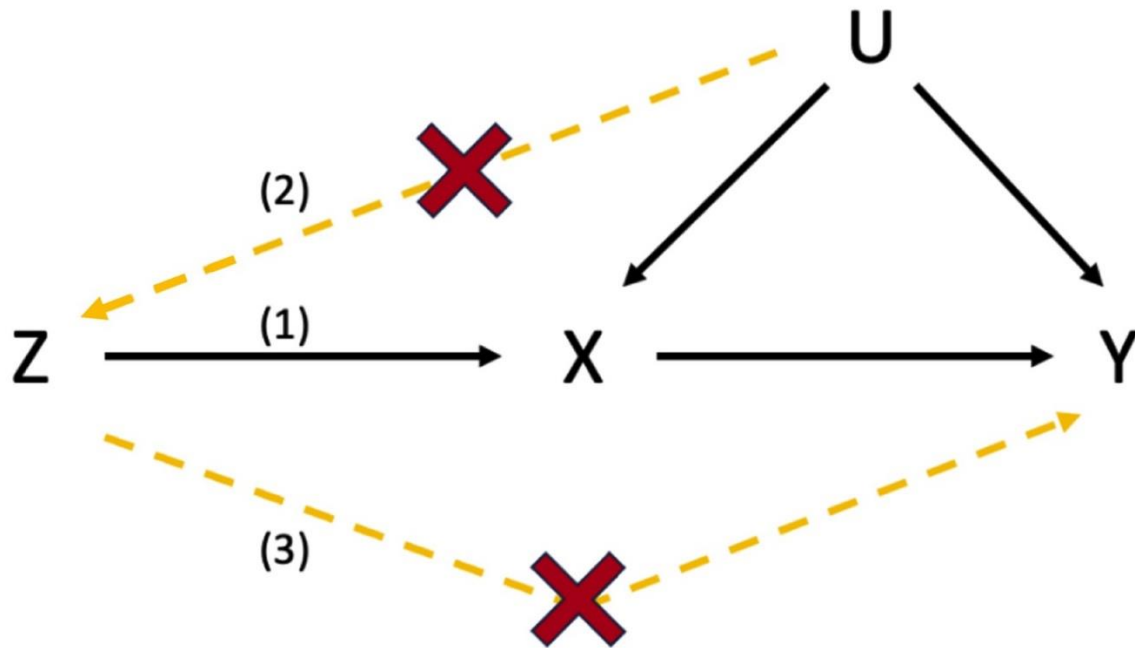


Figure 2: Odds ratio and 95% confidence interval for association of nine stroke risk factors with EOS (blue) and LOS (red) for all ischemic strokes.

** P < 0.01; * P < 0.05; for the heterogeneity test between EOS and LOS

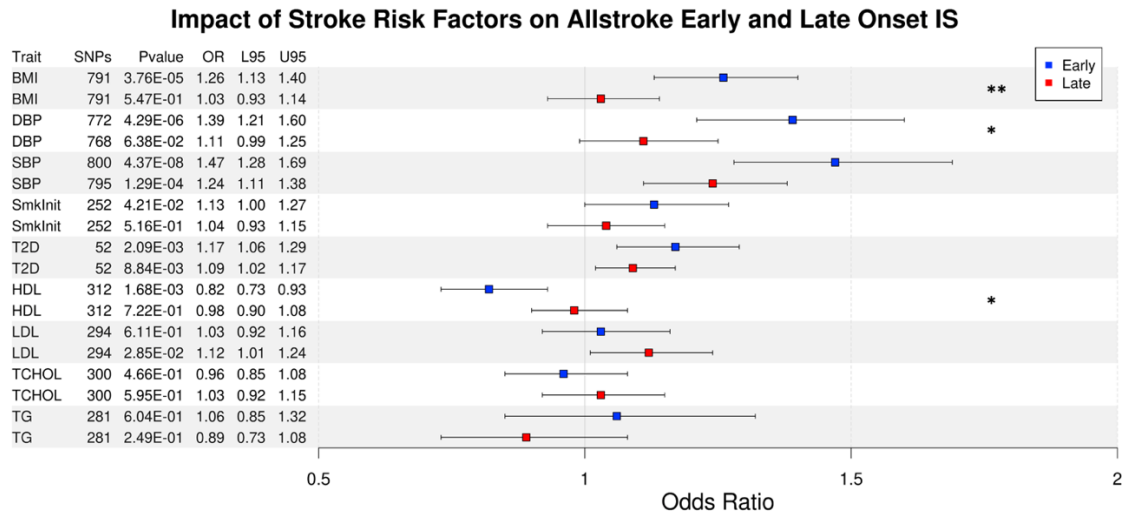


Figure 3: MR Scatterplots and Causal Estimators of BMI, DBP, SBP, and HDL Association with All Stroke EOS and LOS.

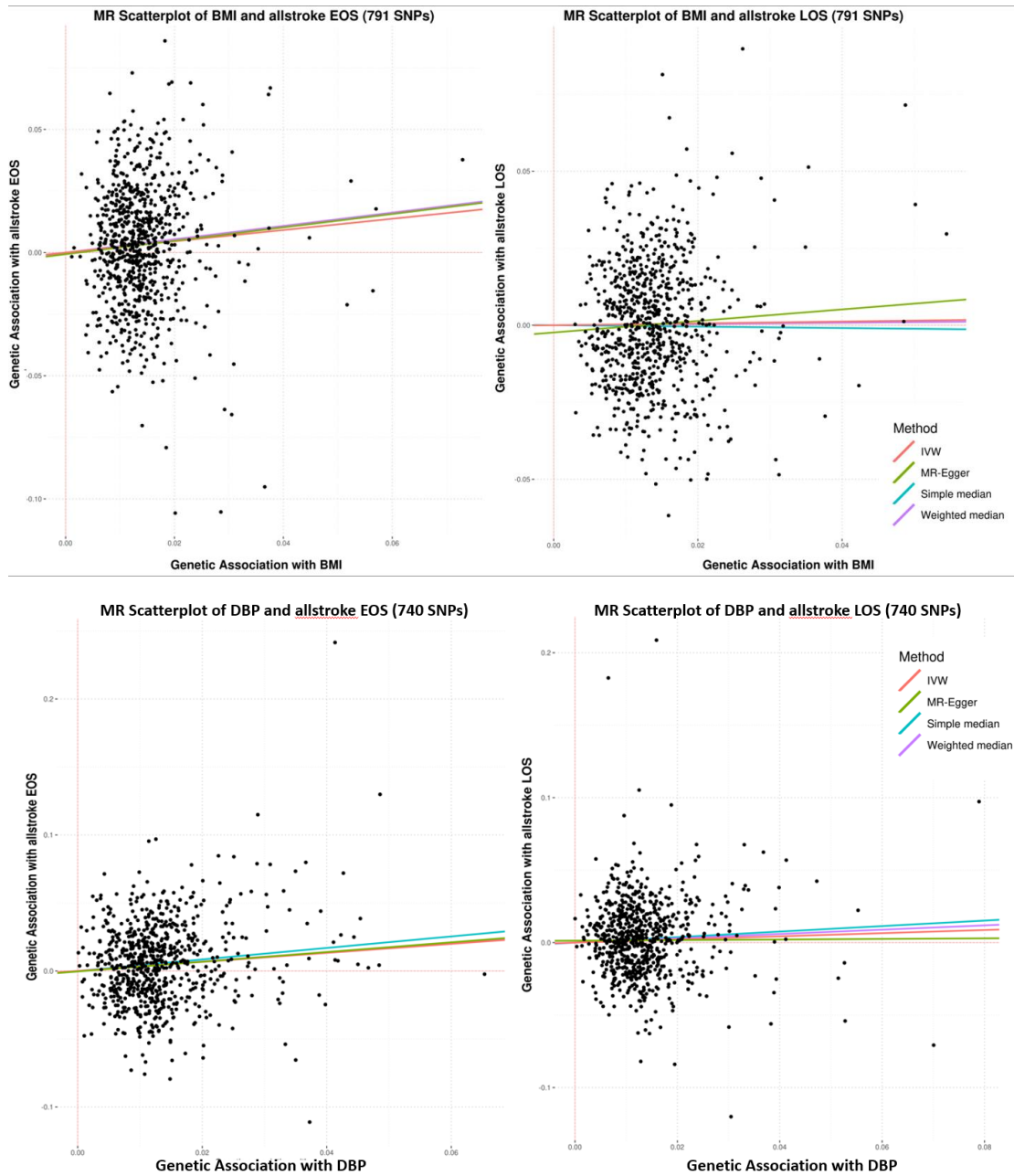
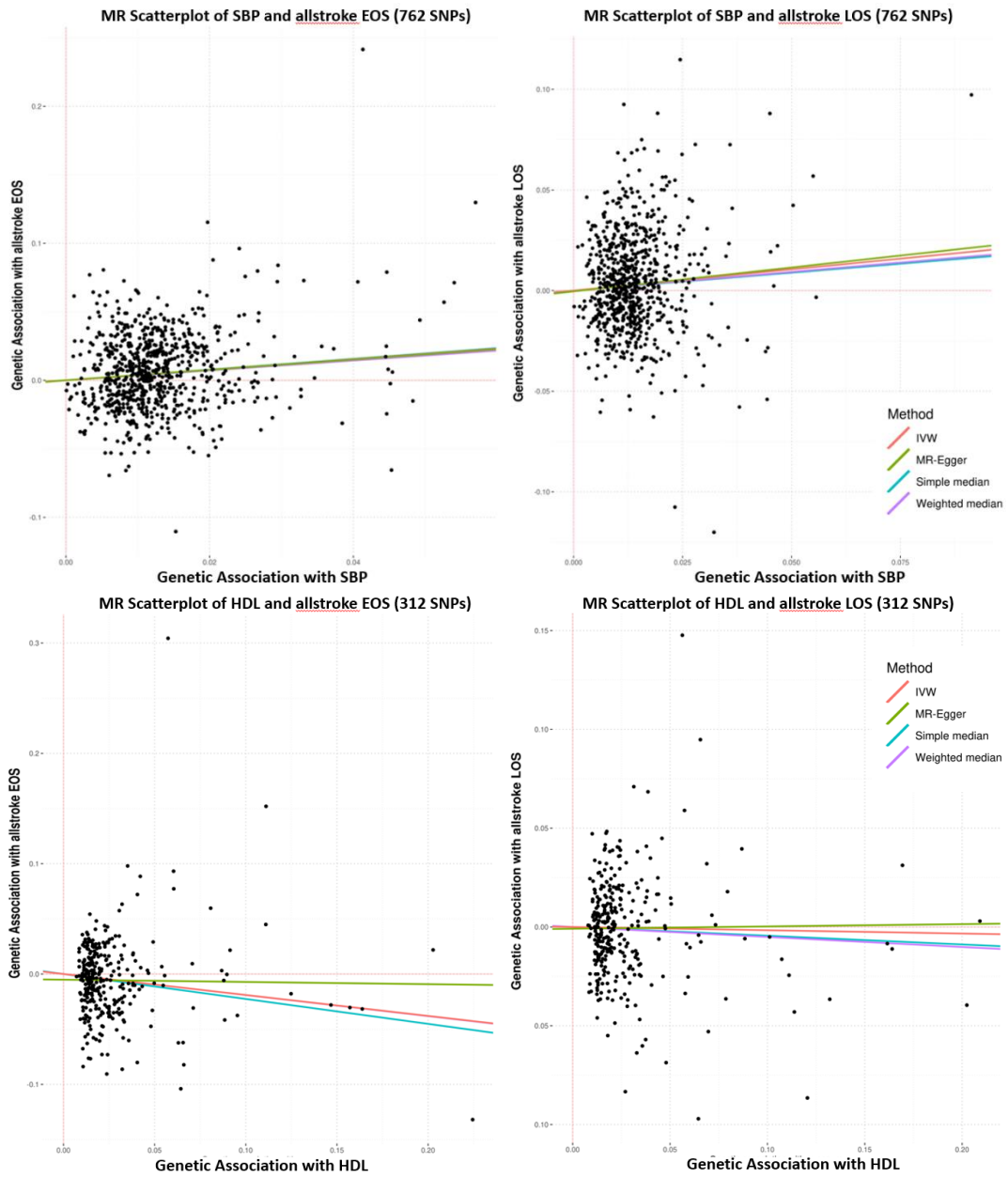


Figure 3 Continued



CHAPTER 3: DIFFERENTIAL EFFECTS OF INFLAMMATORY BIOMARKERS ON ISCHEMIC STROKE ONSET AND SUBTYPES: A MENDELIAN RANDOMIZATION STUDY

3.1 Abstract

Objective: To utilize mendelian randomization to estimate causal effects of inflammatory biomarkers on early and late onset ischemic stroke (EOS, onset 18-59 years; and LOS, onset ≥ 60 years)

Background: Ischemic Stroke (IS) is one of the leading causes of death and disability worldwide. Inflammation biomarkers have been implicated as risk factor for IS through observational studies, but it is difficult disentangling the effects of reverse causality and pleiotropic effects with related cardiometabolic disorders. To assess the causal relationship of these markers to IS, we performed mendelian randomization analyses.

Design/Methods: Genetic instruments were obtained from prior GWAS of 17 inflammatory biomarkers. Genetically predicted marker levels were used to estimate the causal effect between each biomarker and stroke in the Stroke Genetics Network (n = 9,272 cases) and Early Onset Stroke Consortium (n = 6,728 cases).

Results: Lower levels of CD40 were causally associated with EOS and LOS LAA (EOS 0.72 [0.62, 0.84]; p = 1.47E-5; LOS 0.80 [0.71, 0.91]; p = 4.54E-5) while increased levels of CD40 was causal for EOS CE (1.37 [1.17, 1.59]; p = 5.66E-5). Lower levels of adiponectin were causally associated with EOS SAO (0.46 [0.31, 0.66]; p = 3.78E-5), CCL2 was causal for all EOS (1.34 [1.18, 1.51]; p = 2.83E-6), and VWF was causally associated with EOS and LOS CE (EOS 1.57 [1.20, 2.05]; p = 1E-3; LOS 1.33 [1.13, 1.56] p = 5.64E-4). Causal estimates of adiponectin, CCL2, and VWF were attenuated in cis-gene MR analysis to suggestive and null significance.

Conclusions: CD40 has multiple roles in the pathogenesis of IS, being both protective and causal depending on the etiology and age of onset. Further investigation into CD40 is required to disentangle its complex role in inflammation and IS.

3.2 Introduction

The risk factors for stroke encompass non-modifiable elements such as age, sex, and genetic predisposition, as well as potentially modifiable metabolic factors including obesity, hypertension, diabetes, and hyperlipidemia. Among these, inflammation has emerged as a crucial pathophysiological mechanism underpinning ischemic stroke, contributing to the pathogenesis and progression of stroke. For example, inflammation plays a critical role in the development of atherosclerosis, a symptom found in 79% of patients of all TOAST subtypes,^{25,79} and blood coagulation, characterized by platelet rich thrombi that obstruct blood flow in ischemic events and heavily involved in stroke pathogenesis.⁸⁰

Efforts have been made to associate elevated levels of inflammation with risk of stroke by using key players in the inflammatory response include cytokines, chemokines, and adhesion molecules. Observational studies have associated elevated levels of inflammatory markers, such as C-reactive protein (CRP),⁸¹ interleukins (e.g., IL-6),⁸² and tumor necrosis factor-alpha (TNF- α)⁸³ with an increased risk of stroke. However, observational studies are prone to bias due to confounding factors and reverse causality. For example, elevated inflammation could merely be a result of cardiometabolic disorders that predispose individuals to the risk of IS,⁸⁴ and thus it is difficult disentangling inflammation's role in stroke. Furthermore, inflammation plays a significant role in the response to stroke,⁸⁵ running the risk of reverse causality when performing association analyses.

To address these limitations, Mendelian randomization (MR) has emerged as a powerful epidemiological tool that leverages genetic variants as instrumental variables to infer causality between exposures and outcomes.⁸⁶ MR is based on the principle that genetic

variants associated with an exposure (e.g., levels of inflammatory markers) are randomly assorted at conception and thus are less likely to be confounded by environmental and lifestyle factors. We hypothesize that chronic inflammation, represented as elevated levels of inflammatory biomarkers, contribute differently to early and late onset ischemic stroke (EOS and LOS) and stroke subtypes. Using Mendelian randomization, we estimated the causal effects of these biomarkers with IS and compared these estimates between early (age of stroke onset < age 60 yrs.) and late (age of stroke onset \geq 60 yrs) IS.

3.3 Methods

Study sample for Primary Outcome

Study sample for Primary Outcome. This study utilizes stroke cases and controls assembled from two large GWAS consortiums: The Early Onset Stroke Consortium (EOSC)⁵⁸ and the Stroke Genetics Network (SiGN).³⁹ Stroke cases in these Consortia underwent brain imaging at each site to exclude diagnoses other than ischemic stroke and to assist with subtype classification. Additional screening was performed in some, but not all, studies to exclude cases believed to be due to a known monogenic cause (e.g., sickle cell disease) or to a known non-genetic cause (e.g., drug use, complications of procedures). Ischemic stroke subtyping was performed using the TOAST criteria in most, but not all, sites.¹⁸

Consistent with criteria used in the EOSC, we defined early onset stroke for these analyses as cases with stroke onset 18-59 years, and late-onset stroke as those with age at first stroke 60 years or older. Subjects included in this report are restricted to a subset of 6,728 early-onset cases (and 33,764 controls) and 9,272 late-onset stroke cases (and 25,124 controls) who are of European ancestry and for whom individual-level genotypes were available.

The genotype data from stroke cases and controls were based on hg38 and imputed using the TOPMed reference panel on the University of Michigan Imputation Server.

Exposure Genetic instrument selection

For the purpose of this study, we searched the NCBI's Pubmed database for publications that previously implicated inflammatory biomarkers with risk of ischemic stroke, using the search terms: "ischemic stroke" and "inflammatory biomarkers." Longitudinal studies using incidence of stroke as the outcome were prioritized for this analysis due to their direct focus on demonstrating "cause-and-effect" of biomarker to stroke incidence. However, some cross-sectional studies using the prevalent stroke as the outcome were also accepted as they provided evidence of potential biomarker-stroke relationships worth investigating. A total of 17 inflammatory biomarkers, that were identified as statistically significant in their respective study, were selected as the exposures based on the following publications (See **table 5**). Of the biomarkers selected, a wide range of inflammatory functions were represented, including pro-inflammatory signaling, coagulation and hemostasis, and immune cell recruitment and adhesion.

Table 5: Inflammatory biomarkers implicated with risk of ischemic stroke

Function	Biomarker	Supporting Documents (PMID)	Study Type	Summary Statistics Source (PMID)
Pro-Inflammatory Signaling	C-reactive protein (CRP)	9077376	Incidence	35459240
	Interleukin-6 (IL-6)	34969940	Incidence	37563310
	IL-8	10512924	Prevalence	33067605
	Resistin	23329137	Incidence	33067605
	Tumor Necrosis Factor (TNF)	25603656	Prevalence	35078996
Coagulation And Hemostasis	Fibrinogen	36721710	Prevalence	28887542
	Coagulation Factor VIII	10449696	Incidence	29875488
	Von Willebrand factor (vWF)	9409345	Prevalence	35078996
Immune Cell Recruitment and Adhesion	C-C motif ligand 2 (CCL2)	34649381	Prevalence	37563310
	Cluster of Differentiation 40 (Cd40)	12764232	Prevalence	37563310
	CD40 Ligand	12764232	Prevalence	33067605
	E-selectin	23329137	Incidence	33067605
	Intercellular Adhesion Molecule 1 (ICAM-1)	20360547	Incidence	35078996
Inflammatory Response	Adiponectin	24203850	Incidence	37859345
	Lipoprotein-associated phospholipase A2 (Lp-PLA2)	18201705	Incidence	34226706
	N-terminal pro-B-type natriuretic peptide (NT-proBNP)	30786848	Incidence	29237677
	Proenkephalin A	34896817	Incidence	29875488

Once the panel of inflammatory biomarkers was chosen, genetic instruments were developed, using publicly available GWAS summary statistics, derived from populations of European ancestry, representing genetic association with circulating biomarker levels. Genetic variants were chosen based on strength of association with the biomarker of interest, i.e., genome-wide significance ($p < 5 \times 10^{-8}$), followed by LD-clumping to select the most significant variant in an LD bracket (clumping-kb = 10,000 and clumping- $r^2 = 0.1$). Strength of association was also assessed by calculating each variants F-statistic (F-Stat > 10), which is a function of the proportion of the variance explained by the genetic instrument, and the GWAS discovery sample size.

Statistical Analysis

Using the same MR methods described in chapter 2, causal estimates between the inflammatory biomarkers and EOS as well as LOS were calculated, using the IVW as the primary estimator. In addition to the ‘standard’ sensitivity analysis (MR-Egger, median methods, MR-PRESSO), we repeated the MR analysis using a gene-restricted genetic instrument. In the gene-restrict analysis, a subset of each instrument was curated by selectively choosing *cis*-variants to the biomarker’s gene (300 kb upstream and downstream). By focusing the instruments’ specificity to the biomarkers, potential pleiotropic effects due to crosstalk and downstream/upstream signaling pathways may be minimized. To account for multiple testing (17 inflammatory biomarkers), we considered a P-value < 0.00294 ($P < 0.05/17$) to be statistically significant.

3.4 Results

Inflammatory Biomarker Association with EOS and LOS

Using published GWAS listed in table X and our filtering steps, we created 17 instrumental variants with all, but one instrument comprising of 1 to 329 variants. The instrument for CRP stood out as an exception, consisting of 2447 variants. F Statistics for all the variants used ranged from 22.57 to 200.23, demonstrating sufficient strength of association to the biomarkers.

To account for multiple testing (17 biomarkers), we considered a P-value < 0.0029 ($0.05/17$) to be statistically significant. Based on this, elevated levels of CRP and CCL2 were causal for all EOS while elevated levels of factor VIII and lower levels of CD40 were causal for all LOS. Factor VIII and VWF were suggestively causal for all EOS ($p < 0.029$). See table 6 for a complete listing of odds ratio, 95% confidence interval, and p-value with regards to all strokes in EOS and LOS.

Table 6: IVW Causal Estimate of 17 Inflammatory Biomarkers with EOS and LOS

** P < 0.0029 ; * P < 0.029

Trait	# Variants	EOS				LOS			
		OR	L95	U95	P	OR	L95	U95	P
Adiponectin	38	0.87	0.74	1.02	8.57E-02	0.98	0.82	1.18	8.64E-01
CCL2	12	1.34	1.18	1.51	2.83E-06 **	1.16	1.01	1.33	3.45E-02
CD40	24	0.96	0.9	1.01	1.43E-01	0.9	0.85	0.95	7.32E-05 **
CD40L	1	0.96	0.55	1.67	8.76E-01	0.74	0.45	1.23	2.49E-01
CRP	2436	1.18	1.10	1.26	1.80E-06 **	1.00	0.94	1.06	9.12E-01
Eselectin	227	0.97	0.93	1.02	2.38E-01	0.98	0.94	1.02	2.24E-01
factorVIII	6	1.21	1.05	1.38	7.97E-03 *	1.14	1.05	1.24	1.85E-03 **
Fibrinogen	3	0.50	0.15	1.62	2.46E-01	0.54	0.20	1.48	2.29E-01
Icam1	325	0.98	0.96	1.01	1.39E-01	1.01	0.99	1.03	4.77E-01
IL6	1	0.84	0.67	1.07	1.54E-01	0.91	0.73	1.12	3.52E-01
IL8	3	1.16	0.88	1.54	2.91E-01	0.87	0.68	1.12	2.77E-01
Lp-PLA2	2	1.11	0.47	2.65	8.07E-01	0.90	0.41	1.99	8.02E-01
NT-proBNP	2	0.92	0.80	1.06	2.46E-01	0.97	0.85	1.10	6.19E-01
ProenkephalinA	12	0.95	0.90	1.00	6.00E-02	0.98	0.94	1.02	3.90E-01
Resistin	17	1.01	0.91	1.12	8.91E-01	1.04	0.91	1.20	5.76E-01
TNF	1	1.12	0.81	1.55	4.99E-01	1.11	0.84	1.47	4.79E-01
VWF	7	1.17	1.05	1.31	4.09E-03 *	1.08	0.99	1.19	9.33E-02

TOAST Subtype Analysis

TOAST subtype-stratified MR analyses showed differential effects depending on the age of stroke onset and etiology. Adiponectin association with LOS and subtypes were all insignificant ($p > 0.05$). With regards to EOS, lower levels of Adiponectin were causal for SAO EOS (OR = 0.46 [0.31, 0.66]; $p = 3.78E-5$) while elevated levels were instead suggestively causal for CE (**Figure 4A**). Lower CD40 levels were at minimum nominally causal for EOS. However, increased CD40 levels were instead found to be statistically causal for CE EOS (1.37 [1.17, 1.59] $p = 7.32E-5$) while decreased level of CD40 was causal for LAA (0.72 [0.62, 0.84] $p = 1.47E-5$) and OTHER (0.72 [0.62, 0.85] $p = 5.61E-5$) in EOS (**Figure 4B**). Elevated CCL2 levels were statistically causal for CE EOS and LOS (EOS: 1.92 [1.47, 2.51] $p = 1.53E-6$; LOS: 1.38 [1.14, 1.67] $p = 7.91E-3$) (**Figure 4C**) but non-significant for other subtypes. Elevated CRP was only statistically causal for LAA and SAO EOS (LAA: 1.43 [1.21, 1.69] $p = 2.3E-5$; SAO: 1.36 [1.16, 1.59] $p = 2.06E-3$) (**Figure 4D**), factor VIII was only causal for SAO LOS (1.31 [1.11, 1.55] $p = 1.51E-3$; **Figure 4E**), and vWF was only causal for CE EOS (1.57 [1.05, 1.31] $p = 1E-3$; **Figure 4F**). This may suggest that the causal association of biomarkers with EOS and LOS all strokes were mainly driven by their respective causal role in these subtypes.

Figure 4A-F: Odds ratio and 95% confidence interval for association of Adiponectin, CD40, CCL2, CRP, Factor VIII, and VWF with EOS (blue) and LOS (red) for all ischemic strokes and TOAST subtype

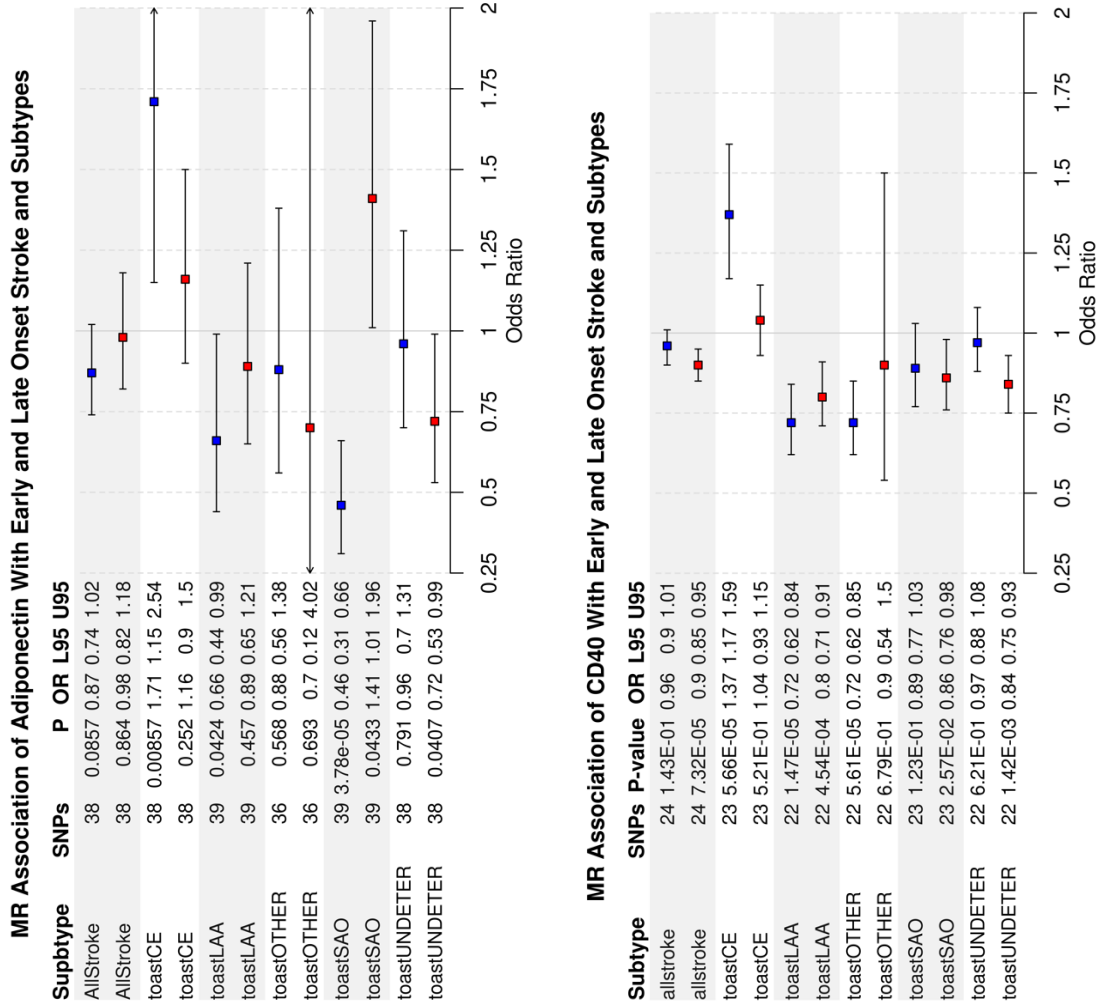


Figure 4A-F Continued

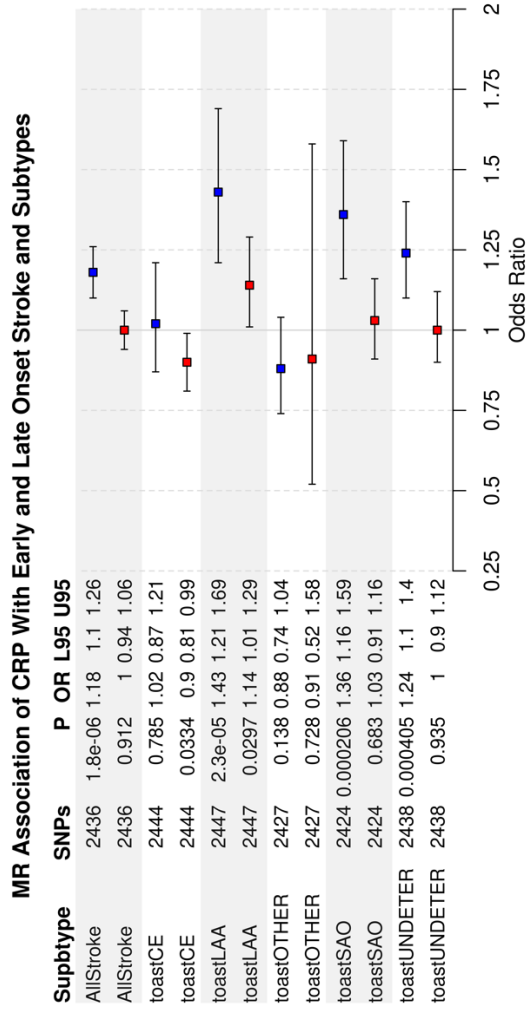
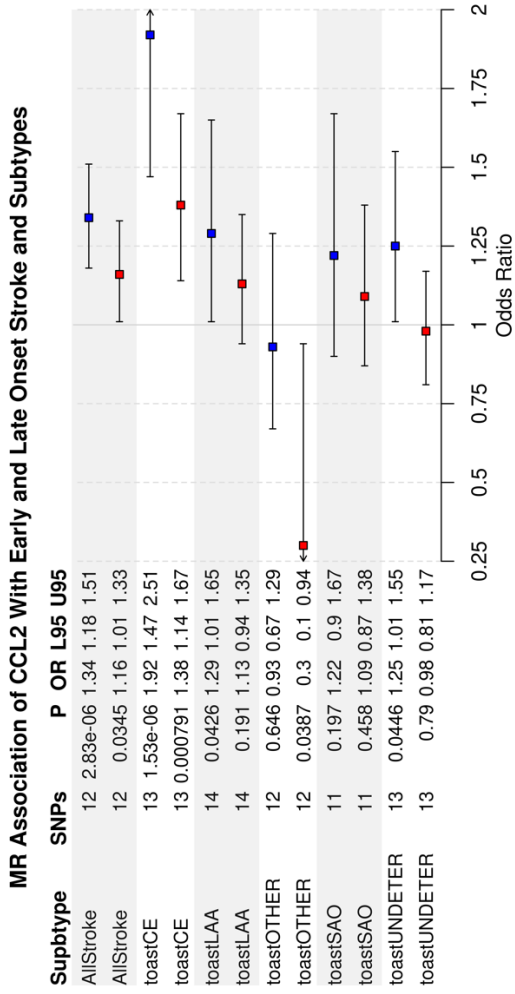
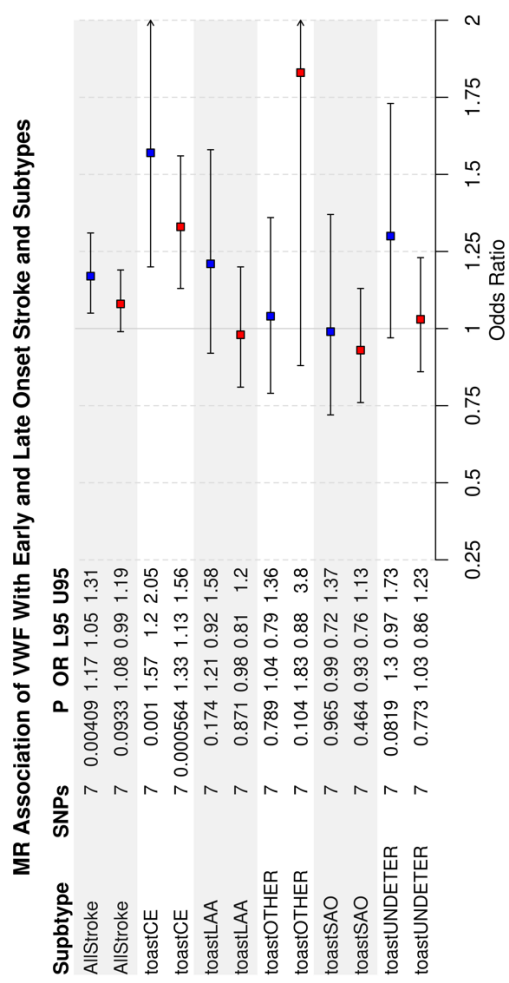
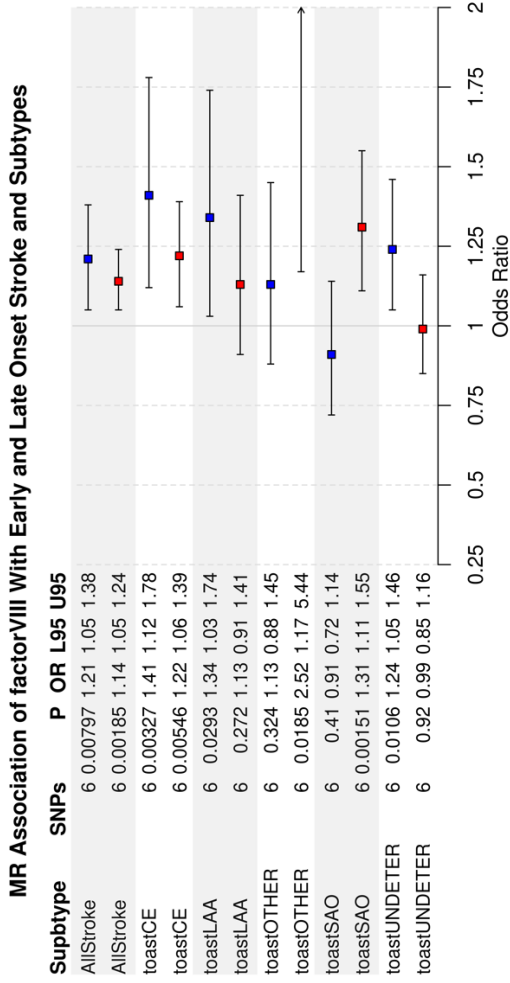


Figure 4A-F Continued



Assessment of the MR assumptions

For the six biomarkers found to be causal in all strokes or specific subtypes (adiponectin, CCL2, CD40, CRP, factor VIII, and vWF), standard MR sensitivity analyses were performed using alternative causal estimators (MR-Egger, Median Method, MRPRESSO). Alternative estimates remained stable among the causal relationships and no indication of significant horizontal pleiotropy was detected. However, this does not entirely rule out undetected pleiotropic effects. Following up the analysis of causal biomarkers, we limited their respective genetic instrument to cis-gene variants to further reduce the risk of pleiotropy. We rationalized that by limiting variants to the specific gene region, we would increase the likelihood that the variants directly influence the expression and concentration of the biomarker of interest. For each causal biomarker-stroke relationship, the cis gene genetic instrument was used to recalculate the IVW, MR-Egger, simple median, and weighed median estimate. A heterogeneity test was performed using a t-test with a significance threshold of $p < 0.05$, indicating the cis-gene analysis deviated from the original MR analysis.

For the majority of the cis-gene causal relationships, the magnitudes of the causal estimates were relatively the same as the original MR analysis, although p-values were attenuated to nominal significance. Notably, the association seen with CRP, EOS, and subtypes (LAA, SAO, and undetermined) diminished to null effect and non-significance, where the heterogeneity test indicated significant deviations, while the effects of CD40 remained consistently strong (**Table 7**).

Table 7: Comparison of inflammatory biomarker IVW estimate using primary genetic instrument versus cis gene instrument.

Cis-Gene MR P-value: ** P < 0.0029, * P < 0.029

Heterogeneity Test P-value: † P < 0.05

Biomarker	Supbtype	Onset	Primary MR					Cis Gene MR					p-value (heterogeneity)
			# Variants	OR	L95	U95	P	SNPs	OR	L95	U95	p	
Adiponectin	toastCE	EOS	38	1.71	1.15	2.54	8.57E-03	13	2.07	1.21	3.53	7.70E-03 *	5.73E-01
Adiponectin	toastSAO	EOS	39	0.46	0.31	0.66	3.78E-05	14	0.50	0.32	0.76	1.18E-03 **	7.74E-01
CCL2	AllStroke	EOS	12	1.34	1.18	1.51	2.83E-06	2	1.80	1.21	2.69	3.94E-03 *	1.64E-01
CCL2	toastCE	EOS	13	1.92	1.47	2.51	1.53E-06	2	2.84	1.00	8.05	4.98E-02 *	4.78E-01
CCL2	toastCE	LOS	13	1.38	1.14	1.67	7.91E-04	2	0.89	0.48	1.66	7.24E-01 *	1.85E-01
CD40	AllStroke	LOS	24	0.90	0.85	0.95	7.32E-05	21	0.91	0.86	0.96	3.53E-04 **	9.85E-01
CD40	toastCE	EOS	23	1.37	1.17	1.59	5.66E-05	20	1.43	1.18	1.74	3.54E-04 **	9.17E-01
CD40	toastLAA	EOS	22	0.72	0.62	0.84	1.47E-05	21	0.74	0.64	0.86	8.25E-05 **	9.72E-01
CD40	toastLAA	LOS	22	0.80	0.71	0.91	4.54E-04	21	0.79	0.69	0.92	2.16E-03 **	9.61E-01
CD40	toastOTHER	EOS	22	0.72	0.62	0.85	5.61E-05	19	0.65	0.52	0.82	1.78E-04 **	8.39E-01
CD40	toastUNDETER	LOS	22	0.84	0.75	0.93	1.42E-03	21	0.85	1.05	0.77	2.52E-03 **	7.95E-01
CRP	AllStroke	EOS	2436	1.18	1.10	1.26	1.80E-06	53	0.97	0.80	1.16	7.16E-01	4.93E-02 †
CRP	toastLAA	EOS	2447	1.43	1.21	1.69	2.30E-05	49	0.62	0.37	1.04	6.85E-02	2.50E-03 †
CRP	toastSAO	EOS	2424	1.36	1.16	1.59	2.06E-04	52	1.38	0.92	2.05	1.16E-01	9.51E-01
CRP	toastUNDETER	EOS	2438	1.24	1.10	1.40	4.05E-04	52	0.80	0.59	1.09	1.65E-01	1.00E-02 †
WWF	AllStroke	EOS	7	1.17	1.05	1.31	4.09E-03	3	1.23	1.06	1.43	6.18E-03 *	6.05E-01
WWF	toastCE	EOS	7	1.57	1.20	2.05	1.00E-03	3	1.47	1.00	2.17	5.09E-02 *	7.94E-01
WWF	toastCE	LOS	7	1.33	1.13	1.56	5.64E-04	3	1.39	1.11	1.75	4.80E-03 *	7.47E-01

3.5 Discussion

In this study, we used two-sample MR analyses to assess the causal relationship of elevated inflammatory biomarkers and ischemic stroke, stratified early and late onset as well as TOAST subtypes. Seventeen biomarkers were chosen, representing different aspects of inflammation, and were found to have a mixed role and significance in ischemic stroke risk. In particular, increased CD40 has directionally discordant effects, being significantly protective for all strokes, LAA, and UNDETER in LOS, LAA and OTHER in EOS, but being causally detrimental for EOS CE stroke.

These results are partially validated by previous MR studies, such as Chong et al. and Chen et al., that have also reported that higher genetically predicted levels of CD40 decreased the risk of large artery atherosclerosis. However, to our knowledge, our finding that increased CD40 is a risk factor of EOS cardioembolic stroke is novel.

CD40, and, by extension, the CD40-CD40 Ligand axis, is central to inflammatory responses and plays a significant role in cardiovascular pathophysiology. CD40, a receptor expressed on immune cells (e.g., B cells, macrophages, and dendritic cells), endothelial cells, and vascular smooth muscle cells, binds to CD40L, which is primarily expressed on activated T cells and platelets.⁸⁷ This binding initiates a cascade of immune and inflammatory signals that promote cell activation, cytokine release, and adhesion molecule expression, contributing to atherosclerosis and thrombosis. With respect to atherosclerosis, CD40-CD40L signaling drives plaque development and instability.⁸⁸ Within the atherosclerotic plaque, CD40L-expressing T cells and platelets interact with CD40-bearing endothelial cells, smooth muscle cells, and macrophages, leading to the activation of these cells and the promotion of an inflammatory plaque environment. Stimulated macrophages and smooth muscle cells produce pro-inflammatory mediators and promote the uptake of lipids into the arterial wall, accelerating plaque progression. The signaling also upregulates matrix metalloproteinases production, inhibiting the formation of stabilizing collagen and thinning the fibrous cap that stabilizes the plaque in the plaque. This process increases plaque vulnerability, making it more prone to rupture, which is a critical event in acute cardiovascular events, such as thromboembolism and thrombosis. Thrombotic risk is further exacerbated by CD40-CD40L role in platelet activation.⁸⁹ When activated, platelets rapidly express CD40L on their surface. Through CD40-CD40L interactions, platelets enhance the release of tissue factor from endothelial cells and monocytes, initializing the formation of cellular aggregates. Combined with ruptured plaques, blood clots are formed, leading to thrombotic events and stroke.

Our MR findings demonstrate a causal association between CD40 and cardioembolic EOS, aligning with the established understanding of CD40 signaling. Both in vivo and clinical studies indicate that cardioembolic thrombi are primarily composed of platelets.⁹⁰ CD40 signaling represents a plausible mechanism for enhanced platelet aggregation, which may elevate the risk of cardioembolic stroke.

We also demonstrated that CD40 is protective of large artery atherosclerosis stroke, contradicting CD40's critical role in atherosclerosis pathogenesis. One explanation is that the exposure used does not just reflect CD40 as just an inflammatory biomarker. The CD40 genetic instrument is based on GWAS summary statistics of plasma proteomic data, derived from immunoassays of plasma. As such, the genetic association is with respect to measured circulating CD40 concentration and without any distinction between soluble from bound protein or active from inactive. This is further complicated due to CD40 existing in multiple forms. CD40 typically functions as a trans-membrane, surface receptor protein that interacts with ligands on other cells in order to produce a signal. However, CD40 can also exist in a soluble form that still mediates signals: cleaved by ADAM17⁹¹ or as isoforms through alternative splicing.⁹² Cleavage of CD40 receptors may directly downregulate signaling. Additionally, it has been hypothesized that soluble CD40 can act as a competitive inhibitor against the membranous form.⁹³ In our analysis, we demonstrated that elevated levels of CD40 as an inflammatory biomarker has the effect on stroke. Even so, it is possible that the genetic instrument used more accurately represents weakened CD40 signaling and may explain why our findings show a protective effect for large artery atherosclerosis stroke.

The causal estimates of the other inflammatory biomarkers (Adiponectin, CCL2, CRP, Factor VIII, and vWF) and stroke were initially found to be statistically significant. However, in the cis-gene MR analysis, where non-cis variants were removed from the instrumental variables, all biomarkers were attenuated to nominal significance except for CRP, which fell to non-significance. This suggests that the excluded variants at least drove the causal effects previously detected and does not necessarily affect stroke risk through the biomarkers exclusively. For example, the CRP instrument originally contained the variants rs60191955 and rs1880241. While they are both significantly associated with elevated CRP levels, they are also significantly associated with elevated interleukin 6 levels, another inflammatory biomarker involved in signaling pathways upstream of CRP. It is possible that these variants effects in CRP-stroke relationships are in actuality the result of confounding, where the variants affect both CRP and stroke risk, or pleiotropic effects, where the variants are more so related to elevated interleukin 6 levels and signaling that inadvertently lead to elevated CRP levels.

These issues highlight one of the limitations of MR, where constructing an instrumental variable specific for the exposure is challenging. Careful consideration must be taken in choosing an appropriate exposure and representative variants in order to avoid estimation bias that may not accurately reflect the causal relationship that MR is attempting to model. There are other limitations to this study. Analyses were restricted to individuals of European ancestry, due to the relatively low contribution of non-European samples to existing genome-wide association studies of inflammation and stroke. Racial/ethnic disparities in inflammation have shown to greatly disfavor individuals of non-European ancestry.⁹⁴ To properly study the effects of inflammation on stroke in these racial groups,

future studies involving non-European samples are urgently needed.⁷⁷ Sample size and power may also be an issue; the number of stroke cases in each subtype classification was relatively small, limiting the ability to accurately measure subtype-specific effect.⁹⁵

In summary, MR analysis was used to assess the causal effects of chronic inflammation, represented as elevated levels of inflammatory biomarkers previously implicated through observational studies, on EOS and LOS. Interestingly, genetically predicted elevated levels of CD40 were protective for all stroke onset and subtypes but causal for cardioembolic EOS. These findings highlight the complex relationship between inflammation and stroke. Additional analysis is required to further validate these findings.

SUMMARY AND IMPLICATIONS OF THE RESEARCH

Ischemic stroke is one of the leading causes of disability and death in the world, but stroke incidence has fallen substantially in high-income countries.¹² However, there have been several reports indicating that incidence among younger adults appears to be increasing instead.^{76,96-98} Stroke in young adults has far-reaching impacts that extend beyond physical disability. Young survivors face a reduction in quality of life during their prime years, often grappling with significant psychological and social challenges as they adjust to new limitations. Alongside the loss of physical function and productivity, the need for long-term healthcare, support, and rehabilitation imposes a substantial financial burden on both individuals and the healthcare system. Addressing this rising public health concern of early onset stroke requires a multifaceted approach, including preventive measures, early detection, effective management, and ongoing support for individuals and their families.

The primary objective of this study was to investigate the effects of conventional risk factors and inflammation in ischemic stroke. We hypothesize some of these risk factors have a greater effect size on EOS than LOS. This chapter summarizes the key findings of this study, as documented in further details in Chapters 2 and 3, presents its strengths and limitations, and discusses the research and policy implications.

4.1 Main findings

In chapter 2, we developed age-specific genetic associations with conventional stroke risk factors and applied mendelian randomization to assess their causal relationships with two groups of stroke patients with different age of onset, EOS and LOS. Our analyses revealed that elevated BMI, DBP, SBP, T2D, and lower levels of HDL-C were found to be causally

associated with the risk of EOS while only SBP was causally associated with the risk of LOS. Direct comparison of the causal estimates showed that the causal effect size of BMI was significantly stronger for EOS than for LOS. DBP, SBP, T2D, and HDL were also stronger in EOS, although they did not achieve statistical significance.

In chapter 3, the analysis focused on the causal relationship of inflammation on risk of EOS and LOS. Once again using mendelian randomization, we tested whether genetic predisposition to elevated levels of inflammatory biomarkers were associated with the two groups of stroke patients stratified by age of onset followed by analysis stratified by TOAST subtypes. In the primary analysis, CRP and CCL2 were found to be causal for EOS while factor VIII and lower levels of CD40 were causal for LOS. When broken down by subtypes, CCL2, Factor VIII, and VWF causal effect were consistent with those found in the primary grouped analysis. Adiponectin, which previously had non-statistically significant association with EOS, was shown to have strong causal effect on SAO specifically. Lastly, elevated levels of CD40 were found to be protective for LAA EOS and LOS, OTHER and UNDETER LOS, but detrimental for EOS CE instead.

4.2 Strength and limitations

This study was the first to evaluate the risk factor profile of EOS and LOS using MR and directly compare and contrast them. Previous studies have examined the association of these risk factors in stroke but not specifically for EOS. The stratified analysis, based on age groups (EOS, LOS) as well as TOAST subtypes, allowed us to determine if genetic burden of conventional risk factors and inflammation varied among these groups. Our

analyses have extended the knowledge of the impact of conventional cardiometabolic risk and chronic inflammation on stroke by highlighting differences between young and old.

While this study offers new insights into the risk factor profile of EOS and LOS, there are several limitations that should be considered. The analysis of this study is constrained to data derived from cohorts of European ancestry. The majority of large-scaled, publicly available GWAS summary statistics used to source the exposure-variant effect sizes are based on European subjects; while there are GWAS performed using mixed ancestry, the discovery sample was still predominantly European.

As such, the outcome dataset, SiGN and EOSC, was also limited to only individuals of European ancestry. In MR studies, matching the ancestry of the exposure and outcome populations is crucial to reduce potential biases and improve the validity of causal estimates.⁹⁹ Genetic allele frequencies vary by ancestry. If the ancestry of the exposure and outcome samples differs, population stratification can lead to confounding, where allele frequency differences between populations could distort the causal estimates.¹⁰⁰ This confounding may bias the calculated estimate and affect the interpretation of the results. Moreover, matching ancestry ensures the chosen genetic instruments for the exposure are applicable to the outcome population as well. However, the lack of non-European representation compromises generalizability of our findings.¹⁰¹ While cardiometabolic dysfunction and subsequent cardiovascular diseases affect all racial and ethnic groups, there is evidence of excess burden among non-whites.¹⁰² The disparity in cardiometabolic health outcomes among racial and ethnic groups highlights a crucial gap in genomic research, particularly in studies focused on cardiovascular disease.^{103,104} Although genetic studies have made significant strides in understanding cardiometabolic risk factors, the

limited representation of diverse populations can lead to findings that are less applicable to non-European groups. As a result, the lack of diversity not only constrains our understanding of the genetic underpinnings of disease but also affects the development of personalized treatments that could address the unique risks faced by underrepresented populations. Expanding representation in genomics research is therefore essential for creating a more inclusive and equitable approach to cardiovascular health, one that acknowledges and addresses the distinct experiences and needs of all racial and ethnic groups.

Another limitation to be mindful of is the issue of power in our MR analysis. When selecting the exposure dataset, GWAS with larger discovery samples size are preferred to maximize power, ensuring results are reliable and minimizing the risk of Type II errors. In our conventional risk factor chapter, initially the largest available summary statistics, with a sample size average of 840,000, were used in the instrumental variable selection process. Once the set of variants was established, the effect sizes were recalculated based on two age groups in the UK Biobank, both groups averaging 190,000. While metric such as the F-statistics indicated that the new effect size weights were still sufficiently strong enough to proxy the effect of the risk factors, the difference in the original sample size and age-specific one may have lowered the power of the overall MR analysis. In order to improve certainty in these findings, replication analysis should be conducted using larger cohorts of age-match groups.

4.3 Research and policy implications

Overall, this study investigates the differential effects of conventional risk factors and chronic inflammation on ischemic stroke in two age groups, EOS and LOS. We found that

BMI, blood pressure, and lower HDL level have stronger causal effect in EOS compared to LOS. In addition, elevated inflammatory biomarkers have a mix of causal and protective functions in ischemic stroke depending on the age of onset and etiology. These results suggest that intervention of these targets may reduce risk of stroke and identifies new areas to consider for future research and policy making.

Cardiometabolic dysfunction and its symptoms are well-established risk factors for cardiovascular diseases, including stroke. Implementing cardiovascular health education programs has play a crucial role in promoting awareness and encouraging lifestyle changes to mitigate cardiovascular disease risks. A major goal of cardiovascular health education is to raise awareness about the risk factors and symptoms of heart disease, empowering individuals to recognize early warning signs and take proactive steps to protect their heart health. Highlighting the results of this thesis in public health efforts may help modifying outlook and behavior towards risk factor control, especially with younger adults. Notably, elevated BMI, blood pressure, and lower levels of HDL were identified as causal risk factors for EOS. Through regular screenings and the early management of high blood pressure and cholesterol as well as health-conscious diet and exercise, individuals can significantly lower their risk of heart-related complications and improve their overall quality of life.

Despite the great strides in disseminating cardiovascular-conscious education and curbing the prevalence of stroke worldwide, stroke incidence has risen among young adults. Multiple factors may account for this, including changes in vascular risk factors as well as lower awareness of the disease and treatment options.⁹⁶ Rising obesity, driven by poor dietary habits and sedentary lifestyles, has been reported among adults 20 to 44 years old.¹⁴

Similarly, uncontrolled hypertension, often exacerbated by stress, poor medication adherence, and lack of regular monitoring, contributes significantly to the risk of stroke.¹⁰⁵

It is possible that the combined effects of rising risk factor prevalence and the heightened impact of these factors in younger adults, as demonstrated in the second chapter, explain the increasing incidence of EOS seen globally. If left unaddressed, young adults will continue to be vulnerable to stroke risk. Health education efforts must increasingly focus on the importance of managing risk factors and promoting early, preventive interventions to combat EOS.

The inflammatory process is understood to have a fundamental role in both the etiology and the pathophysiology consequence of cerebral ischemia. The third chapter provides evidence further supporting this and highlights potential targets for intervention. One such target is CD40, which had directionally discordant effects, being protective for LAA but causal for CE in EOS. CD40 and CD40L are intimately involved in leukocyte recruitment, leading to atherosclerotic pathogenesis that causes LAA, as well as platelet aggregation, leading to the thrombosis mechanism of CE.^{106,107} CD40-CD40L has been previously proposed as an immunotherapeutic target in cardiovascular disease. Preclinical studies showed promising results in slowing atherosclerotic progression, but also showed risk for immunosuppression and thromboembolic events. Continued effort in MR analysis of inflammation and stroke may identify other desirable targets equally as effective as the CD40-CD40L signaling system but with less burdensome side-effects.

4.4 Future work and Considerations

The MR framework is a powerful and elegant scientific method for assessing causality in epidemiology. The systematic process of MR has identified hundreds of strongly causal

variants and promising targets for intervention. However, the MR is limited in that it can only validate causal relationships, rooted in biological plausibility, and lacks the ability to explain mechanistically how exposures relate to outcomes.¹⁰⁸ Improving the translatability of the MR results will require alternative methods of research better suited for addressing these questions. One example is Phenome-wide association study that can identify genes and phenotypes also associated with the instrumental variable. This may add additional context to what biological processes are involved in the overall exposure trait. Another example would be traditional laboratory-based experiments, which would help examine how variants, genes, and protein interact with one another to form the phenotype of the exposure.

The MR analysis from this study may be improved by adjusting the definition of certain exposure/outcomes, leading to novel, alternative results. For example, in chapter 2 we used genetic predisposition to higher BMI and assessed its causal effects on stroke. BMI was surprisingly found to be not causal for LOS, in spite of decades of research stating otherwise. Instead of using BMI, it may be prudent to use other metrics of obesity and repeat the MR analysis. In particular, waist hip ratio has been shown to more accurately reflect body fat distribution as opposed to BMI and may serve as a better predictor for adverse outcome.^{109,110}

Similarly, the MR analysis in chapter 3 could be further applied to identify causal inflammation biomarkers for the outcome of stroke. Several of the inflammatory biomarkers were chosen based on prevalence/case-control studies of stroke. That is, the biomarkers' concentration was measured post-ischemic event, roughly 18-72 hours afterwards. As such, these studies did not necessarily indicate the cause-effect relationship

between chronic inflammation and stroke incidence, which was seen in some of the MR results. Interestingly, these same biomarkers may instead serve as prognostic factors for stroke outcome severity. Fibrinogen, TNF, and IL-6, biomarkers that were non-causal for stroke incidence in the MR analysis, have been observed in patients with severe ischemia and poststroke infections.¹¹¹ It may be interesting to shift the inflammation/stroke analysis and see if the biomarkers identified by the prevalence study are causal for stroke outcome instead.

In conclusion, this study contributes important scientific insights to the understanding of the effects of stroke risk factors. To our knowledge, it is the first to assess the differential effect within EOS and LOS. Using the two-sample MR framework, we found that genetic predisposition to higher BMI, blood pressure, type 2 diabetes, and lower HDL-cholesterol level had stronger effect size over the risk of EOS compared to that of LOS. Chronic inflammation, represented by certain inflammatory biomarkers, exhibited differential causal effects for the risk of stroke depending on the age of onset. Namely, elevated levels of CCL2 were causal for all EOS while elevated levels of factor VIII and lower levels of CD40 were causal for all LOS. Furthermore, adiponectin and CD40 were causal specifically for cardioembolic EOS. These results highlight the importance of early intervention of specific modifiable, cardiometabolic risk factors as well as provides potential drug targets that may reduce the risk of EOS specifically. Dissemination of genetic and stroke research is essential for raising awareness and guiding preventative strategies, and future research is needed to translate results into applicable intervention.

REFERENCES

1. George MG, Fischer L, Koroshetz W, Bushnell C, Frankel M, Foltz J, Thorpe PG. CDC Grand Rounds: Public Health Strategies to Prevent and Treat Strokes. *Morbidity and Mortality Weekly Report*. 2017 May 12;66. doi: 10.15585/mmwr.mm6618a5
2. SC C. Stroke Death Rates Among Adults Ages 45-64 by Region and Race and Hispanic Origin: United States, 2002-2022 - PubMed. *NCHS data brief*. 2024 Aug. doi: 10.15620/cdc/158326
3. Zhou T, Havenon Ad, Sheth KN, Ross JS. Disability Status and Secondary Prevention Among Survivors of Stroke: A Cross-Sectional Analysis of the 2011 to 2018 National Health and Nutrition Examination Survey. *Journal of the American Heart Association: Cardiovascular and Cerebrovascular Disease*. 2023 Nov 28;12. doi: 10.1161/JAHA.123.030869
4. Koton S, Schneider ALC, Rosamond WD, Shahar E, Sang Y, Gottesman RF, Coresh J. Trends in Stroke Incidence and Mortality. *JAMA*. 2014/07/16;312. doi: 10.1001/jama.2014.7692
5. Ananth CV, Brandt JS, Keyes KM, Graham HL, Kostis JB, Kostis WJ. Epidemiology and trends in stroke mortality in the USA, 1975–2019. *International Journal of Epidemiology*. 2022 Nov 7;52. doi: 10.1093/ije/dyac210
6. Ramirez L, Kim-Tenser MA, Sanossian N, Cen S, Wen G, He S, Mack WJ, Towfighi A. Trends in Acute Ischemic Stroke Hospitalizations in the United States. *Journal of the American Heart Association: Cardiovascular and Cerebrovascular Disease*. 2016 May 11;5. doi: 10.1161/JAHA.116.003233
7. Towfighi A, Saver JL. Stroke Declines From Third to Fourth Leading Cause of Death in the United States. *Stroke*. 2011-8;42. doi: 10.1161/STROKEAHA.111.621904
8. Luepker RV, Arnett DK, Jacobs DR, Duval SJ, Folsom AR, Armstrong C, Blackburn H. Trends in Blood Pressure, Hypertension Control, and Stroke Mortality: The Minnesota Heart Survey. *The American Journal of Medicine*. 2006/01/01;119. doi: 10.1016/j.amjmed.2005.08.051
9. Mann D, Reynolds K, Smith D, Muntner P, Devin Mann KR, Donald Smith, Paul Muntner. Trends in Statin Use and Low-Density Lipoprotein Cholesterol Levels Among US Adults: Impact of the 2001 National Cholesterol Education Program Guidelines. *Annals of Pharmacotherapy*. 2008-07-29;42. doi: 10.1345/aph.1L181
10. Cutler JA, Sorlie PD, Wolz M, Thom T, Fields LE, Roccella EJ. Trends in Hypertension Prevalence, Awareness, Treatment, and Control Rates in United States Adults Between 1988–1994 and 1999–2004. *Hypertension*. 2008-11-01;52. doi: 10.1161/HYPERTENSIONAHA.108.113357
11. Mercy UC, Farhadi K, Oguniola AS, Karaye RM, Baguda US, Eniola OA, Yunusa I, Karaye IM. Revisiting recent trends in stroke death rates, United States, 1999–2020. *Journal of the Neurological Sciences*. 2023/08/15;451. doi: 10.1016/j.jns.2023.120724
12. Scott CA, Li L, Rothwell PM. Diverging Temporal Trends in Stroke Incidence in Younger vs Older People. *JAMA Neurology*. 2022/10/01;79. doi: 10.1001/jamaneurol.2022.1520

13. MS E, JI V, I V, KM vN, CJM K, FE dL. Stroke incidence in young adults according to age, subtype, sex, and time trends - PubMed. *Neurology*. 05/21/2019;92. doi: 10.1212/WNL.00000000000007533
14. Aggarwal R, Yeh RW, Maddox KEJ, Wadhwa RK. Cardiovascular Risk Factor Prevalence, Treatment, and Control in US Adults Aged 20 to 44 Years, 2009 to March 2020. *JAMA*. 2023 Mar 5;329. doi: 10.1001/jama.2023.2307
15. Wang L, Li X, Wang Z, Bancks MP, Carnethon MR, Greenland P, Feng Y-Q, Wang H, Zhong VW. Trends in Prevalence of Diabetes and Control of Risk Factors in Diabetes Among US Adults, 1999-2018. *JAMA*. 2021 Jun 25;326. doi: 10.1001/jama.2021.9883
16. Kalra R, Parcha V, Patel N, Bhargava A, Booker KS, Arora G, Arora P. Increased awareness, inadequate treatment, and poor control of cardiovascular risk factors in American young adults: 2005–2016. *European Journal of Preventive Cardiology*. 2021/04/23;28. doi: 10.1177/2047487320905190
17. Joundi RA, Patten SB, Williams JVA, Smith EE. Association Between Excess Leisure Sedentary Time and Risk of Stroke in Young Individuals. *Stroke*. 2021-11;52. doi: 10.1161/STROKEAHA.121.034985
18. H P Adams J, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, E E Marsh r. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke*. 1993-Jan;24. doi: 10.1161/01.STR.24.1.35
19. Ornello R, Degan D, Tiseo C, Di Carmine C, Perciballi L, Pistoia F, Carolei A, Sacco S. Distribution and Temporal Trends From 1993 to 2015 of Ischemic Stroke Subtypes: A Systematic Review and Meta-Analysis. *Stroke*. 2018;49:814-819. doi: 10.1161/STROKEAHA.117.020031
20. George MG. Risk Factors for Ischemic Stroke in Younger Adults: A Focused Update. *Stroke*. 2020;51:729-735. doi: 10.1161/STROKEAHA.119.024156
21. Falk E, Shah PK, Fuster V. Coronary Plaque Disruption. *Circulation*. 1995-08-01;92. doi: 10.1161/01.CIR.92.3.657
22. Spagnoli LG, Mauriello A, Sangiorgi G, Fratoni S, Bonanno E, Schwartz RS, Piepgras DG, Pistolesse R, Ippoliti A, Holmes DR. Extracranial Thrombotically Active Carotid Plaque as a Risk Factor for Ischemic Stroke. *JAMA*. 2004/10/20;292. doi: 10.1001/jama.292.15.1845
23. Howard DPJ, Lammeren GWv, Redgrave JN, Moll FL, Vries J-PPMd, Kleijn DPVd, Borst GJd, Pasterkamp G, Rothwell PM. Histological Features of Carotid Plaque in Patients With Ocular Ischemia Versus Cerebral Events. *Stroke*. 2013-March;44. doi: 10.1161/STROKEAHA.112.678672
24. Redgrave JN, Gallagher P, Lovett JK, Rothwell PM. Critical Cap Thickness and Rupture in Symptomatic Carotid Plaques. *Stroke*. 2008-06-01;39. doi: 10.1161/STROKEAHA.107.507988
25. Marnane M, Duggan CA, Sheehan OC, Merwick A, Hannon N, Curtin D, Harris D, Williams EB, Horgan G, Kyne L, et al. Stroke Subtype Classification to Mechanism-Specific and Undetermined Categories by TOAST, A-S-C-O, and Causative Classification System. *Stroke*. 2010-08-01. doi: 10.1161/STROKEAHA.109.575373

26. Aviles RJ, Martin DO, Apperson-Hansen C, Houghtaling PL, Rautaharju P, Kronmal RA, Tracy RP, Wagoner DRV, Psaty BM, Lauer MS, et al. Inflammation as a Risk Factor for Atrial Fibrillation. *Circulation*. 2003-12-16;108. doi: 10.1161/01.CIR.0000103131.70301.4F
27. Conen D, Ridker PM, Everett BM, Tedrow UB, Rose L, Cook NR, Buring JE, Albert CM. A multimarker approach to assess the influence of inflammation on the incidence of atrial fibrillation in women. *European Heart Journal*. 2010 May 25;31. doi: 10.1093/eurheartj/ehq146
28. Marott SCW, Nordestgaard BG, Zacho J, Friberg J, Jensen GB, Tybjaerg-Hansen A, Benn M. Does Elevated C-Reactive Protein Increase Atrial Fibrillation Risk?: A Mendelian Randomization of 47,000 Individuals From the General Population. *Journal of the American College of Cardiology*. 2010/08/31;56. doi: 10.1016/j.jacc.2010.02.066
29. Marcus GM, Whooley MA, Glidden DV, Pawlikowska L, Zaroff JG, Olgin JE. Interleukin 6 and Atrial Fibrillation in Patients with Coronary Artery Disease: Data from the Heart and Soul Study. *American heart journal*. 2007 Oct 25;155. doi: 10.1016/j.ahj.2007.09.006
30. Esmon CT. Inflammation and thrombosis. *J Thromb Haemost*. 2003;1:1343-1348. doi: 10.1046/j.1538-7836.2003.00261.x
31. CT E. The impact of the inflammatory response on coagulation - PubMed. *Thrombosis research*. 2004;114. doi: 10.1016/j.thromres.2004.06.028
32. Tuttolomondo A, Sciacca RD, Raimondo DD, Serio A, D'Aguanno G, Placa SL, Pecoraro R, Arnao V, Marino L, Monaco S, et al. Plasma levels of inflammatory and thrombotic/fibrinolytic markers in acute ischemic strokes: Relationship with TOAST subtype, outcome and infarct site. *Journal of Neuroimmunology*. 2009/10/30;215. doi: 10.1016/j.jneuroim.2009.06.019
33. Fumagalli S, Perego C, Pischiutta F, Zanier ER, De Simoni MG. The ischemic environment drives microglia and macrophage function. *Front Neurol*. 2015;6:81. doi: 10.3389/fneur.2015.00081
34. Rohde LEP, Hennekens CH, Ridker PM. Cross-Sectional Study of Soluble Intercellular Adhesion Molecule-1 and Cardiovascular Risk Factors in Apparently Healthy Men. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 1999-07;19. doi: 10.1161/01.ATV.19.7.1595
35. Hackman A, Abe Y, William Insull J, Pownall H, Smith L, Dunn K, Antonio M, Gotto J, Ballantyne CM. Levels of Soluble Cell Adhesion Molecules in Patients With Dyslipidemia. *Circulation*. 1996-04-01;93. doi: 10.1161/01.CIR.93.7.1334
36. Ferri C, Desideri G, Valenti M, Bellini C, Pasin M, Santucci A, Mattia GD. Early Upregulation of Endothelial Adhesion Molecules in Obese Hypertensive Men. *Hypertension*. 1999-10;34. doi: 10.1161/01.HYP.34.4.568
37. Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafò MR, Palmer T, Schooling CM, Wallace C, Zhao Q, et al. Mendelian randomization. *Nature reviews Methods primers*. 2022 Feb 10;2. doi: 10.1038/s43586-021-00092-5
38. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;37:658-665. doi: 10.1002/gepi.21758

39. Network NSG, International Stroke Genetics C. Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurol.* 2016;15:174-184. doi: 10.1016/S1474-4422(15)00338-5
40. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, Rutten-Jacobs L, Giese A-K, van der Laan SW, Gretarsdottir S, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nature Genetics* 2018 50:4. 2018-03-12;50. doi: 10.1038/s41588-018-0058-3
41. Larsson SC, Scott RA, Traylor M, Langenberg CC, Hindy G, Melander O, Orho-Melander M, Seshadri S, Wareham NJ, Markus HS, et al. Type 2 diabetes, glucose, insulin, BMI, and ischemic stroke subtypes: Mendelian randomization study. *Neurology.* 2017 Aug 1;89. doi: 10.1212/WNL.0000000000004173
42. Georgakis MK, Gill D, Webb AJS, Evangelou E, Elliott P, Sudlow CLM, Dehghan A, Malik R, Tzoulaki I, Dichgans M. Genetically determined blood pressure, antihypertensive drug classes, and risk of stroke subtypes. *Neurology.* 2020;95:e353-e361. doi: 10.1212/WNL.0000000000009814
43. Georgakis MK, Harshfield EL, Malik R, Franceschini N, Langenberg C, Wareham NJ, Markus HS, Dichgans M. Diabetes Mellitus, Glycemic Traits, and Cerebrovascular Disease: A Mendelian Randomization Study. *Neurology.* 2021 Mar 30;96. doi: 10.1212/WNL.0000000000011555
44. Larsson SC, Burgess S, Michaëlsson K. Smoking and stroke: A mendelian randomization study. *Annals of Neurology.* 2019/09/01;86. doi: 10.1002/ana.25534
45. Yuan S, Tang B, Zheng J, Larsson SC. Circulating Lipoprotein Lipids, Apolipoproteins and Ischemic Stroke. *Annals of Neurology.* 2020/12/01;88. doi: 10.1002/ana.25916
46. Hindy G, Engström G, Larsson SC, Traylor M, Markus HS, Melander O, Orho-Melander M. Role of Blood Lipids in the Development of Ischemic Stroke and its Subtypes: A Mendelian Randomization Study. *Stroke.* 2018 Mar 26;49. doi: 10.1161/STROKEAHA.117.019653
47. Lin J, Wang Y, Wang Y, Pan Y. Inflammatory biomarkers and risk of ischemic stroke and subtypes: A 2-sample Mendelian randomization study. *Neurological Research.* 2020-2-1;42. doi: 10.1080/01616412.2019.1710404
48. Georgakis MK, Malik R, Gill D, Franceschini N, Sudlow CLM, Consortium I, Group CIW, Dichgans M. Interleukin-6 Signaling Effects on Ischemic Stroke and Other Cardiovascular Outcomes: A Mendelian Randomization Study. *Circulation Genomic and Precision Medicine.* 2020 May 12;13. doi: 10.1161/CIRCGEN.119.002872
49. Georgakis MK, Gill D, Rannikmäe K, Traylor M, Anderson CD, (ISGC) McoTISGC, Lee J-M, Kamatani Y, Hopewell JC, Worrall BB, et al. Genetically Determined Levels of Circulating Cytokines and Risk of Stroke: Role of Monocyte Chemoattractant Protein-1. *Circulation.* 2019 Jan 8;139. doi: 10.1161/CIRCULATIONAHA.118.035905
50. Yuan S, Carter P, Bruzelius M, Vithayathil M, Kar S, Mason AM, Lin A, Burgess S, Larsson SC. Effects of tumour necrosis factor on cardiovascular disease and

- cancer: A two-sample Mendelian randomization study. *EBioMedicine*. 2020 Aug 14;59. doi: 10.1016/j.ebiom.2020.102956
51. Chong M, Sjaarda J, Pigeyre M, Mohammadi-Shemirani P, Lali R, Shoamanesh A, Gerstein HC, Paré G. Novel Drug Targets for Ischemic Stroke Identified Through Mendelian Randomization Analysis of the Blood Proteome. *Circulation*. 2019-09-03;140. doi: 10.1161/CIRCULATIONAHA.119.040180
 52. Tsao CW, Aday AW, Almarzooq ZI, Anderson CAM, Arora P, Avery CL, Baker-Smith CM, Beaton AZ, Boehme AK, Buxton AE, et al. Heart Disease and Stroke Statistics-2023 Update: A Report From the American Heart Association. *Circulation*. 2023;147:e93-e621. doi: 10.1161/CIR.0000000000001123
 53. Rohr J, Kittner S, Feeser B, Hebel JR, Whyte M-G, Weinstein A, Kanarak N, Buchholz D, Earley C, Johnson C, et al. Traditional Risk Factors and Ischemic Stroke in Young Adults: The Baltimore-Washington Cooperative Young Stroke Study. *Archives of Neurology*. 1996/07/01;53. doi: 10.1001/archneur.1996.00550070041010
 54. Aigner A, Grittner U, Rolfs A, Norrving B, Siegerink B, Busch MA. Contribution of Established Stroke Risk Factors to the Burden of Stroke in Young Adults. *Stroke*. 2017;48:1744-1751. doi: 10.1161/STROKEAHA.117.016599
 55. Putaala J, Metso AJ, Metso TM, Konkola N, Kraemer Y, Haapaniemi E, Kaste M, Tatlisumak T. Analysis of 1008 consecutive patients aged 15 to 49 with first-ever ischemic stroke: the Helsinki young stroke registry. *Stroke*. 2009;40:1195-1203. doi: 10.1161/STROKEAHA.108.529883
 56. Ebrahim S, Davey Smith G, Ebrahim S, Davey Smith G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Human Genetics* 2007 123:1. 2007-11-23;123. doi: 10.1007/s00439-007-0448-6
 57. Skrivankova VW, Richmond RC, Woolf BAR, Davies NM, Swanson SA, VanderWeele TJ, Timpson NJ, Higgins JPT, Dimou N, Langenberg C, et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ*. 2021;375:n2233. doi: 10.1136/bmj.n2233
 58. Jaworek T, Xu H, Gaynor BJ, Cole JW, Rannikmae K, Stanne TM, Tomppo L, Abedi V, Amouyel P, Armstrong ND, et al. Contribution of Common Genetic Variants to Risk of Early-Onset Ischemic Stroke. *Neurology*. 2022;99:e1738-e1754. doi: 10.1212/WNL.0000000000201006
 59. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48:1284-1287. doi: 10.1038/ng.3656
 60. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, Groza T, Gunes O, Hall P, Hayhurst J, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res*. 2023;51:D977-D985. doi: 10.1093/nar/gkac1010
 61. Burgess S, Thompson SG, Collaboration CCG. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol*. 2011;40:755-764. doi: 10.1093/ije/dyr036

62. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7. doi: 10.1186/s13742-015-0047-8
63. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol*. 2016;40:304-314. doi: 10.1002/gepi.21965
64. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol*. 2017;32:377-389. doi: 10.1007/s10654-017-0255-x
65. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet*. 2018;50:693-698. doi: 10.1038/s41588-018-0099-7
66. <Simons et al. - 1998 - Risk Factors for Ischemic Stroke.pdf>.
67. <Singer et al. - 2019 - Independent ischemic stroke risk factors in older .pdf>.
68. Georgakis MK, Gill D. Mendelian Randomization Studies in Stroke: Exploration of Risk Factors and Drug Targets With Human Genetic Data. *Stroke*. 2021;52:2992-3003. doi: 10.1161/STROKEAHA.120.032617
69. Mitchell AB, Cole JW, McArdle PF, Cheng YC, Ryan KA, Sparks MJ, Mitchell BD, Kittner SJ. Obesity increases risk of ischemic stroke in young adults. *Stroke*. 2015;46:1690-1692. doi: 10.1161/STROKEAHA.115.008940
70. Horn JW, Feng T, Morkedal B, Strand LB, Horn J, Mukamal K, Janszky I. Obesity and Risk for First Ischemic Stroke Depends on Metabolic Syndrome: The HUNT Study. *Stroke*. 2021;52:3555-3561. doi: 10.1161/STROKEAHA.120.033016
71. Markidan J, Cole JW, Cronin CA, Merino JG, Phipps MS, Wozniak MA, Kittner SJ. Smoking and Risk of Ischemic Stroke in Young Men. *Stroke*. 2018;49:1276-1278. doi: 10.1161/STROKEAHA.117.018859
72. AS R, JE M, IM L, S S, CH H. Cigarette smoking and stroke in a cohort of U.S. male physicians - PubMed. *Annals of internal medicine*. 03/15/1994;120. doi: 10.7326/0003-4819-120-6-199403150-00002
73. Asplund K, Karvanen J, Giampaoli S, Jousilahti P, Niemela M, Broda G, Cesana G, Dallongeville J, Ducimetriere P, Evans A, et al. Relative risks for stroke by age, sex, and population based on follow-up of 18 European populations in the MORGAM Project. *Stroke*. 2009;40:2319-2326. doi: 10.1161/STROKEAHA.109.547869
74. RL S, RT B, DE K, B B-A, C T, IF L, JF C, MC P, S S, L B. High-density lipoprotein cholesterol and ischemic stroke in the elderly: the Northern Manhattan Stroke Study - PubMed. *JAMA*. 06/06/2001;285. doi: 10.1001/jama.285.21.2729
75. Marini S, Merino J, Montgomery BE, Malik R, Sudlow CL, Dichgans M, Florez JC, Rosand J, Gill D, Anderson CD. Mendelian randomization study of obesity and cerebrovascular disease. *Annals of neurology*. 2020 Feb 19;87. doi: 10.1002/ana.25686
76. George MG, Tong X, Bowman BA. Prevalence of Cardiovascular Risk Factors and Strokes in Younger Adults. *JAMA Neurol*. 2017;74:695-703. doi: 10.1001/jamaneurol.2017.0020

77. Adebamowo CA, Adeyemo A, Ashaye A, Akpa OM, Chikowore T, Choudhury A, Fakim YJ, Fatumo S, Hanchard N, Hauser M, et al. Polygenic risk scores for CARDINAL study. *Nat Genet.* 2022;54:527-530. doi: 10.1038/s41588-022-01074-3
78. Smit RAJ, Trompet S, Dekkers OM, Jukema JW, Cessie SI. Survival Bias in Mendelian Randomization Studies: A Threat to Causal Inference. *Epidemiology (Cambridge, Mass).* 2019 Sep 30;30. doi: 10.1097/EDE.0000000000001072
79. Kelly PJ, Lemmens R, Tsvigoulis G. Inflammation and Stroke Risk: A New Target for Prevention. *Stroke.* 2021-08. doi: 10.1161/STROKEAHA.121.034388
80. Sang Y, Roest M, Laat Bd, Groot PGd, Huskens D. Interplay between platelets and coagulation. *Blood Reviews.* 2020 Jul 12;46. doi: 10.1016/j.blre.2020.100733
81. Rost NS, Wolf PA, Kase CS, Kelly-Hayes M, Silbershatz H, Massaro JM, D'Agostino RB, Franzblau C, Wilson PWF. Plasma Concentration of C-Reactive Protein and Risk of Ischemic Stroke and Transient Ischemic Attack. *Stroke.* 2001-11-01;32. doi: 10.1161/hs1101.098151
82. Zhu H, Hu S, Li Y, Sun Y, Xiong X, Hu X, Chen J, Qiu S. Frontiers | Interleukins and Ischemic Stroke. *Frontiers in Immunology.* 2022/01/31;13. doi: 10.3389/fimmu.2022.828447
83. Xue Y, Zeng X, Tu W-J, Zhao J. Tumor Necrosis Factor- α : The Next Marker of Stroke. *Disease Markers.* 2022 Feb 27;2022. doi: 10.1155/2022/2395269
84. Wenzl FA, Ambrosini S, Mohammed SA, Kraler S, Lüscher TF, Costantino S, Paneni F. Frontiers | Inflammation in Metabolic Cardiomyopathy. *Frontiers in Cardiovascular Medicine.* 2021/10/04;8. doi: 10.3389/fcvm.2021.742178
85. Simats A, Liesz A. Systemic inflammation after stroke: implications for post-stroke comorbidities. *EMBO Molecular Medicine.* 2022 Aug 15;14. doi: 10.15252/emmm.202216269
86. Pierce BL, Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am J Epidemiol.* 2013;178:1177-1184. doi: 10.1093/aje/kwt084
87. Ots HD, Tracz JA, Vinokuroff KE, Musto AE. CD40–CD40L in Neurological Disease. *International Journal of Molecular Sciences.* 2022 Apr 8;23. doi: 10.3390/ijms23084115
88. Lacy M, Bürger C, Shami A, Ahmadsei M, Winkels H, Nitz K, van Tiel CM, Seijkens TTP, Kusters PJH, Karshovka E, et al. Cell-specific and divergent roles of the CD40L-CD40 axis in atherosclerotic vascular disease. *Nature Communications* 2021 12:1. 2021-06-18;12. doi: 10.1038/s41467-021-23909-z
89. Cognasse F, Duchez AC, Audoux E, Ebermeyer T, Arthaud CA, Prier A, Eyraud MA, Mismetti P, Garraud O, Bertoletti L, et al. Frontiers | Platelets as Key Factors in Inflammation: Focus on CD40L/CD40. *Frontiers in Immunology.* 2022/02/03;13. doi: 10.3389/fimmu.2022.825892
90. Jolugbo P, Ariëns R. Thrombus composition and efficacy of thrombolysis and thrombectomy in acute ischaemic stroke. *Stroke.* 2021 Feb 10;52. doi: 10.1161/STROKEAHA.120.032810
91. A K, S M, F L, T K, M H, AH W. Ectodomain Shedding by ADAM17 Increases the Release of Soluble CD40 from Human Endothelial Cells under Pro-

- Inflammatory Conditions - PubMed. *Cells*. 07/25/2023;12. doi: 10.3390/cells12151926
92. Tone M, Tone Y, Fairchild PJ, Wykes M, Waldmann H. Regulation of CD40 function by its isoforms generated through alternative splicing. *Proceedings of the National Academy of Sciences of the United States of America*. 2001 Feb 13;98. doi: 10.1073/pnas.98.4.1751
 93. Esposito P, Rampino T, Canton AD. Soluble CD40 as a modulator of CD40 pathway. *Immunology Letters*. 2012/09/01;147. doi: 10.1016/j.imlet.2012.06.003
 94. Schmeer KK, Tarrence J. Racial/Ethnic Disparities in Inflammation: Evidence of Weathering in Childhood? *Journal of health and social behavior*. 2018 Jun 27;59. doi: 10.1177/0022146518784592
 95. Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *International Journal of Epidemiology*. 2014 Mar 6;43. doi: 10.1093/ije/dyu005
 96. Béjot Y, Daubail B, Jacquin A, Durier J, Osseby G-V, Rouaud O, Giroud M. Trends in the incidence of ischaemic stroke in young adults between 1985 and 2011: the Dijon Stroke Registry. *Journal of Neurology, Neurosurgery & Psychiatry*. 2014-05-01;85. doi: 10.1136/jnnp-2013-306203
 97. Kissela BM, Khoury JC, Alwell K, Moomaw CJ, Woo D, Adeoye O, Flaherty ML, Khatri P, Ferioli S, Rosa FDLRL, et al. Age at stroke: Temporal trends in stroke incidence in a large, biracial population. *Neurology*. 2012 Oct 23;79. doi: 10.1212/WNL.0b013e318270401d
 98. Tibæk M, Dehlendorff C, Jørgensen HS, Forchhammer HB, Johnsen SP, Kammersgaard LP. Increasing Incidence of Hospitalization for Stroke and Transient Ischemic Attack in Young Adults: A Registry-Based Study. *Journal of the American Heart Association*. 2016-05-11;5. doi: 10.1161/JAHA.115.003158
 99. Burgess S, Smith GD, Davies NM, Dudbridge F, Gill D, Glymour MM, Hartwig FP, Kutalik Z, Holmes MV, Minelli C, et al. Guidelines for performing Mendelian randomization investigations: update for summer 2023. *Wellcome Open Research*. 2023 Aug 4;4. doi: 10.12688/wellcomeopenres.15555.3
 100. Hellwege J, Keaton J, Giri A, Gao X, Edwards DRV, Edwards TL. Population Stratification in Genetic Association Studies. *Current protocols in human genetics*. 2017 Oct 18;95. doi: 10.1002/cphg.48
 101. K B-D, A H. Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups - PubMed. 05/17/2022. doi: 10.17226/26479
 102. Javed Z, Maqsood MH, Yahya T, Amin Z, Acquah I, Valero-Elizondo J, Andrieni J, Dubey P, Jackson RK, Daffin MA, et al. Race, Racism, and Cardiovascular Health: Applying a Social Determinants of Health Framework to Racial/Ethnic Disparities in Cardiovascular Disease. *Circulation: Cardiovascular Quality and Outcomes*. 2022-01;15. doi: 10.1161/CIRCOUTCOMES.121.007917
 103. Clarke SL, Assimes TL, Tcheandjieu C. The Propagation of Racial Disparities in Cardiovascular Genomics Research. *Circulation Genomic and precision medicine*. 2021 Aug 31;14. doi: 10.1161/CIRCGEN.121.003178
 104. Fernandez-Rhodes L, Young KL, Lilly AG, Raffield LM, Highland HM, Wojcik GL, Agler C, Love S-AM, Okello S, Petty LE, et al. Importance of Genetic

- Studies of Cardiometabolic Disease in Diverse Populations. *Circulation research*. 2020 Jun 4;126. doi: 10.1161/CIRCRESAHA.120.315893
105. Zhang Y, Moran AE. Trends in the Prevalence, Awareness, Treatment, and Control of Hypertension among Young Adults in the United States, 1999–2014. *Hypertension (Dallas, Tex : 1979)*. 2017 Aug 28;70. doi: 10.1161/HYPERTENSIONAHA.117.09801
 106. Bosmans LA, Bosch L, Kusters PJ, Lutgens E, Seijkens TT. The CD40-CD40L Dyad as Immunotherapeutic Target in Cardiovascular Disease. *Journal of Cardiovascular Translational Research*. 2020 Mar 28;14. doi: 10.1007/s12265-020-09994-3
 107. Lutgens E, Lievens D, Beckers L, Donners M, Daemen M. CD40 and Its Ligand in Atherosclerosis. *Trends in Cardiovascular Medicine*. 2007/05/01;17. doi: 10.1016/j.tcm.2007.02.004
 108. Nguyen K, Mitchell BD. A Guide to Understanding Mendelian Randomization Studies. *Arthritis Care & Research*. 2024/11/01;76. doi: 10.1002/acr.25400
 109. Elsayed EF, Tighiouart H, Weiner DE, Griffith J, Salem D, Levey AS, Sarnak MJ. Waist Hip Ratio and Body Mass Index as Risk Factors for Cardiovascular Events in Chronic Kidney Disease. *American journal of kidney diseases : the official journal of the National Kidney Foundation*. 2008 Jun 2;52. doi: 10.1053/j.ajkd.2008.04.002
 110. Harris E. Study: Waist-to-Hip Ratio Might Predict Mortality Better Than BMI. *JAMA*. 2023/10/24;330. doi: 10.1001/jama.2023.19205
 111. Tuttolomondo A, Raimondo DD, Sciacca Rd, Pinto A, Licata G. Inflammatory Cytokines in Acute Ischemic Stroke. *Current Pharmaceutical Design*. 2008;14. doi: 10.2174/138161208786848739