



HHS Public Access

Author manuscript

Stroke. Author manuscript; available in PMC 2021 November 01.

Published in final edited form as:

Stroke. 2020 November ; 51(11): 3406–3416. doi:10.1161/STROKEAHA.120.030424.

Approaches to Studying Determinants of Race-Ethnic Disparities in Stroke and its Sequelae

Mitchell S. V. Elkind, MD^{1,2}, Lynda Lisabeth, PhD³, Virginia J. Howard, PhD⁴, Dawn Kleindorfer, MD⁵, George Howard, DrPH⁶

¹Department of Neurology, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA

²Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA

³Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA

⁴Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA

⁵Department of Neurology, University of Michigan, Ann Arbor, MI, USA

⁶Department of Biostatistics, UAB School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA

Abstract

Disparities are differences in health outcomes among groups that originate from sources including historically experienced social injustice and broadly-defined environmental exposures. Large health disparities exist, defined by many factors including race/ethnicity, sex, age, geography and socioeconomic status. Studying disparities relies on measures of disease burden. Traditional measures, such as mortality, may be less applicable to neurological disorders, which often lead to substantial morbidity and lower quality of life, without necessarily causing death. Measures such as disability adjusted life years or healthy life expectancy may be more appropriate for assessing neurological disease, and permit comparisons across diseases and communities. There are many approaches that can be used to study disparities. Analyses of population-based observational studies, patient registries, and administrative data all contribute to the understanding of disparities in humans. Animal and other experimental designs, including clinical trials, may be used to identify mechanisms and strategies to reduce disparities. All of these approaches have strengths and weaknesses. Ultimately, understanding and mitigating disparities will require use of all of these methods. Crucially, a focus on not only improving outcomes among all individuals in society, but minimizing or eliminating differences between those with better outcomes and those who have historically been disadvantaged, should drive the ongoing investigations into disparities. This review is focused on epidemiologic approaches to examining the depth and determinants of race-ethnic disparities in the United States related to stroke, stroke care, and stroke outcomes.

Keywords

Cerebrovascular Disease; Cohort study; Disparities; Epidemiology; Race and ethnicity; Stroke

Introduction

The coronavirus pandemic of 2020 has brought into stark relief the inequities in the American healthcare system. Black Americans, as well as other disadvantaged groups, have contracted coronavirus disease at higher rates and disproportionately suffered death from the pandemic.¹ Geographic differences have also played a powerful role, with the pandemic initially having the heaviest burden in the New York City area, but with this burden subsequently shifting to other regions of the country. These disparities in outcomes reflect non-biological factors, such as housing quality, environmental pollution, employment, education, governmental approaches and policies, and other social determinants of disease. Structural racism, the subject of protests across the US and globally in the summer of 2020, sparked by the police killing of George Floyd and others, is itself a social determinant of health.² These inequities and the resulting protests indicate the need for increased investment in study of disparities.

Epidemiologists have long recognized similar disparities in, and similar fundamental contributors to, the occurrence of stroke and outcomes among stroke patients. This invited review grew out of the Health Equity and Actionable Disparities in Stroke: Understanding and Problem-solving (HEADS UP) Symposium at the 2020 International Stroke Conference. It is aimed at early stage investigators with an interest in stroke disparities, and more established epidemiological investigators with a new interest in disparities, particularly those focused on race. We will first describe basic concepts important to the study of determinants of race-ethnic disparities in stroke, including measures of the burden of disease, such as disability-adjusted life years and healthy life expectancy. We will consider the essential distinction between strategies to improve stroke-related outcomes among all groups, which may have benefits for disparity populations but still preserve the disparity, and those strategies which actually reduce the differences between groups. We will also discuss different study designs and resources available to disparities researchers, including observational studies, patient registries, and administrative datasets; advantages and disadvantages to these approaches will be considered. Finally, experimental study designs will be briefly addressed, including animal studies and human clinical trials.

Basic concepts

The United States Department of Health and Human Services, in its Healthy People 2020 document, defined health disparities as "...a particular type of health difference that is closely linked with social, economic, and/or environmental disadvantage."³ The document further states that "Health disparities adversely affect groups of people who have systematically experienced greater obstacles to health based on their racial or ethnic group... or other characteristics historically linked to discrimination or exclusion." Thus, disparities are not only differences in health outcomes between groups, but a *difference with a*

difference. For example, the difference could be linked to membership in a particular group that has historically experienced social injustice of one type or another. One of the largest health disparities in the US is that based on race. Investigators have traced some race disparities in stroke mortality back to historical injustices rooted in slavery.⁴ In Canada, differences in cardiovascular risk factors, particularly type 2 diabetes in youth, have analogously been associated with the multigenerational legacy of colonization.⁵ Other types of disparities exist, including those based on ethnicity, age, sex, sexual orientation, gender identity, physical disability, regional or geographic location, religious affiliation, educational attainment, socioeconomic status, immigration and documentation status, acculturation, language proficiency, literacy, housing stability, mental health, legal needs, and others.

Overall, health disparities pose an enormous public health problem, affecting the majority of Americans. Approximately 1/3 of people in the US, or more than 100 million people, identify as belonging to a racial or ethnic minority population. Just over 50% of the US population are women (>160 million people), and approximately 45 million (15%) live in the southeastern “stroke belt” states. Approximately 12%, or 36 million people not living in nursing homes or other residential care facilities, have a disability. Another estimated 4% of the US population ages 18 to 44 identify as lesbian, gay, bisexual, or transgender. Approximately a quarter of the population, or 71 million people, live in rural areas. All of these groups, and others, are subject to disparities. Being part of a large group, therefore, does not preclude being subject to disparities. Health disparities do not reflect numbers or minority status alone, but rather the systematic and historical challenges to adequate health and healthcare.

Assessing disparities in stroke and other neurological disorders requires using a metric that captures the burden of disease. Several measures may be used to evaluate disease burden, including prevalence, incidence, case fatality, mortality, disability, quality of life, cost, and others. Incidence and prevalence refer to how common a disease is. The number of cases of a disease over a defined time interval in a defined population free of disease at the start of the time interval represents its cumulative incidence, and is commonly calculated for first cases only (initial incidence). Prevalence measures the total number of cases, new or old, at a particular time or over a period of time. Both indices depend on the accurate and complete enumeration of cases and adequate knowledge of the underlying population at risk. Determination of both cases and the size of the underlying population may be more difficult in under-resourced populations, as diagnostic testing to establish the diagnosis may not be available, or individuals, such as immigrant populations, may not allow themselves to be readily counted.

Mortality refers to the overall number of deaths due to a disease in a given time period, while case fatality refers to the proportion of patients who die from the disease in a shorter time window after disease onset, typically 30 days. Traditional approaches to estimating the impact of illness have focused on mortality because it is relatively easy to measure, at least in developed countries. While mortality is generally considered a “hard” outcome and thus subject to less bias than other measures, attributing a death to a specific cause, such as stroke, may be difficult in some circumstances, since diagnosis and documentation are still required. Establishing cause of death is particularly challenging among individuals with

multiple comorbidities, such as the elderly. In race-ethnic groups that also have multiple comorbidities at younger ages, it may also become difficult to establish stroke as the cause of death, potentially leading to underestimates of disparities in stroke mortality. Nonetheless, the impact of some neurologic disorders may be captured by consideration of their mortality. For example, malignant brain tumors, severe strokes, and head injuries frequently lead to death. A metric focused exclusively on mortality, however, fails to capture the impact of many neurological disorders that are chronic and slowly progressive, or intermittent and disabling, but that do not cause their sufferers to die. For example, cerebral small vessel disease, vascular dementia, and migraine have low mortality but moderate to high disability and are extremely common. Attempts to measure the disease burden of these disorders therefore requires a more versatile metric than mortality alone.

One common approach to measuring burden of disease is the disability-adjusted life year (DALY). The DALY is a time-based metric incorporating both premature mortality (the number of years of life lost due to premature death, based on an expected life span) and disability (years of healthy life lost as a result of disability weighted by the severity of the disability). DALYs are a well-established measure of disease burden that includes the number of healthy years of life lost as a result of both death and disability caused by a particular disease. An advantage to using the DALY is that it permits comparisons of burden across very different disease states, in effect serving as a common measure for acute severe illness (hemorrhagic stroke, malignant infarction, myocardial infarction) and chronic, debilitating, yet rarely fatal illnesses (vascular dementia, migraine). The DALY metric thus reflects the impact of disease on both early mortality and on disability, both of which are particularly important for assessing the overall burden of neurologic disease. It is important to note, moreover, that the “weighting” of disability (or utilities of a given health state) may themselves depend on cultural values regarding how disability is perceived, so that disparities may also influence the measured number of DALYs.

While DALYs emphasize the number of healthy years lost to illness, another related metric, Health Adjusted Life Expectancy, or HALE, focuses instead on the healthy lifespan. Life expectancy is the average number of years that a person is expected to live based on current mortality rates in the population. The HALE is simply life expectancy discounted for disease and disability. Years of life lived with a hemiparesis, for example, would count less to HALE than years lived in full health, while both would count equally toward life expectancy. In order to estimate HALE, both how many people a disease affects and its impact on health and function of the disease on the people who have it must be measured. Combining these two effectively creates a measure of disease-free life expectancy weighted by the severity of the disease. The HALE was used recently by the American Heart Association (AHA) in a statement of its goals for 2030, which included to equitably increase health-adjusted life expectancy in the US by 2 years, from 66 to 68 years.⁶ Of note, the statement explicitly called for doing so equitably, reflecting the mission of the AHA to mitigate health disparities. Arguably, increasing HALE for all Americans by 2 years, while of benefit for all, would not necessarily reduce disparities, since the benefits would be seen evenly, and differences would not necessarily be diminished, as illustrated in the next section.

Mitigating disparities versus lifting all boats: Examples from comparisons of Black-White Stroke Disparities

It is important to note that improving the health of a particular population that has been historically disadvantaged does not necessarily imply a reduction in the disparity, or difference between that group and others. Over the past 20 years, for example, there has been a dramatic decline in stroke mortality that has been shared by all race-ethnic groups. For those aged 45 years or greater, between 1999 and 2018, age-adjusted stroke mortality declined 40% for White people, and 37% for Black populations.⁷ While we should celebrate that this decline has lowered the burden of stroke for both Whites and Black populations, we should also be concerned that it represents a total failure to reduce the relative disparity of stroke mortality in Black people. In fact, in 1999 Black individuals suffered a 36% higher age-adjusted burden of stroke death, and this relative burden increased to 43% in 2018. In fact, the Black-White disparity in stroke mortality was first noted in 1960 in a report documenting disparities dating back to 1950;⁸ in the subsequent 70 years there has been little change in the relative Black-White disparity in stroke.⁹

The higher stroke mortality in Black populations could be a product of: 1) a higher incidence of stroke in Black people, 2) a higher case-fatality in Black stroke patients, or 3) both a higher incidence and higher case-fatality (Figure). Whether the Black excess is a product of a higher incidence or higher case-fatality has profound implications for tailoring interventions to reduce the racial stroke disparity. If the disparity in stroke mortality is a result of higher stroke incidence in Black populations, then interventions need to target the disparity in stroke incidence and would likely include community-based efforts in primary stroke prevention. Conversely, if the disparity is a product of higher case-fatality, then interventions would need to focus on hospital-based efforts to improve stroke outcomes in Black populations. Both the Greater Cincinnati/Northern Kentucky Stroke Study (GCNKSS) and the REasons for Geographic And Racial Differences in Stroke (REGARDS) study have shown the overwhelming contributor to Black-White differences in stroke mortality is racial differences in stroke incidence, with racial differences in case-fatality playing a very minor role.^{10,11} In REGARDS the Black-to-White incidence ratio is 2.39 [95% CI: 1.71 – 3.32] between ages 45–54, and then decreases 1.90 [95% CI: 1.53 – 2.36] for ages 55–64, 1.51 [95% CI: 1.32 – 1.72] for ages 65–74, 1.20 [95% CI: 1.05 – 1.37] for ages 75–84, and 0.95 [95% CI: 0.76 – 1.94] for those 85 years and older.¹⁰ There was no evidence of an age-related change in the Black-to-White case fatality rates ($p>0.01$), and the Black-to-White odds of death after stroke was not significant [OR = 1.20; 95% CI: 0.89 – 1.62]. The GCNKSS showed a similar decreasing Black-to-White incidence ratio from 2.64 [95% CI: 2.53 – 2.76] for ages 45–54, 1.82 [95% CI: 1.73 – 1.91] for ages 55–64, 1.18 [95% CI: 1.11 – 1.25] for ages 65–74, 0.92 [95% CI: 0.86 – 0.97] for ages 75–84, and 0.77 [95% CI: 0.73 – 0.82] for ages 85 and older.² The case-fatality data failed to show any Black-to-White difference in the risk of death after stroke. Therefore, it seems most promising to focus efforts to reduce the racial disparity on stroke prevention efforts.

Mitigating the racial disparity in incidence will require additional highly-focused research to understand and intervene on the specific contributors to the Black-White disparity. Although

a simplification, there are two major pathways through which a risk factor could contribute to racial disparity: 1) a factor could be a risk factor for stroke shared by Black and White populations, with a higher prevalence in the Black population, or 2) a factor could have a differential association with stroke risk by race with a higher impact in Black populations.

As an example of the first pathway, many traditional stroke risk factors (for example, the Framingham Heart Study stroke risk profile¹²) have a higher prevalence in Black than White communities. For instance, in the REGARDS Study, Black people had a prevalence of diabetes of 29% and a mean systolic blood pressure of 130 mm Hg, with 62% treated with antihypertensives; Whites, in contrast, had a prevalence of diabetes of 15% and a mean systolic blood pressure of 125 mm Hg, with only 42% on treatment.¹³ One would therefore presume that a part of the higher stroke risk in Black individuals is a result of the higher prevalence of stroke risk factors in the Black population. Research can focus specifically on the contribution of racial disparities in stroke risk factor prevalence to the racial disparities in stroke risk through mediation analysis,¹⁴ which provides a direct estimate of the proportion of the racial disparity in stroke risk that is attributable to the racial disparity in the prevalence of stroke risk factors. The use of such mediation analyses assumes that race is causally associated with the risk factor (e.g, hypertension). It is more likely, however, that race itself is not causally associated with a higher prevalence of risk factors, but associated with it due to common upstream fundamental causes of disparities, such as lifestyle choices including diet and exercise, early life experiences, residential segregation, racism, etc. Analyses should therefore be optimally adjusted for such potential confounders, which may be difficult with some existing datasets.

Whether the associations of race and risk factors are causal or not, these analytic approaches may provide insights into potential interventions by identifying the specific risk factors where a reduction in the disparity in risk factor prevalence would be more likely to result in a reduction in the disparity in stroke risk. For example, in the REGARDS study, at age 55, the stroke risk was estimated to be 2.20 times (95% confidence interval (CI) 1.15 – 3.12) higher for Black people than White people. Adjustment for the traditional risk factors (systolic blood pressure, use of antihypertensive medications, diabetes, atrial fibrillation, heart disease, left ventricular hypertrophy and smoking) resulted in a 41% decrease in the Black excess risk to 1.70 (95% CI 1.18–2.45), suggesting that slightly less than half of the Black excess stroke risk is attributable to a higher prevalence of these risk factors.⁸ For the first pathway, the focus of interventions to reduce excess stroke risk in Black populations should focus on preventing the heavier burden of stroke risk factors. If interventions are not specifically targeted at Black populations, however, general population-wide improvement of risk factor control (without reducing disparities in risk factor prevalence) would result in improvement in both Black and White populations, which would not necessarily have an impact on the disparity in stroke risk.

Even trials of interventions that are limited to Black populations will not necessarily confirm the hypothesis that such interventions have an impact on reducing disparities if they are not similarly tested in non-Black populations and found less effective. As an example, Black-owned barbershops can identify Black individuals with hypertension and refer them to treatment that will reduce blood pressure effectively,¹⁵ but these results by themselves do not

necessarily demonstrate that such interventions reduce disparities. As a public health matter, these approaches are important, and their basis in population-specific theoretical models of behavior change are likely to lead to a differential effect in the Black population that will narrow the gap between populations. As a scientific matter, however, they do not demonstrate approaches that reduce disparities *per se*. Trials that address strategies to reduce stroke disparities would ideally focus not only on reducing the prevalence of risk factors or the incidence of disease among one population or another, but should demonstrate that some interventions have a *differential* effect in one population versus another. Only interventions that have a greater impact in reducing disease among Black individuals than among White people, for example, can legitimately be said to reduce Black-White disparities. Statistically speaking, such interventions would interact with race-ethnicity and provide a greater impact among minority groups. Studies should be designed with this concern in mind. Interventions that are broadly beneficial may help to drive overall rates down but may still fall short of decreasing disparities *per se*, as discussed above.

Interventions would differ substantially for factors that contribute to stroke disparities through having a higher impact in Black populations (the second pathway above). These are factors where the association with stroke risk differs by race, which can be assessed by analyses focusing on effect modification. For example, in REGARDS, overall a 10 mm Hg difference in systolic blood pressure is associated with a 14% (95% CI: 8% – 21%) difference in risk of incident stroke.¹⁶ This difference proved to differ significantly between Black and White people ($p = 0.049$), with an 8% (95% CI: 0% - 16%) difference in White individuals and a 3-times greater difference of 24% (95% CI: 14% - 35%) in Black people. Unlike factors impacting the racial disparity in stroke risk through the first pathway, this suggests that a similar prevalence of hypertension in Black and White people would have a larger impact for the Black than White population. Additional factors that could influence ascertainment of prevalence, such as failure to screen or diagnose, and the earlier onset and longer duration of hypertension in Black populations, will also play a role. For factors falling into this paradigm, a reduction in the prevalence would likely be associated with an associated reduction in the disparity of stroke risk. For those factors with a larger impact in the Black population, interventions could target the pathways resulting in the more potent effect of the risk factor in Black people (for example, potentially focusing on salt sensitivity for a larger impact of hypertension in Black patients). For example, there is evidence that measurement of plasma renin and aldosterone to physiologically target Black patients for identification of salt sensitivity can improve blood pressure control.¹⁷

Some risk factors, of course, could contribute to the stroke disparity in Black populations through both pathways. For example, in addition to the more potent contribution of hypertension in Black patients (the second pathway), hypertension also is likely to contribute to the disparity in stroke through a higher prevalence in Black populations (the first pathway). Statistical methods, including decomposition methods, are an increasingly used tool to understand the contributions of the two pathways to underlying health disparities. These approaches have rarely been used in stroke research.

Observational epidemiological studies

Longstanding observational epidemiological studies are essential to studying and monitoring progress towards reducing disparities. The Framingham Heart Study began in 1948, and remains the paradigmatic observational cohort study of risk factors for cardiovascular disease, including stroke. The study has evolved over the years to include a wide range of risk factor exposures and outcomes, including brain imaging and neuropsychological testing, that may be of interest to those interested in disparities in neurological disease. The population (n=5209) of the initial Framingham Study, however, was entirely white. Thus, other cohorts were created to study risk factors and cardiovascular outcomes in more diverse populations (Table 1). As illustrative examples, the Jackson Heart Study¹⁸ comprised 5306 Black people in Jackson, MS; the Strong Heart Study¹⁹ included 3516 American Indians from Arizona, Oklahoma, North Dakota, and South Dakota; and the Hispanic Community Heart Study/Study of Latinos (HCHS/SOL)²⁰ included 16,415 Hispanics/Latinos from San Diego, CA, Chicago, IL, Bronx, NY, and Miami, FL. Each of these studies permits the study of cohorts of Americans from historically disadvantaged racial and ethnic groups, but comparisons among race-ethnic groups based on these studies are limited due to the cohorts utilizing different study designs, inclusion and exclusion criteria, and methods, and variability in environments of each cohort. Subsequent observational epidemiological studies that include multiple race-ethnic groups within a single study, using the same methodology, may permit more reliable data regarding disparities, depending on how the cohorts are selected. Studies that include multiple race-ethnic groups within a single study design include the Cardiovascular Health Study,²¹ the Atherosclerosis Risk in Communities (ARIC) study,²² and the Multi-ethnic Study of Atherosclerosis (MESA).²³ A persistent limitation of these cohorts, though, is that the populations under study may still live in disparate communities. Additionally, these studies sometime confound race with geography, for example with the majority of the Black participants in ARIC being from the all-Black clinical center in Jackson, MS, and with select sites in MESA recruiting specific race/ethnic groups. Thus, it becomes impossible to disentangle the effects of race/ethnicity from place. An alternative, exemplified by the Northern Manhattan Study²⁴ and other studies,^{25, 26, 27} is to include members of different race-ethnicities living within a single geographic location. The REGARDS study allows national representation of individuals, but race/ethnic heterogeneity limited to only Whites and Black people.^{10,28} Other cohorts and national registries have been used to study disparities among indigenous,²⁹ immigrant,³⁰ and multinational³¹ populations outside the US.

These different studies reflect different ways of investigating disparities (Table 2): comparing distinct race-ethnic population in different locations using different study designs; comparing distinct race-ethnic communities in different locations but within the same study design; comparing individuals from different race-ethnic groups within one location using the same study design; and comparing individuals representing different race-ethnic groups from different locations using a common study design. Finally, meta-analyses and studies of combinations of these cohorts provide opportunities for investigators to achieve larger sample sizes to address less common exposures or to explore questions that may require participants of different ages (Table 2).³² The BP-Cog multi-cohort study, for

example, was designed to explore the effects of blood pressure across the lifecourse on risk of cognitive decline and dementia.³³

Access to these cohorts is a tremendous resource for young investigators planning a career in disparities research, and mentorship and networking opportunities from members of these research groups are generally readily available³⁴ (Table 3). Research teams often include specialists across many disciplines, including neurology, cardiology, neuropsychology, neuroimaging, epidemiology, genetics, biostatistics and others. Junior investigators can often apply for ancillary grants to study these cohorts. Advantages to using large cohort studies to study disparities include the availability of extensive, detailed data, often including granular data on stroke subtype and stroke severity; and availability of previously collected blood and other samples in many cases. There are some disadvantages to studying the cohorts, as well. It may be difficult to study rare diseases, for example. From a practical point of view, these studies may also have bureaucracies to manage, and investigators will need to learn how data are collected and interpreted. In many cases, investigators may need to provide funding for analyses or to obtain data, though often junior investigators may be able to ask for preliminary or first analyses to be conducted. There is also the challenge of working with mentors who may be available only remotely (“telementoring”).

Several cross-sectional surveys (notably the nationally-representative National Health and Nutrition Examination Survey (NHANES)), provide the opportunity for the direct assessment of prevalence of stroke and stroke risk factors.³⁵ NHANES has been conducted in several two-year “cycles,” providing the opportunity to assess temporal changes in prevalence. Additionally, follow-up has been performed on specific cycles of the NHANES assessment, allowing assessment of the association of risk factors with incident stroke. Similarly, the Behavioral Risk Factor Surveillance System provides a very large (150,000 – 200,000 annually) telephone assessment of self-reported health conditions (including stroke and stroke risk factors), allowing a very precise estimate of these self-reported conditions.³⁶

Registries and administrative databases

Patient registries may also provide opportunities for stroke disparities research (Table 3). The AHA Get With The Guidelines-Stroke registry, for example, has grown tremendously since 2000 when it began. It currently captures over 80% of hospitalized strokes that occur annually in the US, and the database includes more than 2000 hospitals and over 6 million patient records.³⁷ Registries are particularly powerful for the characterization of stroke events, examination of some clinical predictors of stroke outcomes, and quality of care because of their large sample size and the inclusion of many care process measures. This design is limited in the ability to estimate associations with risk factors because of the lack of a stroke-free control group. In addition, disadvantages include the potential for a biased sample of hospitals and an absence of population-based data, the capture of patients at the time of a clinical event rather than before, and absence of granular research data on laboratory results or imaging data. In addition, smaller registries may be biased by the absence of minority patients or unknown race/ethnicity of patients. Nonetheless, important observations about disparities in stroke care may be captured by appropriately designed studies. For example, in an analysis of Get With The Guidelines-Stroke, it was found that

White women were most likely to arrive at the hospital via ambulance (62%), while Hispanic men were least likely to do so (52%).³⁸ After adjustment for patient characteristics, Hispanic and Asian men and women had a 20 to 29% lower odds of using an ambulance compared to Whites; Black women were 25% less likely than White women to use EMS.

Large administrative databases may also provide opportunities to study disparities (Table 3). These databases, well reviewed elsewhere,³⁹ include information generated from routine encounters between patients, providers, and healthcare systems. The data may include information on hospitalizations, emergency department visits, outpatient appointments, and medication prescriptions and pharmacy dispensing. These data are generally collected primarily for administrative, regulatory, or billing purposes, and thus may be limited in clinical detail. The sources of information include hospitals; local, state or federal health departments or governmental agencies; and payers, such as Medicare, Medicaid, the Veterans Administration, and commercial insurers.

There are several scientific advantages to using these databases for analyses (Table 3). The numbers included are often extremely large, thus enhancing analytic power and permitting study of rare diseases or diseases in subgroups of patients.^{40,41,42} These databases also include a broadly representative population, allowing greater generalizability of findings. They may also be geographically broad, permitting study of variation in practice, care, and outcomes. Databases that have been operational for many years also allow analyses of temporal trends in data.⁴³ For example, analyses of temporal trends in blood pressure and diabetes control show a reduction over time in disparities in risk factors after age 65, when Medicare coverage becomes universal in the US.⁴⁴ In many cases, they can also be linked to other datasets, including observational cohort studies discussed above, to provide additional exposure or outcome data that may not be available in an initial dataset.⁴⁵ This may, for example, allow regional estimates for the effects of social determinants of disease, such as pollution, green space, crime, and other measures.^{46,47} Practical strengths of these administrative databases include their availability and accessibility; the low cost of use (with some exceptions); the ability to link to other databases; and the absence of identifying information about patients, which will generally allow institutional review board of the waiver of the need for review.

The use of administrative databases is not without limitations. Scientifically, these weaknesses include missing data and systematic biases in reporting. Because diagnoses are based on coding for administrative or billing purposes, there may be concerns about their accuracy and variation by geographic region. It is important for investigators to provide assessments of the validity of the data, including exposures and outcomes. Race-ethnic identification, in particular, may be subject to bias. Conducting sensitivity analyses that vary the use of diagnostic codes or other assumptions help to round out papers using these data and improve their acceptability. There is also often a lack of granularity of data. While procedure codes may permit assessment of the frequency and type of diagnostic testing, it will not provide information about the results of such tests. Thus, these analyses may be useful for assessing processes, quality and costs of care, but may be less useful for determining biological associations. For instance, investigators found that among 19,639 Medicare fee-for-service beneficiaries with ischemic stroke, Black patients were

significantly less likely than Whites to receive carotid imaging after risk adjustment.⁴⁸ Despite their limitations, analyses of administrative databases can be used to provide “hypothesis-generating” preliminary data that set the stage for the development of further studies and funding opportunities. They may be considered an important component of a well-rounded stroke disparities research program, if not the entire buffet.

Experimental studies

In addition to clinical and population research, laboratory-based research and animal models can play a role in understanding stroke disparities just as they do for other biologically-mediated aspects of neurological disease. For example, experiments have shown that social isolation, animal housing arrangements, and depression contribute to stroke severity, recovery, and cognition.^{49,50} These experiments do more than document the existence of these pathways in animals; they also provide mechanistic understanding, potentially pointing the way toward therapies that could be used to ameliorate the adverse effects of disparities. Specific microRNAs, for example, appear to play a role in mediating the effects of social isolation on adverse outcomes.⁵¹ Animal experiments may thus be used to address disparities using similar mechanistic models.

Clinical trials can also address disparities. Trials that address strategies to reduce stroke disparities must focus not only on reducing the prevalence of risk factors or the incidence of disease among one population or another, but must strive to identify interventions that have a *differential* effect in one population versus another, as described above. As an example of what such a trial could look like, the Baby’s First Years (BFY, <https://clinicaltrials.gov/ct2/show/NCT03593356>) trial is the first randomized controlled trial to systematically assess the role of economic resources in early childhood development by evaluating whether unconditional cash gift payments have a causal effect on the cognitive, socio-emotional and brain development of infants and toddlers in low-income U.S. families. In BFY, 1,000 new mothers of infants with incomes below the federal poverty line from four diverse U.S. communities will receive monthly cash gift payments by debit card for the first 40 months of the child’s life. Parents in the experimental group will receive \$333 per month while those in the active control group will receive a nominal monthly payment of \$20. Outcomes at age 3 include measures of cognitive, language, memory, self-regulation and socio-emotional development. Although not explicitly included in the study design, comparison of outcomes among those with higher versus lower resources could provide evidence that the intervention of cash payments has a differential effect on those with limited resources.

Conclusions

There are many approaches that can be used to study race-ethnic disparities in stroke. Analyses of population-based observational studies, patient registries, and administrative data, may all be used to understand these disparities. Animal and other experimental designs, including clinical trials, may be used to identify mechanisms and strategies to reduce disparities. All of these approaches have strengths and weaknesses. Ultimately, understanding and mitigating disparities will require use of all of these methods. Additionally, a focus on not only improving outcomes among all patients, but minimizing or

eliminating differences between those with better outcomes and those who have historically been under-resourced should drive these approaches.

Acknowledgments

Conflicts of Interest

Dr. Elkind receives royalties from UpToDate for chapters on stroke and ancillary support from Roche and BMS-Pfizer Alliance for Eliquis for an NINDS-funded trial of stroke prevention.

Dr. Lisabeth has support from NINDS and NHLBI for stroke disparities related research.

Dr. Virginia Howard reports no disclosures.

Dr. Kleindorfer reports no disclosures.

Dr. George Howard has support from NINDS for population-based studies of race-ethnic disparities.

Non-standard Abbreviations and Acronyms

AHA	American Heart Association
ARIC	Atherosclerosis Risk in Communities
BFY	Baby's First Years
CHS	Cardiovascular Health Study
DALY	Disability-adjusted life year
HALE	Health Adjusted Life Expectancy
GCNKSS	Greater Cincinnati/Northern Kentucky Stroke Study
HCHS-SOL	Hispanic Community Heart Study/Study of Latinos
MESA	Multiethnic Study of Atherosclerosis
NHANES	National Health and Nutrition Examination Survey
NHLBI	National Heart Lung and Blood Institute
NINDS	National Institute of Neurological Disorders and Stroke
NOMAS	Northern Manhattan Study
REGARDS	REasons for Geographic And Racial Differences in Stroke

References

1. Yancy CW. COVID-19 and African Americans. *JAMA*. 2020;323:1891–1892. [PubMed: 32293639]
2. Cogburn CD. Culture, Race, and Health: Implications for Racial Inequities and Population Health. *Milbank Q*. 2019;97:736–761. [PubMed: 31512293]
3. Office of Disease Prevention and Health Promotion, United States Department of Health and Human Services. Disparities. Healthy People 2020 Available at <https://www.healthypeople.gov/2020/about/foundation-health-measures/Disparities> Accessed June 20, 2020.

4. Esenwa C, Ilunga Tshiswaka D, Gebregziabher M, Ovbiagele B. Historical Slavery and Modern-Day Stroke Mortality in the United States Stroke Belt. *Stroke*. 2018;49:465–469. [PubMed: 29335332]
5. Tobe SW, Maar M, Roy MA, Warburton DE. Preventing Cardiovascular and Renal Disease in Canada's Aboriginal Populations. *Can J Cardiol*. 2015;31:1124–9. [PubMed: 26321434]
6. Angell SY, McConnell MV, Anderson CAM, Bibbins-Domingo K, Boyle DS, Capewell S, Ezzati M, de Ferranti S, Gaskin DJ, Goetzel RZ, et al. The American Heart Association 2030 Impact Goal: A Presidential Advisory from the American Heart Association. *Circulation*. 2020;141:e120–e138. [PubMed: 31992057]
7. Compressed Mortality File 1968–2018. CDC WONDER Online Database, compiled from Underlying Cause of Death Mortality File 1968–2018. Centers for Disease Control and Prevention Accessed March 14, 2020.
8. Wylie CM. Recent trends in mortality from cerebrovascular accidents in the United States. *J Chronic Dis* 1960;14:213–220.
9. Howard G Ancel Keys Lecture: Adventures (and Misadventures) in Understanding (and Reducing) Disparities in Stroke Mortality. *Stroke*. 2013 44: p. 3254–9. [PubMed: 24029634]
10. Howard G, Moy CS, Howard VJ, McClure LA, Kleindorfer DO, Kissela BM, Judd SE, Unverzagt FW, Soliman EZ, Safford MM, et al. Where to Focus Efforts to Reduce the Black-White Disparity in Stroke Mortality: Incidence Versus Case Fatality?. *Stroke*. 2016;47:1893–1898. [PubMed: 27256672]
11. Kleindorfer D, Broderick J, Khoury J, Flaherty M, Woo D, Alwell K, Moomaw CJ, Schneider A, Miller R, Shukla R, et al. The unchanging incidence and case-fatality of stroke in the 1990s: a population-based study. *Stroke*. 2006;37:2473–2478. [PubMed: 16946146]
12. Wolf PA, D'Agostino RB, Belanger AJ, Kannel WB. Probability of stroke: a risk profile from the Framingham Study. *Stroke*. 1991;22:312–318. [PubMed: 2003301]
13. Howard G, Cushman M, Kissela BM, Kleindorfer DO, McClure LA, Safford MM, Rhodes JD, Soliman EZ, Moy CS, Judd SE, et al. Traditional risk factors as the underlying cause of racial disparities in stroke: lessons from the half-full (empty?) glass. *Stroke*. 2011;42:3369–3375. [PubMed: 21960581]
14. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. *Annu Rev Psychol*. 2007;58:593–614. [PubMed: 16968208]
15. Victor RG, Lynch K, Li N, Blyler C, Muhammad E, Handler J, Brettler J, Rashid M, Hsu B, Foxx-Drew D, et al. A Cluster-Randomized Trial of Blood-Pressure Reduction in Black Barbershops. *N Engl J Med*. 2018;378:1291–1301. [PubMed: 29527973]
16. Howard G, Lackland DT, Kleindorfer DO, Kissela BM, Moy CS, Judd SE, Safford MM, Cushman M, Glasser SP, Howard VJ. Racial differences in the impact of elevated systolic blood pressure on stroke risk. *JAMA Intern Med*. 2013;173:46–51. [PubMed: 23229778]
17. Akintunde A, Nondi J, Gogo K, Jones ESW, Rayner BL, Hackam DG, Spence JD. Physiological Phenotyping for Personalized Therapy of Uncontrolled Hypertension in Africa. *Am J Hypertens*. 2017;30:923–930. [PubMed: 28472315]
18. Jackson Heart Study. Available at <https://Jacksonheartstudy.org/> Accessed July 3, 2020.
19. Strong Heart Study. Available at <https://strongheartstudy.org/> Accessed July 3, 2020.
20. Hispanic Community Heart Study/Study of Latinos. Available at <https://www.nhlbi.nih.gov/science/hispanic-community-health-studystudy-latinos-hchssol#:~:text=The%20HCHS%2FSOL%20study%20found,among%20Hispanics%20of%20various%20backgrounds> Accessed July 3, 2020.
21. The Cardiovascular Health Study. Available at: <https://chs-nhlbi.org/> (Accessed July 3, 2020.)
22. Atherosclerosis Risk in Communities Study Description. Available at https://sites.csc.unc.edu/atic/desc_pub Accessed July 3, 2020.
23. Multi-ethnic Study of Atherosclerosis. Available at: <https://www.nhlbi.nih.gov/science/multi-ethnic-study-atherosclerosis-mesa#:~:text=The%20Multi%2DEthnic%20Study%20of,other%20parts%20of%20the%20body> Accessed July 3, 2020.

24. Gardener H, Sacco RL, Rundek T, Battistella V, Cheung YK, Elkind MSV. Race and Ethnic Disparities in Stroke Incidence in the Northern Manhattan Study. *Stroke*. 2020;51:1064–1069. [PubMed: 32078475]
25. Kleindorfer DO, Khoury J, Moomaw CJ, Alwell K, Woo D, Flaherty ML, Khatri P, Adeyoye O, Ferioli S, Broderick JP, et al. Stroke incidence is decreasing in whites but not in blacks: a population-based estimate of temporal trends in stroke incidence from the Greater Cincinnati/Northern Kentucky Stroke Study. *Stroke*. 2010;41:1326–1331. [PubMed: 20489177]
26. Greater Cincinnati / Northern Kentucky 5 County Area Population-Based Epidemiology of Stroke Study. Available at: <https://www.gcnkss.com/> Accessed July 3, 2020.
27. Smith MA, Risser JM, Moyé LA, Garcia N, Akiwumi O, Uchino K, Morgenstern LB. Designing multi-ethnic stroke studies: the Brain Attack Surveillance in Corpus Christi (BASIC) project. *Ethn Dis*. 2004;14:520–526. [PubMed: 15724771]
28. REGARDS Study. Available at: <https://www.uab.edu/soph/regardsstudy/> Accessed July 3, 2020.
29. Kapral MK, Shah BR, Green ME, Porter J, Griffiths R, Frymire E, Slater M, Jacklin K, Sutherland R, Walker JD. Hospital admission for stroke or transient ischemic attack among First Nations people with diabetes in Ontario: a population-based cohort study. *CMAJ Open*. 2020;8:E156–E162.
30. Ramadan H, Patterson C, Maguire S, Melvin I, Kain K, Teale E, Forster A. Incidence of first stroke and ethnic differences in stroke pattern in Bradford, UK: Bradford Stroke Study. *Int J Stroke*. 2018;13:374–378. [PubMed: 29192873]
31. Owolabi MO, Sarfo F, Akinyemi R, Gebregziabher M, Akpa O, Akpalu A, Wahab K, Obiako R, Owolabi L, Ovbiagele B, et al. Dominant modifiable risk factors for stroke in Ghana and Nigeria (SIREN): a case-control study. *Lancet Glob Health*. 2018;6:e436–e446. [PubMed: 29496511]
32. Hyacinth HI, Carty CL, Seals SR, Irvin MR, Naik RP, Burke GL, Zakai NA, Wilson JG, Franceschini N, Winkler CA, et al. Association of Sickle Cell Trait with Ischemic Stroke Among African Americans: A Meta-analysis. *JAMA Neurol*. 2018;75:802–807. [PubMed: 29710269]
33. Levine DA, Gross AL, Briceño EM, Tilton N, Kabeto MU, Hingtgen SM, Giordani BJ, Sussman JB, Hayward RA, Burke JF, et al. Association Between Blood Pressure and Later-Life Cognition Among Black and White Individuals. *JAMA Neurol*. 2020;77:810–819. [PubMed: 32282019]
34. Aparicio HJ. Stroke Research with Longitudinal Cohort Studies. *Stroke*. 2019;50:e103–e105. [PubMed: 30841822]
35. Roger VL, Sidney S, Fairchild AL, Howard VJ, Labarthe DR, Shay CM, Tiner AC, Whitsel LP, Rosamond WD; American Heart Association Advocacy Coordinating Committee. Recommendations for Cardiovascular Health and Disease Surveillance for 2030 and Beyond: A Policy Statement From the American Heart Association. *Circulation*. 2020;141:e104–e119. [PubMed: 31992050]
36. Behavioral Risk Factor Surveillance System. Available at: <https://www.cdc.gov/brfss/index.html> Accessed July 3, 2020.
37. Get With The Guidelines® Stroke. Available at <https://www.heart.org/en/professional/quality-improvement/get-with-the-guidelines/get-with-the-guidelines-stroke> Accessed July 3, 2020.
38. Mochari-Greenberger H, Xian Y, Hellkamp AS, Schulte PJ, Bhatt DL, Fonarow GC, Saver JL, Reeves MJ, Schwamm LH, Smith EE. Racial/Ethnic and Sex Differences in Emergency Medical Services Transport Among Hospitalized US Stroke Patients: Analysis of the National Get With The Guidelines-Stroke Registry. *J Am Heart Assoc*. 2015;4:e002099. [PubMed: 26268882]
39. Merkler AE, Parikh NS, Kamel H. Jump Starting Your Clinical Research Career Using Administrative Data Sets for Stroke Research. *Stroke*. 2018;49:e303–e305. [PubMed: 30355122]
40. Gardener H, Leifheit EC, Lichtman JH, Wang Y, Wang K, Gutierrez CM, Ciliberti-Vargas MA, Dong C, Oluwole S, Robichaux M. FL-PR CReSD Investigators and Collaborators. Racial/Ethnic Disparities in Mortality Among Medicare Beneficiaries in the FL - PR CR eSD Study. *J Am Heart Assoc*. 2019;8:e009649. [PubMed: 30587062]
41. Miller EC, Gatollari HJ, Too G, Boehme AK, Leffert L, Marshall RS, Elkind MSV, Willey JZ. Risk Factors for Pregnancy-Associated Stroke in Women with Preeclampsia. *Stroke*. 2017;48:1752–1759. [PubMed: 28546324]

42. Chen ML, Parikh NS, Merkler AE, Kleindorfer DO, Bhavne PD, Levitan EB, Soliman EZ, Kamel H. Risk of Atrial Fibrillation in Black Versus White Medicare Beneficiaries With Implanted Cardiac Devices. *J Am Heart Assoc.* 2019;8:e010661. [PubMed: 30741594]
43. Allen NB, Holford TR, Bracken MB, Goldstein LB, Howard G, Wang Y, Lichtman JH. Trends in one-year recurrent ischemic stroke among the elderly in the USA: 1994–2002. *Cerebrovasc Dis.* 2010;30:525–532. [PubMed: 20881382]
44. McWilliams JM, Meara E, Zaslavsky AM, Ayanian JZ. Differences in control of cardiovascular disease and diabetes by race, ethnicity, and education: U.S. trends from 1999 to 2006 and effects of medicare coverage. *Ann Intern Med.* 2009;150:505–515. [PubMed: 19380852]
45. Navi BB, Howard G, Howard VJ, Zhao H, Judd SE, Elkind MSV, Iadecola C, DeAngelis LM, Kamel H, Okin PM, et al. New diagnosis of cancer and the risk of subsequent cerebrovascular events. *Neurology.* 2018;90:e2025–e2033. [PubMed: 29728524]
46. Henneman LRF, Choirat C, Zigler ACM. Accountability Assessment of Health Improvements in the United States Associated with Reduced Coal Emissions Between 2005 and 2012. *Epidemiology.* 2019;30:477–485. [PubMed: 31162280]
47. Wang A, Kho AN, French DD. Association of the Robert Wood Johnson Foundations' social determinants of health and Medicare hospitalisations for ischaemic strokes: a cross-sectional data analysis. *Open Heart.* 2020;7:e001189. [PubMed: 32076565]
48. Martin KD, Naert L, Goldstein LB, Kasl S, Molinaro AM, Lichtman JH. Comparing the use of diagnostic imaging and receipt of carotid endarterectomy in elderly black and white stroke patients. *J Stroke Cerebrovasc Dis.* 2012;21:600–606. [PubMed: 21411337]
49. Aron AW, Staff I, Fortunato G, McCullough LD. Prestroke living situation and depression contribute to initial stroke severity and stroke recovery. *J Stroke Cerebrovasc Dis.* 2015;24:492–499. [PubMed: 25524014]
50. Friedler B, Crapser J, McCullough L. One is the deadliest number: the detrimental effects of social isolation on cerebrovascular diseases and cognition. *Acta Neuropathol.* 2015;129:493–509. [PubMed: 25537401]
51. Verma R, Ritzel RM, Harris NM, Lee J, Kim T, Pandi G, Vemuganti R, McCullough LD. Inhibition of miR-141–3p Ameliorates the Negative Effects of Poststroke Social Isolation in Aged Mice. *Stroke.* 2018;49:1701–1707. [PubMed: 29866755]

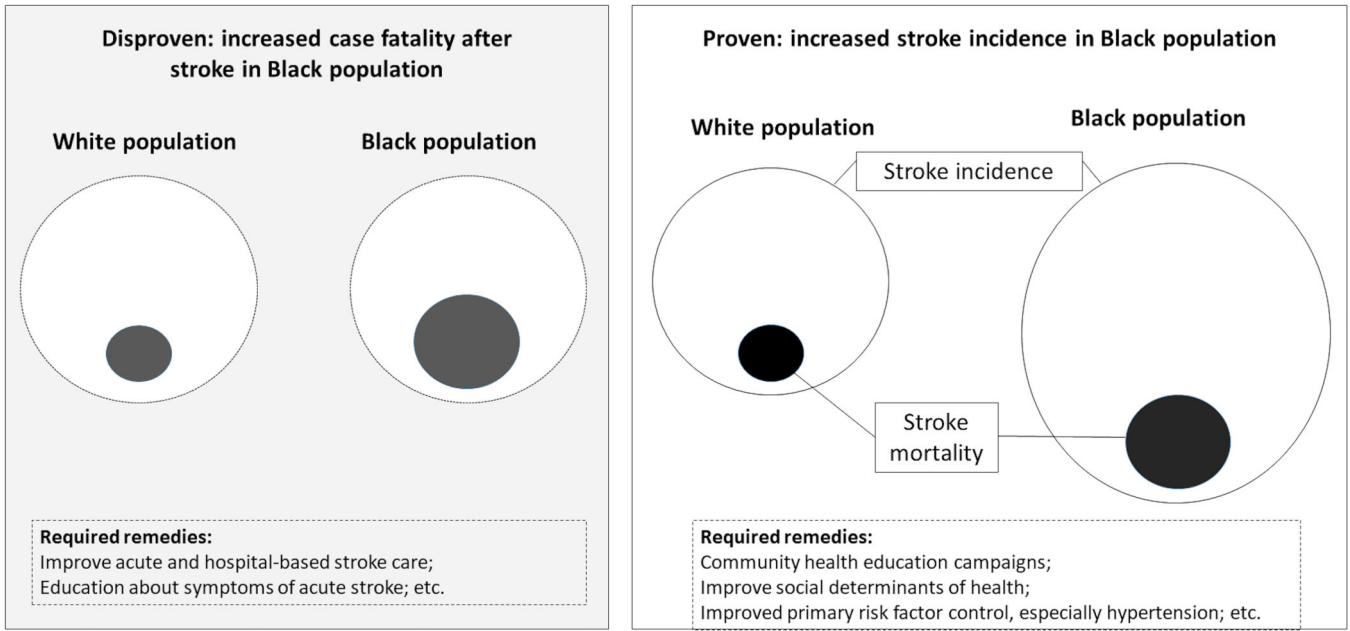


Figure: Distinguishing increased incidence from case-fatality as causes of disparities in stroke mortality

The figure illustrates the difference between increases in stroke incidence versus increases in stroke case-fatality as potential mechanisms for increased stroke mortality in Black populations compared to White populations, and implications for required remedies for mitigating the increase in mortality. The outer circles represent stroke incidence and the smaller shaded circles stroke mortality. Evidence shows (“Proven” box in figure; see text) that stroke incidence is increased in Black populations (larger outer circle on the right of the figure) with a proportional increase in mortality (the size of the inner circle). Another potential way in which increased mortality could occur (“Disproven” shaded box on the left of figure) would be to have stroke incidence remain the same for both populations (outer dashed circles), with an increase in proportional size of the mortality among the Black population (inner gray-shaded circle). Implications for required remedies to reduce disparities in mortality are provided in the dashed boxes.

Table 1. Selected cohort studies of cardiovascular risk factors and outcomes (including stroke) that may be used in disparities research

Study	Primary Funder	Recruitment period	Location	Ages	N	% White	% Black	% American Indian	% Hispanic/Latinx	% Asian
Single study design—individual race-ethnic groups										
Jackson HS	NHLBI	2000–4	Jackson, MS	20–95	5306	--	100%	--	--	--
Strong HS	NHLBI	1989–1991	Phoenix, AZ;OK, ND, SD	45–74	3516	--	--	100%	--	--
HCHS-SOL	NHLBI	2008–2011	San Diego, Chicago, Bronx, Miami	18–74	16,415	--	--	--	100% (distinct subgroups)	--
Single study design—multiple race-ethnic groups										
CHS	NHLBI	1989–1993	Multiple cities	65	5888	84%	16%	--	--	--
ARIC	NHLBI	1987–1989	Multiple cities	45–64	15,368	73%	27%	--	--	--
MESA	NHLBI	2000–2002	Multiple cities	45–84	7071	39%	27%	0%	23%	11%
NOMAS	NINDS	1993–2001	New York, NY	40	3298	25%	21%	1%	52%	1%
REGARDS	NINDS	2003–2008	National	45	30,239	52%	48%	0%	0%	0%

Abbreviations: HS=Heart Study; NHLBI=National Heart Lung and Blood Institute; HCHS-SOL=Hispanic Community Heart Study/Study of Latinos; CHS = Cardiovascular Health Study; ARIC = Atherosclerosis Risk in Communities; MESA = Multiethnic Study of Atherosclerosis; NOMAS=Northern Manhattan Study; NINDS=National Institute of Neurological Disorders and Stroke; REGARDS=Reasons for Geographic And Racial Differences in Stroke

Table 2. Comparisons among different types of epidemiological study designs with stroke outcomes

Type of analysis	Example(s)
Within a race-ethnic population	Jackson Heart Study ¹⁵ Strong Heart Study ¹⁶ Hispanic Community Heart Study/Study of Latinos (HCHS/SOL) ¹⁷
Comparison of populations in different studies	Comparison of Framingham Heart Study and others
Comparison of distinct race-ethnic communities in different locations but within same study design	Cardiovascular Health Study (CHS) ¹⁸ Atherosclerosis Risk in Communities (ARIC) ¹⁹ Multi-Ethnic Study of Atherosclerosis (MESA) ²⁰
Comparison of individuals within one location using same study design	Northern Manhattan Study (NOMAS) ²¹ Greater Cincinnati-Northern Kentucky Stroke Study (GCNKSS) ^{25,26} Brain Attack Surveillance in Corpus Christi (BASIS) ²⁷ Bradford Stroke Study ²⁵ Canadian Institute for Health Information Discharge Abstract Database ²⁴
Comparison of individuals representing different race-ethnic groups but from different locations	REasons for Geographic And Racial Differences in Stroke (REGARDS) ²² Stroke Investigative Research and Educational Network (SIREN) ²⁶
Meta-analysis/consortia	BP-Cog ²² NHLBI Pooled cohorts study Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE)

Abbreviations: NHLBI=National Heart Lung and Blood Institute

Table 3. Advantages and Disadvantages to Working with Different Clinical and Epidemiological Databases

Study Type	Scientific advantages	Practical advantages	Disadvantages
Prospective Cohort Studies	<ul style="list-style-type: none"> - Extensive data available - Granular data on stroke subtype, stroke severity, other characteristics - Availability of previously collected blood and other samples - Ability to study multiple exposures and outcomes 	<ul style="list-style-type: none"> - Senior investigators to offer mentorship and networking opportunities - Opportunity for ancillary grant proposals - Research teams across multiple clinical and scientific backgrounds (e.g., neurology, cardiology, neuropsychology, neuroimaging, epidemiology biostatistics, genetics, etc.) 	<ul style="list-style-type: none"> - Expensive to conduct - Loss to follow-up - Hard to study rare diseases - Bureaucracy - Need to learn how data collected and interpreted - May need to provide funding for analyses or data availability - Absence of mentor available in person (“telementoring”)
Patient registries	<ul style="list-style-type: none"> - Patients captured at time of clinical event - Potentially rich data - Opportunity to study processes of care - May be able to link to other data sources 	<ul style="list-style-type: none"> - Availability of databases - Existence of analytic resources 	<ul style="list-style-type: none"> - Bureaucracy - Need to learn how data collected and interpreted - May need to provide funding for analyses or data availability
Administrative data	<ul style="list-style-type: none"> - Large sample size/ excellent power - Ability to study rare diseases - Can study small subpopulations (race-ethnic groups) - Broad, representative population/high generalizability - Geographically broad; permits study of variation in practice, care, and outcomes - Temporal trends - Potential to link to other datasets 	<ul style="list-style-type: none"> - Accessible/publicly available - Relatively low cost - Deidentified data/IRB waiver - Can obtain “hypothesis-generating” data to support grants 	<ul style="list-style-type: none"> - Missing data / biases - Validity of diagnoses and other data (including race-ethnicity) - Misclassification of exposures and outcomes - Lack of granularity of data (imaging, labs, etc.) - Analyses/papers potentially held in lower regard for promotions - Data considered less scientifically robust