Emerging Risk Factors for Stroke
What Have We Learned From Mendelian Randomization Studies?

Jemma C. Hopewell, PhD; Robert Clarke, FRCP

Establishing new approaches for the prevention and treatment of stroke relies on identifying modifiable risk factors that contribute to the development of this complex disease. Mendelian randomization (MR) studies, analogous to naturally occurring randomized trials, can assess causality of potentially modifiable biomarkers and offer new insights into biological pathways.

Stroke is the second leading cause of death worldwide and the chief determinant of long-term disability. Stroke is a heterogeneous disease arising from several distinct underlying pathologies and is typically classified as ischemic or hemorrhagic, and further subclassified using imaging data. Ischemic stroke (IS), including its 3 main subtypes: small vessel disease, large vessel disease, and cardioembolic stroke, accounts for ≈80% of stroke and is the result of an interrupted blood supply, leading to localized areas of ischemia in the brain. Small vessel disease may be a consequence of nonatherosclerotic, as well as atherosclerotic, mechanisms that result in an occlusion of the small perforating arteries, whereas large vessel disease results from occlusions or emboli from plaque rupture in larger vessels, such as a carotid artery. Cardioembolic stroke arises typically from emboli from the heart. By contrast, hemorrhagic stroke is a consequence of intracerebral hemorrhage (bleeding into the brain) or subarachnoid hemorrhage (bleeding into the subarachnoid space). These diverse stroke subtypes have distinct underlying pathologies reflecting different risk factor distributions. MR studies, using genetic variants as instrumental variables, afford a powerful approach to assessing causality of risk factors and avoid biases inherent in observational studies, including confounding and reverse causation. This review considers the contribution of MR studies to stroke epidemiology and their relevance to understanding risk factors and new therapeutic targets for stroke.

Classical Risk Factors for Stroke

Meta-analyses of large prospective studies have enhanced our knowledge of classical and emerging risk factors for stroke.\(^1\)\(^-\)\(^4\) Classical risk factors for stroke include nonmodifiable characteristics, such as age and ethnicity, and modifiable risk factors, such as coronary heart disease (CHD), atrial fibrillation, diabetes mellitus, hypertension, smoking, dyslipidemia, and adiposity. Risk factor associations may be mediated solely or partially through other factors. For example, the effect of body mass index (BMI) on stroke may be mediated through blood pressure, lipids, and diabetes mellitus. Observational studies are constrained by the effects of confounding and reverse causality and are unable to assess causality. In contrast, genetic studies offer the opportunity to assess the causal relevance of risk factors for stroke. Moreover, given that different risk factors are likely to contribute to distinct mechanisms and individual stroke subtypes, the availability of detailed phenotyping is important. However, many studies can only examine presumed ischemic or hemorrhagic strokes or total stroke, which may mask underlying differences.

Genetic Architecture of Stroke and Its Risk Factors

Previous studies of common genetic variation suggest that total IS is ≈38% heritable (ranging from 40% for large vessel disease to 33% for cardioembolic stroke, to only 16% for small vessel disease subtypes).\(^3\) As with CHD, it is likely that most of the genetic contribution to stroke is polygenic and reflects the effects of multiple genes each exerting small effects. Genome-wide meta-analyses of IS involving >10,000 IS cases have to date, only identified a few susceptibility loci.\(^6\)\(^-\)\(^7\) Furthermore, most genetic associations with IS are subtype specific for which power is more limited. Genetic risk factors for hemorrhagic stroke are summarized elsewhere.\(^8\) Summary level data from genome-wide meta-analyses, commonly based on tens of thousands of individuals, are increasingly available for vascular risk factors (eg, blood lipids, blood pressure, and BMI). These data can be used to select genetic instruments and to construct genetic risk scores (GRSs), which are the features of most MR studies.

Understanding MR

MR is a technique for assessing causal associations in observational data and can be considered analogous to a randomized
trial with genetic variants (influencing a risk factor) akin to a randomized treatment. Under certain conditions, if a genetic variant is associated with the risk factor of interest and is also associated with the outcome of interest, then the risk factor is a cause of the outcome.5,10

MR has strengths and limitations for assessing causality (Figure 1). Pleiotropy that occurs when a genetic locus influences multiple risk factors, makes it unclear which of the risk factors, if any, are causally associated with disease and is therefore problematic for the interpretation of MR studies (Figure 2). Linkage disequilibrium between variants influencing stroke through different pathways leads to similar problems. Population stratification can result in spurious associations when the distributions of the risk factor or disease outcome differ between populations. Similarly, gene–environment and gene–gene interactions may influence conclusions for example, where only certain subgroups are affected. Canalization refers to biological adaptation in response to genetic variation and may also result in misleading estimates. The effects of genetic variants on risk factors and on disease outcomes are often small, and large sample sizes may be required to provide sufficient statistical power to detect effects on stroke. This lack of suitable genetic variants for MR may lead to weak instrument bias. GRSS that combine information across multiple genetic variants associated with the risk factor of interest are useful to maximize power and avoid such bias. However, care must be taken to select appropriate variants. MR studies can be undertaken using individual level data or summary level data (eg, from published meta-analyses), which are similarly efficient when considering uncorrelated variants. However, individual level data offer additional benefits when testing assumptions, undertaking conditional analyses, and exploring the relevance of comorbidities that may bias results when imbalanced between genotype groups.

MR has been used extensively in CHD research, for example, to demonstrate that C-reactive protein levels are not causally associated with CHD, whereas lipoprotein(a) (Lp[a]) levels are similarly efficient when considering uncorrelated variants. However, individual level data offer additional benefits when testing assumptions, undertaking conditional analyses, and exploring the relevance of comorbidities that may bias results when imbalanced between genotype groups.

MR has been used extensively in CHD research, for example, to demonstrate that C-reactive protein levels are not causally associated with CHD, whereas lipoprotein(a) (Lp[a]) levels are causally associated with CHD, and to develop our understanding of cardiovascular therapies.11,12 MR provides estimates of the impact of genetically determined risk factor levels, representing life-long differences, whereas randomized trials provide the effects of shorter-term interventions. Thus, although MR may provide insight into the relevance of the risk factor

or biological pathway of interest, estimates from MR studies and clinical trials may reasonably differ.12 Furthermore, drugs may have off-target effects and thus the impact of the biological mechanism on safety and efficacy of interventions tested in clinical trials cannot be assumed to be interchangeable.

Further details of the MR framework, methodological approaches for causal estimation, and approaches to mitigating the limitations of MR studies are presented elsewhere.10,13

What Have We Learned From Genetic Studies of Modifiable Risk Factors for Stroke?

A PubMed search (on terms [stroke AND Mendelian randomization]) gave rise to 30 results, whereas, in contrast, a similar search for CHD gave rise to 121 results. The following sections provide an overview of MR studies alongside the observational evidence.

Lipids

Randomized statin trials have demonstrated that lowering low-density lipoprotein cholesterol by 1 mmol/L with a statin reduces the risk of CHD and IS by ≈20%.14 By contrast, observational studies have reported stronger effects of low-density lipoprotein cholesterol (and non–high-density lipoprotein cholesterol) on risk of CHD than on IS.1 Thus, MR studies are needed to examine the causal relationship between low-density lipoprotein cholesterol and IS ensuring careful consideration of potential pleiotropic effects on other lipid fractions. The role of high-density lipoprotein cholesterol for IS also remains unclear. For example, the Framingham study examined the impact on IS using a GRS including 47 determinants of high-density lipoprotein cholesterol levels (describing ≈6% of the variation), but found that it had no association with IS. However, the power of this study was low and therefore a possible association could not be excluded.15 MR studies have also provided insight into biological pathways associated with lipid-modifying treatments. For example, genetic variants in the CETP gene that lower cholesterol transfer protein (CETP) activity (and influence various lipid fractions) and thus mimic the effect of CETP inhibition have been associated with lower risk of IS (and CHD), suggesting that this pathway of lipid modification may be beneficial.16

Figure 1. Key strengths and limitations of Mendelian randomization studies. Strengths are shown in green and limitations in red.

Figure 2. Illustration of pleiotropy in a hypothetical Mendelian randomization study of C-reactive protein and ischemic stroke. Pleiotropic associations with the chosen genetic instrument are shown in boxes with dashed lines.
Lipoprotein(a)
Genetic studies have established a causal relationship between Lp(a) and CHD. A meta-analysis of prospective studies, involving 1900 IS events, demonstrated a similar strength of association of Lp(a) levels for IS and CHD. In contrast, a genetic study reported that an LPA genotype score that explains about a third of the variation in Lp(a) levels, and was strongly associated with CHD, was not associated with prevalent or incident IS in those with or without previous CHD. In a subsequent study, Helgadottir et al reported that LPA alleles associated with higher Lp(a) levels were associated with higher risk of large artery IS, but not with either cardioembolic or small vessel IS. Thus, any impact of Lp(a) levels on IS may be subtype specific. Furthermore, consideration of concomitant CHD may be important for establishing any independent impact of Lp(a) on risk of stroke.

Inflammatory Biomarkers
Inflammation has been linked with both incident IS and hemorrhagic stroke. The Emerging Risk Factors Collaboration reported a 27% higher risk of IS per SD higher log C-reactive protein levels, comparable with the effect for CHD. However, genes encoding C-reactive protein were not associated with risk of CHD. Genetic studies of upstream markers occurring earlier in the inflammatory cascade, such as interleukin-6, suggest causal relationships with CHD. However, genetic variants that produce inhibition of interleukin-1, a key regulator of inflammation, were not associated with IS. Thus, it remains unclear whether inhibition of systemic inflammation or enhancement of systemic anti-inflammatory response could represent therapeutic strategies for stroke.

Renal Function
Poor renal function has been associated with higher risk of major vascular events, although effects on stroke subtypes are unclear. Holliday et al demonstrated that genetic scores associated with higher estimated glomerular filtration rate and higher urine albumin:creatinine ratio were also associated with lower risk of large artery stroke, thereby suggesting that the shared genetic components between renal disease and IS may be subtype specific.

Blood Pressure
Hypertension is a major cause of cardiovascular disease mortality worldwide, as a consequence of its high prevalence and concomitant vascular risks. Observational studies have demonstrated strong positive associations of systolic blood pressure (SBP) with both stroke and CHD mortality, with a 20 mm Hg lower usual SBP being associated with a 50% lower risk of stroke and CHD mortality. Genome-wide association studies have identified multiple genetic variants associated with various measures of blood pressure. A GRS, based on 29 genetic variants strongly associated with blood pressure traits, was associated with a =5 mm Hg difference in SBP (top to bottom fifth), and with both total stroke and CHD risk. Likewise, visit-to-visit blood pressure variability may have additional relevance for stroke independent of the usual SBP. However, genetic variants in NLGN1 that were associated with blood pressure variability were not associated with IS or its subtypes, casting doubt on the causal relevance of visit-to-visit variability in SBP for stroke.

Obesity
Observational studies have reported that higher BMI is associated with greater risk of total stroke. In those with higher BMI (25–50 kg/m²), BMI was associated positively with ischemic, hemorrhagic, and total stroke, with each of these associations largely accounted for by the effects of BMI on blood pressure. However, for those with lower BMI (15–25 kg/m²), there was no evidence of a positive association, despite the strong positive association between BMI and blood pressure. MR studies have examined the causal relevance of BMI for stroke but with varying conclusions. Holmes et al examined 3813 total stroke cases and reported no association between a BMI GRS and risk of stroke (or CHD). However, Hågg et al found a 15% increase in IS risk per SD increase in BMI (P=0.0008), suggesting a causal relationship (and similarly for CHD). More studies are needed to resolve these differences and to examine effects related to different stroke subtypes and measures of adiposity.

Alcohol
Observational studies indicate that moderate alcohol consumption is associated with a lower risk of CHD and stroke in different ethnic populations but several studies have reported a J-shaped association. A meta-analysis of 84 cohort studies reported that light to moderate alcohol consumption was associated with 25% lower risk of CHD, but had no protective effect on stroke mortality. However, it is unclear the extent to which the reported inverse associations with alcohol consumption are artifacts of confounding or bias resulting from reverse causation. A genetic study involving 261,991 individuals of European descent, including 10,164 total stroke events, reported no association of ADH1B (rs1229984) with total stroke, albeit the genotype associated with lower alcohol consumption was associated with a lower risk of IS. Thus, in addition to suggesting a causal association between alcohol consumption and IS, this MR study also demonstrated the importance of stroke subtyping.

Homocysteine
Moderately elevated blood levels of homocysteine are weakly correlated with CHD and stroke risk. When folate levels are low, the C677T polymorphism of the MTHFR gene explains =2.5 μmol/L difference in homocysteine levels. A meta-analysis of C677T and CHD reported marked heterogeneity between the results of published and unpublished studies (P=0.001), where unpublished data suggested little or no effect on CHD, whereas published data suggested a positive association (with these differences reflecting publication bias). However, the relevance of MTHFR for stroke in Asian populations is not entirely resolved as a previous meta-analysis of MTHFR suggested a possible causal effect on total stroke in low-folate Asian populations. The predicted effect of homocysteine reduction from large genetic studies in low folate regions (Asia) was larger than in Western populations.
Thus, further studies of MTHFR and stroke are required in low-folate populations to confirm the causal relevance of homocysteine and exclude methodological artifacts (controlling for population stratification and publication bias) before advocating changes in public health policy on use of folic acid for stroke prevention in Asian populations.

Lipoprotein-Associated Phospholipase A2

Lipoprotein-associated phospholipase A2 (Lp-PLA2) produces the proinflammatory mediators lysophosphatidylcholine and oxidized free fatty acids through hydrolysis of oxidized phospholipids carried on low-density lipoproteins. Although increased Lp-PLA2 activity has been associated with higher risks of occlusive vascular diseases, recent randomized trials of the Lp-PLA2 inhibitor darapladib in individuals with CHD failed to establish a protective role of darapladib for prevention of further major vascular diseases. The V279F loss-of-function variant in the PLA2G7 gene encoding Lp-PLA2 is common in East Asians (but not in Western populations) and allows an MR-based assessment of the causal effects of lifelong lower Lp-PLA2 activity, analogous to the randomized trials of Lp-PLA2 inhibition. Studies in East Asian populations have examined variants in the PLA2G7 gene and, consistent with the trials, reported no significant association with CHD or with IS. Additional studies are needed to assess the effects on IS subtypes but the evidence available currently casts considerable doubt on its relevance for stroke.

Vitamin D

Prospective observational studies have reported that higher plasma levels of 25-hydroxyvitamin D [25(OH)D] are associated with lower risks of stroke and CHD. Genetic variants encoding vitamin D–binding protein levels, a key determinant of plasma 25(OH)D, allow assessment of the causal relevance of vitamin D for stroke. In a MR study involving 12 389 IS cases and 62 004 controls, rs2282679 was associated with a 27 mg/L difference in age- and sex-adjusted vitamin D–binding protein levels (per variant allele) and a strong association with 25(OH)D levels (P=3.2×10−19), but was not associated with IS (P=0.92).

Atrial Fibrillation

Atrial fibrillation is associated with higher risk of IS typically because of an excess risk of cardioembolic events. A GRS based on genetic variants associated with atrial fibrillation has been shown to identify 20% of individuals who are at ≥2-fold higher risk of incident atrial fibrillation and 23% higher risk of IS between top and bottom fifths. Furthermore, genetic variants associated with atrial fibrillation have been more specifically associated with cardioembolic stroke. Additional genetic studies are needed to assess the interplay between atrial fibrillation and stroke.

Coronary Heart Disease

CHD is also a risk factor for IS and has many common risk factors and thus the potential for shared genetic architecture. Dichgans et al reported a substantial overlap in the genetic risk of IS, and particularly large artery stroke, with CHD. However, Cheng et al reported that major common loci associated with myocardial infarction risk do not have effects of similar magnitude on total IS, although this does not preclude moderate associations restricted to specific subtypes. These studies suggest some common, but many distinct underlying mechanisms in the pathogenesis of IS subtypes and CHD.

Diabetes Mellitus

Meta-analysis of cohort studies demonstrate over 2-fold higher risk of IS in diabetic versus nondiabetic individuals, but only a weak positive association with hemorrhagic stroke. Measures of glycemia, such as fasting glucose and glycated hemoglobin, have also been associated with an increased risk of cardiovascular disease in the absence of diabetes mellitus. Genetic variants that increase risk of diabetes mellitus have been associated with higher risk of CHD, but MR studies that assess causality of diabetes mellitus and glycated hemoglobin for stroke and its subtypes are needed.

Multiple Risk Factors

Further studies have shown that GRSs based on multiple risk factors are associated with higher risks of IS. Malik et al reported that a GRS based on SNPs associated with atrial fibrillation, CHD, hypertension, and SBP was associated with increased IS risk. Thus, genetic predisposition to the main risk factors for stroke is associated with higher risk of stroke events. Exploring intermediate phenotypes such as white matter hyperintensity and carotid intima thickness may also offer additional insights into stroke pathogenesis.

Future Promises of MR for Stroke

MR offers considerable potential for assessing causality of traditional and emerging risk factors, providing insight into biological pathways, and improving our understanding of cardiovascular drug targets as well as facilitating identification of new targets for cardiovascular and stroke-specific therapies. MR also provides a useful tool for wider omics research, such as developing our understanding of proteomic markers (and because proteins are directly encoded by genes, the potential for pleiotropy is likely to be small).

Stroke research is complicated by phenotypic heterogeneity and new recommendations for appropriate biological sample collection and standardization of phenotypic data collection will play a key role in developing future studies of stroke genetics. The comparability between Causative Classification of Stroke (CCS) and Trial of ORG 10172 in Acute Stroke Treatment (TOAST) classification of IS subtypes has also been examined and will further facilitate large-scale genetic meta-analyses. Exploring genetic factors in different ethnic populations, for example in China where stroke is more common than in the West, will be important for determining the underlying distribution of different stroke subtypes and their risk factors. Studies such as the China Kadoorie Biobank, a prospective study of 500 000 individuals from across China, and the Risk Assessment of Cerebrovascular Events Study (RACE), a study of stroke in South Asians, will facilitate this effort. The MEGASTROKE international collaborative study, which focuses on advancing...
genetic research of stroke and its subtypes, is likely to acquire over 70000 stroke cases across a range of ethnicities in the coming months and will greatly enhance the statistical power of genetic studies for stroke, leading to new genetic discoveries and greater potential for MR experiments. Such studies will inevitably improve our understanding of the biological pathways involved and contribute to the development of novel therapeutic targets as well as enable us to anticipate their potential impact on stroke ahead of outcome trials. MR studies of stroke have had limited impact, to date, and have yet to achieve their full potential. However, with the advent of larger well-phenotyped studies, we are likely to see rapid growth in genetic-based stroke research and corresponding translational impact.

**Sources of Funding**
The Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford receives grants from the pharmaceutical industry for research conducted independently of all sources of funding (http://www.ctsu.ox.ac.uk/about-ctsu/documents/independent-research).

**Disclosures**
Associate Professor Hopewell holds a British Heart Foundation Intermediate Basic Science Research Fellowship (FS/14/55/30806). Professor Clarke is a coapplicant on research grants from the UK Medical Research Council (Clinical Trial Service Unit A310) and Wellcome Trust (104085/Z/14/Z).

**References**


