Can Survival Bias Explain the Age Attenuation of Racial Inequalities in Stroke Incidence?

A Simulation Study

Elizabeth Rose Mayeda,[a,b] *Hailey R. Banack,*[c] *Kirsten Bibbins-Domingo,*[b,d] *Adina Zeki Al Hazzouri,*[e] *Jessica R. Marden,*[f] *Rachel A. Whitmer,*[g,b] *and M. Maria Glymour*[b]

Background: In middle age, stroke incidence is higher among black than white Americans. For unknown reasons, this inequality decreases and reverses with age. We conducted simulations to evaluate whether selective survival could account for observed age patterning of black–white stroke inequalities.
Methods: We simulated birth cohorts of 20,000 blacks and 20,000 whites with survival distributions based on US life tables for the 1919–1921 birth cohort. We generated stroke incidence rates for ages 45–94 years using Reasons for Geographic and Racial Disparities in Stroke (REGARDS) study rates for whites and setting the effect of black race on stroke to incidence rate difference (IRD) = 20/10,000 person-years at all ages, the inequality observed at younger ages in REGARDS. We compared observed age-specific stroke incidence across scenarios, varying effects of *U*, representing unobserved factors influencing mortality and stroke risk.
Results: Despite a constant adverse effect of black race on stroke risk, the observed black–white inequality in stroke incidence attenuated at older age. When the hazard ratio for *U* on stroke was 1.5 for both blacks and whites, but *U* only directly influenced mortality for blacks (hazard ratio for *U* on mortality =1.5 for blacks; 1.0 for whites), stroke incidence rates in late life were lower among blacks (average observed IRD = −43/10,000 person-years at ages 85–94 years versus causal IRD = 20/10,000 person-years) and mirrored patterns observed in REGARDS.
Conclusions: A relatively moderate unmeasured common cause of stroke and survival could fully account for observed age attenuation of racial inequalities in stroke.

Keywords: Survival bias; Selection bias; Racial disparities; Stroke; Lifecourse Epidemiology; Simulation

(*Epidemiology* 2018;29: 525–532)

Stroke is a leading cause of death and disability in the United States, and stroke incidence and prevalence is higher among black Americans than among other racial/ethnic groups in the United States.[1] The magnitude of the inequality in stroke incidence is large in midlife but decreases with age and is eliminated in the oldest age bands.[2–5] For example, in the national Reasons for Geographic and Racial Differences in Stroke (REGARDS) cohort, stroke incidence was four times higher among blacks than whites 45–54 years of age, but the incidence rate ratio gradually attenuated at older ages, and by ≥85 years of age, incidence rates were lower for blacks.[2]

From a public health perspective, it is important to understand whether the observed age attenuation of racial inequalities in stroke incidence represents elimination of the social disadvantage that leads to higher stroke incidence among black versus white Americans or whether the observed trend may instead be an artifact of selective survival. In the present situation, it is especially pertinent to evaluate the role of survivor bias due to collider stratification, an association between two variables arising from conditioning on a "collider"—a factor influenced by both variables.[6,7] In the context of selective survival, collider-stratification bias potentially arises because analyses are restricted to living

Submitted March 7, 2017; accepted March 28, 2018.
From the [a]Department of Epidemiology, University of California, Los Angeles Fielding School of Public Health, Los Angeles, CA; [b]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA; [c]Department of Epidemiology and Environmental Health, State University of New York at Buffalo, Buffalo, NY; [d]Department of Medicine, University of California, San Francisco, San Francisco CA; [e]Division of Epidemiology and Population Health, Department of Public Health Sciences, Miller School of Medicine, University of Miami, Miami, MI; [f]Department of Epidemiology, Harvard School of Public Health, Boston, MA; and [g]Kaiser Permanente Division of Research, Oakland, CA.
This work was supported by grants U54NS081760 from the National Institute of Neurological Disorders and Stroke, 15POST25090083 from the American Heart Association, K99AG053410, RF1AG052132, RF1AG050782, and K01AG047273 from the National Institute on Aging, K24DK103992 from the National Institute of Diabetes and Digestive and Kidney Diseases, T32 MH017119 from the National Institute on Mental Health, and a Banting Postdoctoral Fellowship from Canadian Institute for Health Research.
Description of the process by which someone else could obtain the data and computing code required to replicate the results reported: Computing code for generating and analyzing simulation data are available online: https://github.com/ermayeda/stroke_inequalities_simulation.
The authors report no conflicts of interest.
SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).
Correspondence: Elizabeth Rose Mayeda, Department of Epidemiology, University of California, Los Angeles Fielding School of Public Health, BOX 951772, 46-070B CHS, Los Angeles, CA 90095. E-mail: ermayeda@ph.ucla.edu

ISSN: 1044-3983/18/2904-0525
DOI: 10.1097/EDE.0000000000000834

individuals, and any two characteristics that influence survival may become associated among survivors. In the demography literature, this phenomenon is attributed to unobserved population heterogeneity or frailty and routinely invoked as a plausible explanation for mortality crossovers: in populations with unobserved heterogeneity in vulnerability to death, the more robust subgroup will come to predominate among survivors.[8,9] Because black Americans have higher mortality than white Americans through early and middle life,[10,11] black Americans who survive stroke-free to old age could represent a more selected, healthier population than their white counterparts. Consequently, stroke incidence rates may be lower among older black than white Americans due to this selection process. Although this explanation is possible, no previous work demonstrates that it could plausibly account for the observed age attenuation of racial differences in stroke incidence, and the extent to which collider stratification induces substantial bias is controversial.[12–19]

Another methodologic factor that could contribute to the age attenuation of racial inequalities in stroke relates to the scale (additive or multiplicative) on which racial inequalities in stroke incidence are reported. Racial inequalities in stroke are often expressed in relative terms (e.g., incidence rate ratios).[2–5] Mathematically, relative measures of inequality can decrease even when absolute measures remain constant if the rate in the reference group increases.[20–22] Because stroke incidence rates increase drastically with age among whites, an age-constant *relative* effect of race on stroke risk implies substantially higher stroke incidence rates in older blacks than would an age-constant *additive* effect of race on stroke risk.[20–22]

The objective of this study was to assess the extent to which selective survival could explain the age attenuation of racial inequalities in stroke incidence. Using stroke incidence rates from REGARDS, survival rates from US life tables, and considering associations between race and stroke incidence on both additive and multiplicative scales, we conducted simulations to evaluate the potential impact of selective survival on observed black–white inequalities in stroke incidence by age under a range of possible underlying causal scenarios.

METHODS

Hypothetical Cohort Study of Racial Inequalities in Stroke Incidence

We considered a hypothetical (simulated) cohort study of black–white inequalities in stroke incidence. Elevated mortality among black versus white Americans is present from birth.[11] We begin our hypothetical cohort population with n = 20,000 blacks and n = 20,000 whites at birth and follow the hypothetical cohort population for first incident stroke from 45–94 years of age. We selected this sample size for the birth cohort so that there would be sufficient numbers of survivors of each race through 94 years of age, while being computationally feasible. To quantify the extent to which selective survival could plausibly explain the age attenuation of racial inequalities in stroke incidence, we generated the data assuming the effect of race on stroke incidence is age constant. Thus, associations between race and stroke incidence that deviate from this prespecified effect of race reflect survival bias in our simulations. Throughout the paper, we use the terms "inequality" and "difference" to refer to statistical patterns showing different rates of stroke between blacks and whites, which would be directly observable in real data, and we use "disparity" to invoke the assumption that these statistical patterns are attributable to unjust mechanisms, which may not be observable.

Simulation Study Outline

Our simulation study procedures can be outlined as follows. (1) Select causal scenarios for investigation; our causal scenarios are described by Figure 1, which are detailed later in the methods; (2) specify data-generating process for hypothetical cohort populations corresponding with each causal scenario. Since we are generating the data, we prespecify the "true" age-constant effect of race on stroke incidence; (3) run 2000 iterations of sample generation under each causal scenario and estimate the racial difference in stroke incidence in each age band in each sample; (4) quantify the magnitude of bias in each scenario by comparing the observed racial difference in stroke in each age band averaged across the 2000 samples with the known "true" effect of race on stroke risk in our simulations. By comparing observed associations with the "true" effect in our simulations, we are able to quantify the extent to which selective survival contributes to the age attenuation of racial differences in stroke incidence in each causal scenario. This simulation study process is outlined graphically in eAppendix 1; http://links.lww.com/EDE/B339 and is described below.

Data-Generating Process

We simulated data for our hypothetical cohort under several causal scenarios. In all scenarios, survival rates were calibrated to match US life tables for race-specific survival for the 1919–1921 birth cohort,[23] and age-specific stroke incidence rates were set to match incidence rates in REGARDS for whites.[2] We selected this birth cohort for survival data because it corresponds with the birth cohort for which the black–white crossover of stroke incidence was observed in REGARDS (i.e., people who would have been in their mid- to late-80s in the mid-2000s). REGARDS investigators provided race- and age-specific stroke incidence rates that are updates of previously published results[2] (personal communication with Dr. George Howard, December 2016). We define the "true" racial disparity in stroke incidence in our simulations as the causal effect of race on stroke incidence,[24,25] which we assume to be age constant in our simulations. We set the effect of black race on stroke incidence to 20 excess strokes per 10,000 person-years in all age bands (i.e., stroke incidence rate$_{black}$ = stroke

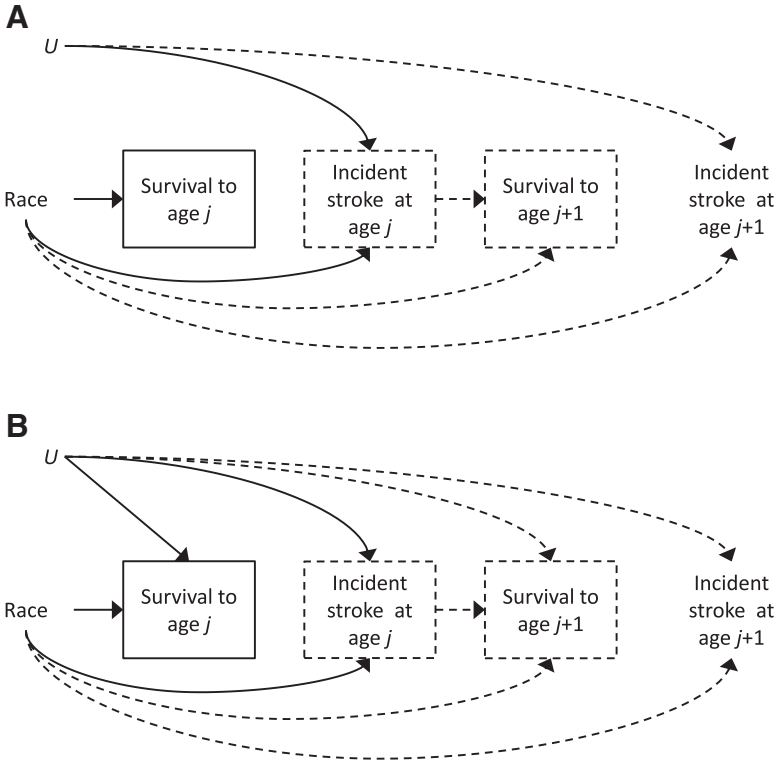

FIGURE 1. Causal scenarios under investigation. In all scenarios, race influences mortality (following US life table estimates from birth) and stroke risk (assuming a constant effect of black race on stroke risk ages 45–94 years). *U* represents a factor, or set of factors, that influences risk of stroke and mortality and is independent of race at birth. History of stroke influences mortality risk, and only people who are alive with no history of stroke are at risk for first incident stroke. Solid lines represent the data-generating process at age *j*, and dashed lines represent the data-generating process at age *j* + 1. The box around "Survival to age *j*" represents that only people alive at age *j* are at risk for stroke at age *j*. The dashed boxes around "Incident stroke at age *j*" and "Survival to age *j* + 1" indicate that any analyses of stroke incidence at age *j* + 1 are conditional on these variables: only people who are alive without a history of stroke at age *j* + 1 are at risk for first incident stroke at age *j* + 1. (A) Scenario A, where *U* influences stroke but has no direct effect on mortality risk. (B) Scenarios B and C. In Scenario B, *U* influences stroke risk and directly affects mortality risk. In Scenario C, *U* influences stroke risk for both blacks and whites; *U* only directly affects mortality risk for blacks.

incidence rate$_{white}$ + 20/10,000 person-years). We used the incidence rate difference (IRD) instead of a ratio measure for the effect of race on stroke risk because in analyses of the REGARDS results, although neither additive nor multiplicative models fit well, an additive model fit the observed data better (see eAppendix 2; http://links.lww.com/EDE/B339 for calculations). In REGARDS, the stroke incidence rates were approximately 20/10,000 person-years higher among blacks than whites in the youngest age band (45–54 years), and the inequality attenuated with age. See eAppendix 3; http://links.lww.com/EDE/B339 for additional details of the data-generating process.

Causal Scenarios

In all scenarios (illustrated in Figure 1), race influences survival and stroke. While we construct race as a single binary variable, we conceptualize race as a social construct that represents many factors that contribute to disparities in survival and stroke risk.[26–28] To represent the selection process that could potentially give rise to lower stroke incidence rates among black compared with white Americans who survive stroke-free to old age, we generated a variable, *U*, as a normally distributed continuous variable with mean = 0 and standard deviation = 1.0 that represents a time-invariant factor, or set of factors, that influences both survival and stroke risk and is independent of race at birth. We selected this specification for *U* so that a one-unit difference in *U* would be easily interpretable; for examples of *U* that are strictly positive, it can be conceptualized as a mean-centered version of the variable. Numerous factors are plausible candidates for *U*, including genetic variants related to vascular disease, personality differences that shape behavioral patterns, or chance exposures to environmental risk factors. In all scenarios, history of stroke increases mortality risk, and only people who are still alive with no history of stroke are at risk for incident stroke.

To ensure that observed bias arises from collider-stratification bias, we begin with a base scenario with no anticipated survivor bias (Scenario A). In this scenario, *U* directly influences stroke risk but has no direct effect on mortality risk (hazard ratio [HR] for one-unit higher *U* on stroke =1.5). Scenarios B and C are causal structures consistent with collider-stratification bias.[6,7] In Scenario B, *U* directly influences stroke risk and mortality risk for both blacks and whites (HR for one-unit higher *U* on stroke =1.5; HR for one-unit higher *U* on mortality =1.5). In Scenario C, *U* directly influences both stroke risk and mortality risk for blacks (HR for one-unit higher *U* on stroke =1.5; HR for one-unit higher *U* on mortality =1.5) but has no direct effect on mortality for whites (HR for one-unit higher *U* on stroke =1.5; HR for one-unit higher *U* on mortality =1.0).

As a supplemental analysis, we investigated an extension of causal scenarios investigated in the primary simulation study, where race influences *U*, that is, a data-generating model where *U* is a partial mediator of the effect of race on stroke risk (in all scenarios) and mortality risk (in scenarios consistent with collider-stratification bias). The methods for this extended simulation study are described in eAppendix 4; http://links.lww.com/EDE/B339.

Quantifying the Magnitude of Survival Bias

As previously described, the causal effect of black race on stroke incidence was set to IRD = 20/10,000 person-years, so estimates for the difference in stroke incidence rates between blacks and whites at any age in our study that deviate from IRD = 20/10,000 person-years reflect bias. For each causal scenario, we estimated the black–white stroke IRD among surviving stroke-free cohort members for 10-year age bands between 45 and 94 years of age, the age bands for which stroke incidence was reported in REGARDS.

Across the 2000 simulated samples, we calculated the average observed black–white stroke IRD (and incidence rate ratio, IRR) for each age band and the percentage bias, calculated as $100 \times (\text{IRD}_{\text{observed}} - \text{IRD}_{\text{causal}}) / \text{IRD}_{\text{causal}}$. We were primarily interested in the bias, but to convey that the deviation from truth will vary in finite samples, we assessed variability in the observed IRD across replications as the empirical standard error, which is the standard deviation of the observed IRD across the 2000 simulated samples. In our simulations, empirical standard errors are larger for older ages due to the smaller sample size. We assessed accuracy as root mean square error, which is the square root of the mean squared deviation of the observed IRD from the causal IRD. See the appendix; http://links.lww.com/EDE/B339 for further explanation on the IRR comparisons.

In Scenarios B and C, survival bias is expected to operate through *U*. Even though *U* is independent of race at birth, because black race and *U* increase mortality and stroke risk, the population of black Americans surviving stroke-free to old age will have more protective (lower) values of *U* than the population of white Americans surviving stroke-free to old age (i.e., an association between race an *U* will be induced among the stroke-free older adult population). This is the process that drives the age attenuation of racial differences in stroke incidence when there is an age-constant effect of race on stroke risk. We examine the race-*U* association by age by plotting the difference in mean *U* between blacks and whites surviving stroke-free to ages 45, 50, 55, …, 85, 90 years.

The simulation code, in Stata SE version 13.1 (StataCorp LP, College Station, Texas), is available online (https://github.com/ermayeda/stroke_inequalities_simulation).

RESULTS

Survival distributions in our study population were consistent with US life tables for the 1919–1921 birth cohort[23] (eTable 2; http://links.lww.com/EDE/B339, eFigure 2; http://links.lww.com/EDE/B339). Median survival was approximately 50 years for blacks and 65 years for whites; approximately 56% of blacks and 73% of whites survived from birth to 45 years of age, the age at which our hypothetical study of stroke incidence began.

In REGARDS, the observed black–white stroke IRD was approximately 20/10,000 person-years at 45–54 years of age and attenuated with age; at ≥75 years of age stroke incidence rates were lower in blacks than in whites (Figure 2A, Table). In Scenario A (*U* increases stroke risk but has no direct effect on mortality risk), there was minimal bias: the observed black–white stroke IRD was approximately 20/10,000 person-years across all age bands, matching the causal effect of race on stroke incidence specified in the simulation. In Scenarios B and C, however, selective survival resulted in age attenuation in the observed black–white stroke IRD. In Scenario B (*U* increases stroke risk and mortality risk), the observed black–white stroke IRD attenuated with age, although the average observed stroke incidence rate was higher in blacks than in whites in all age bands (average observed IRD for ages 85–94 years = 9.6/10,000 person-years). The magnitude of bias was significantly greater in Scenario C (*U* increases stroke risk for both blacks and whites but only directly affects mortality for blacks): the age attenuation of the observed black–white IRD was similar to REGARDS estimates, with the observed IRD less than 0 at ≥75 years of age (average observed IRD for ages 85–94 = −43.0/10,000 person-years). The observed black–white stroke IRD and 95% confidence intervals for 85–94 years of age for Scenario C across each of the 2000 simulated samples is displayed in eFigure 3; http://links.lww.com/EDE/B339.

Results followed similar patterns for the observed black–white stroke IRR (Figure 2B, eTable 3; http://links.lww.com/EDE/B339). In REGARDS, the observed black–white stroke IRR was approximately 3.4 at 45–54 years of age and attenuated with age, and the observed stroke IRR was less than 1.0 for ≥75 years of age. In our simulations, because we set the effect of black race on stroke incidence to be age constant on the IRD in our simulations, the "true" black–white stroke IRR attenuates with age. In simulation Scenario A, there was essentially no bias; the average observed IRR was 3.51 for 45–54 years of age (true IRR = 3.39) and 1.15 for 85–94 years of age (true IRR = 1.13). In Scenario B, the average observed IRR attenuated with age but remained greater than 1.0 across all age bands. In Scenario C, the average observed IRR also attenuated with age and was less than 1.0 for ≥75 years of age. The age attenuation of the black–white stroke IRR in Scenario C was similar to the results from the REGARDS study.

As described in the methods, the age attenuation of racial inequalities in stroke incidence in Scenarios B and C is driven by the presence of *U*, a common cause of mortality and stroke risk which is independent of race at birth. However, since race influences stroke risk and mortality risk, black Americans who survive to old age tend to have more protective values of *U* than white Americans who survive to old age. The consequence of this selection, the association between race and *U*, is illustrated by Figure 3. In Scenario A, *U* influences stroke risk but has no direct effect on mortality risk. As a result, the mean of *U* was similar in blacks and whites across age bands. In Scenarios B and C, *U* influences both stroke risk and mortality risk. As a result, a black–white difference in mean *U* was present by 45 years of age in both scenarios. In

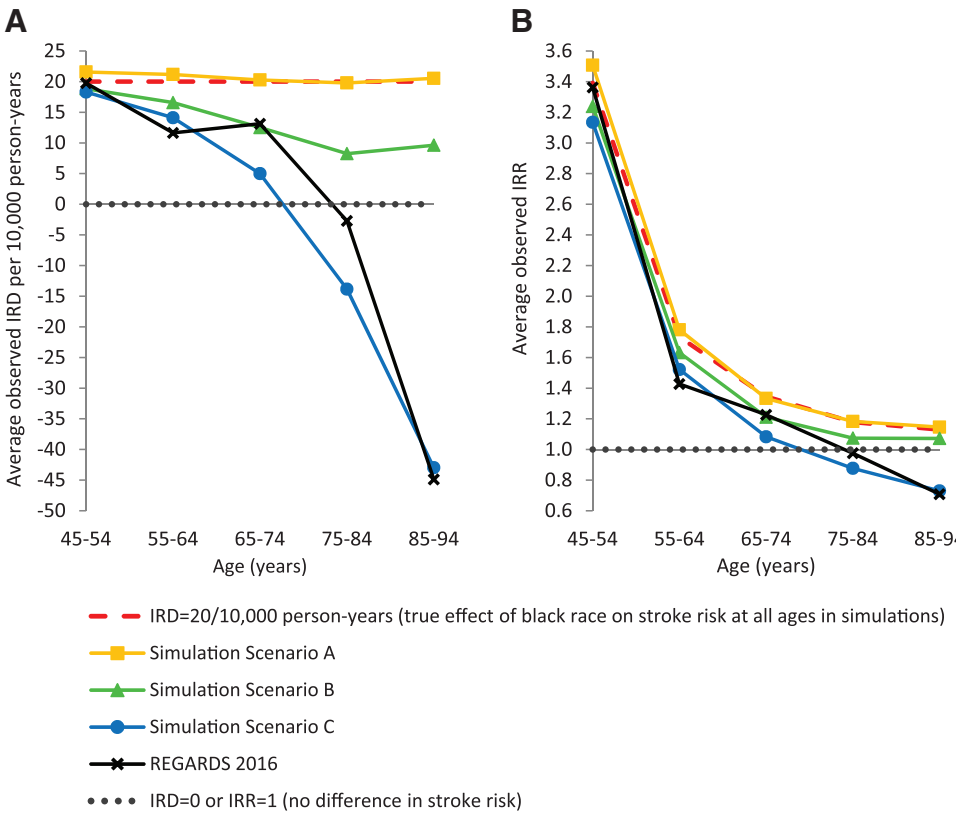

FIGURE 2. Average observed black–white stroke incidence rate difference (IRD) (A) and stroke incidence rate ratio (IRR) (B) by age band, calculated by taking the average of the observed IRD or IRR across 2000 simulated samples per scenario. For comparison, results observed in the Reasons for Geographic and Racial Disparities in Stroke (REGARDS) cohort are also displayed. The dashed line shows that the true effect of black race on stroke rates at all ages is IRD = 20/10,000 person-years in our simulation study; the dotted line shows IRD = 0 or IRR = 1 (no difference in stroke rate). Note that in our simulations, because we set the effect of black race on stroke incidence to be age constant on the IRD scale, the black–white stroke IRR consistent with the true IRD = 20/10,000 person-years attenuates with age. Figure is available in color online.

Scenario B, the difference in *U* grew until 75–80 years of age, the age band where mortality rates in whites surpass mortality rates in blacks. In Scenario C, this difference in *U* grew monotonically with age. As a result of the difference in the distribution of *U* among blacks and whites who are still alive and at risk for stroke in the oldest age bands in Scenarios B and C, the observed difference in black–white stroke inequality attenuated with age, even though the causal effect of race on stroke risk persisted with age.

In the supplemental simulation study in which *U* mediates effects of race, we found that the causal structure where race and *U* interact to influence mortality reproduced the age attenuation of racial inequalities in stroke consistent with REGARDS, similar to our primary simulations where race and *U* were independent (Appendix 6; http://links.lww.com/EDE/B339).

DISCUSSION

Using simulations with parameter inputs guided by real data for race-specific survival and stroke incidence, we found that a relatively moderate unmeasured common cause of stroke and survival could fully account for observed age attenuation of racial inequalities in stroke. Even in the context of persistent racial disparities—that is, a true adverse effect of black race on stroke risk that persists through old age—incidence rates in the simulations were lower among blacks than among whites in the highly selected sample of stroke-free survivors after 75 years of age.

Although selective survival has been posited as a potential explanation for the observed age attenuation in racial inequalities in stroke,[29] to our knowledge, this is the first study to evaluate the plausibility of this explanation. We focus on results from REGARDS because it is a large, national, and up-to-date source of data on stroke incidence in black and white Americans,[2] but age attenuation of black–white inequalities in stroke incidence has been reported in other settings, such as the Greater Cincinnati/Northern Kentucky Stroke Study[3,4] and the National Health and Nutrition Examination Survey Epidemiologic Follow-up Study.[5]

In our simulations, the scenario that reproduced the age patterning of racial inequalities in stroke incidence in REGARDS was the scenario where there was a multiplicative

TABLE. Black–White Stroke Incidence Rate Difference (IRD) Results From the REGARDS Cohort and From Simulation Scenarios (2000 Simulated Samples)

	Age 45–54 years			Age 55–64 years			Age 65–74 years			Age 75–84 years			Age 85–94 years		
	IRD[a] (SE[b])	% Bias[c]	RMSE[d]	IRD[a] (SE[b])	% Bias[c]	RMSE[d]	IRD[a] (SE[b])	% Bias[c]	RMSE[d]	IRD[a] (SE[b])	% Bias[c]	RMSE[d]	IRD[a] (SE[b])	% Bias[c]	RMSE[d]
REGARDS[e]	19.8 (5.4)	—	—	11.6 (3.9)	—	—	13.1 (6.2)	—	—	−2.7 (13.1)	—	—	−44.9 (42.7)	—	—
Truth in simulations	20.0	—	—	20.0	—	—	20.0	—	—	20.0	—	—	20.0	—	—
Scenario A[f]	21.6 (1.9)	8%	2.5	21.2 (3.1)	6%	3.3	20.3 (5.1)	1%	5.1	19.8 (10.1)	−1%	10.1	20.5 (25.5)	3%	25.5
Scenario B[g]	18.8 (1.8)	−6%	2.1	16.6 (2.8)	−17%	4.4	12.5 (4.8)	−37%	8.9	8.3 (9.3)	−59%	15.0	9.6 (24.4)	−52%	26.5
Scenario C[h]	18.3 (1.9)	−9%	2.5	14.1 (2.8)	−29%	6.5	5.0 (4.6)	−75%	15.7	−13.8 (8.9)	−169%	35.0	−43.0 (22.7)	−315%	66.9

See eTable 4; http://links.lww.com/EDE/B339 for average stroke incidence rate estimates for blacks and whites in each scenario.

[a]IRD = stroke incidence rate difference per 10,000 person-years for blacks versus whites in each age band. For REGARDS, the IRD is the observed age-specific IRD in REGARDS. For truth in simulations, the causal IRD = 20/10,000 person-years. For simulation scenario results, the IRD is the average of the observed age-specific IRD across 2000 simulated samples for each simulation scenario.

[b]For REGARDS, the SE is the SE for REGARDS age-specific IRD estimates. For simulations, the SE is the empirical SE, calculated as the standard deviation of observed age-specific IRDs across the 2000 simulated samples for each simulation scenario. The SEs (for REGARDS) and empirical SEs (simulations) increase across age bands because the number of people remaining alive and at risk for stroke declines with age (see proportion of birth cohort alive at each age in eTable2; http://links.lww.com/EDE/B339).

[c]Percentage bias in the simulation estimates is defined as deviation from the truth, $100 \times (\overline{\text{IRD}}_{\text{observed}} - \text{IRD}_{\text{causal}}) / \text{IRD}_{\text{causal}}$

[d]RMSE for simulations is defined as the square root of the mean squared deviation of the observed IRD from the causal IRD. The RMSE is influenced by both the bias and the variance of the observed IRD. Thus, the decrease in sample size across age bands contributes to the increase in RMSE across age bands.

[e]Personal communication with Dr. George Howard, PI of REGARDS Study, December 2016.

[f]Scenario A: *U* increases stroke risk but has no direct effect on mortality risk.

[g]Scenario B: *U* increases stroke risk and mortality risk.

[h]Scenario C: *U* increases stroke risk for both blacks and whites but only directly affects mortality for blacks.

IRD indicates incidence rate difference; REGARDS, Reasons for Geographic and Racial Disparities in Stroke; RMSE, root mean square error; SE, standard error.

interaction between race and *U* on mortality. For simplicity, we expressed this interaction such that *U* influenced mortality for blacks but had no direct effect on mortality for whites. However, we expect that the magnitude of bias would be similar or larger if *U* directly affected mortality for both blacks and whites but interacted with race such that *U* had larger direct effects on mortality for blacks. More empirical research is needed to examine the likely magnitude of the interaction between race and possible *U* variables on mortality.

Recent work has evaluated the likely role of collider-stratification bias in phenomena such as the obesity paradox[14,16,17,30–33] and the potential magnitude of bias arising from selective study participation/attrition.[12,18,34,35] In the present simulations, a relatively modest multiplicative interaction between race and *U* on mortality resulted in an age attenuation of the race–stroke association that mirrors the observed age attenuation of racial inequalities in stroke incidence in REGARDS. Our finding that the magnitude of bias in our simulations is greatest in the scenario where there was a multiplicative interaction between race and *U* on mortality is consistent with prior work demonstrating that the magnitude of selection bias is larger when there is a multiplicative interaction between the exposure and the common cause of selection and the outcome.[12,14,17,19,36,37]

Our decision to use an additive scale for the effect of race on stroke risk in our simulations was likely important. Racial inequalities are often reported in terms of ratios, despite frequent calls to focus on differences or absolute effect estimates.[20–22] An additive model fit REGARDS data on racial inequalities in stroke better than a multiplicative model, but specifying an additive model for stroke incidence is unusual.[38,39] The estimated effect of risk factors on stroke is commonly modeled on a multiplicative scale (e.g., using Cox regression models),[1] and while we modeled the effect of race on stroke risk as an additive effect, we modeled the effect of *U* on stroke risk as a multiplicative effect. Because stroke incidence increases dramatically with age, the age-constant additive effect of race on stroke specified in our simulations represented a fairly small relative effect in older age groups, in contrast to the effect of *U* on stroke, which was age constant on the relative scale. Understanding the most plausible data-generating structures and parameterizations of variables is critical for evaluating the credibility of simulation studies.

The assumption of whether race has a constant additive effect or a constant multiplicative effect is particularly critical when modeling *U* as a partial mediator of the effect of race on stroke. Given substantial racial stratification in socioeconomic conditions, encounters with discrimination, environmental exposures, and health care access in the United States, there are probably many common causes of stroke and mortality that are influenced by race. For simplicity, our primary simulations modeled *U* as independent of race at birth. In supplemental simulations, we considered causal structures where *U* partially mediates the effect of race on stroke risk and mortality risk and found patterns similar to our primary results, provided there was interaction on a multiplicative scale between *U* and race in determining mortality. To allow *U* to partially mediate the effect of race on stroke, we generated the effect of *U* on stroke as additive in the supplemental simulations. To maintain *U*'s additive effect on stroke entails, however, that in older ages, race accounts for only a small fraction of the variance in *U*.

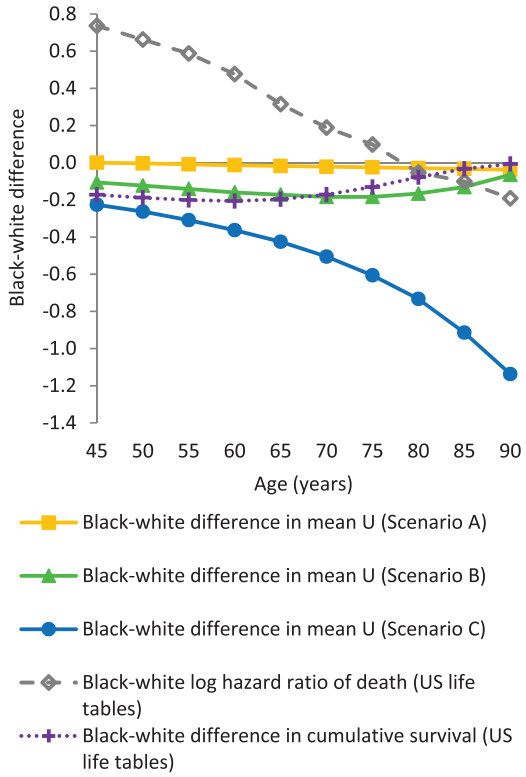

FIGURE 3. Black–white difference in mean *U* in population at risk for incident stroke at each age band for each simulation scenario across 2000 simulated samples, black–white log hazard ratio of death (from US life tables), and black–white difference in cumulative survival (from US life tables). We generated *U* in the hypothetical population at birth as a normally distributed continuous variable with mean = 0 and standard deviation 1.0 that was independent of race. In Scenario A, *U* influenced stroke risk but had no direct effect on mortality. Thus, there was approximately no mean difference in *U* between blacks and whites in Scenario A. In Scenarios B and C, *U* increased stroke risk and directly influenced mortality. Thus, a black–white difference in *U* was present by 45 years of age. In Scenario B, the black–white difference in *U* increased with age until it began to narrow at older ages (around 75–80 years of age)—the approximate age at which mortality rates in whites surpass mortality rates in blacks based on US life tables. In Scenario C, the black–white difference in *U* increased monotonically with age. Figure is available in color online.

Age attenuation of risk factor associations are observed for many exposures and health outcomes.[40,41] Explanations for such age attenuation are equivocal in many cases, and true age heterogeneity in effects may co-exist with biases such as selective survival or reverse causality. For example, midlife obesity predicts higher mortality risk, whereas the association is weaker among older adults.[42,43] Confounding due to chronic conditions that cause weight loss likely contributes to the age attenuation of the obesity–mortality association,[43] but there is little evidence on whether other sources of bias contribute.

Although the present study demonstrates that selective survival is a plausible explanation for the age attenuation of racial inequalities in stroke incidence, we cannot rule out other potential explanations, which include both causal (i.e., the effect of black race on stroke incidence attenuates with age) and noncausal (i.e., sources of bias other than survival bias) phenomena. A potential causal explanation is that at 65 years of age, access to government-sponsored social and healthcare services such as Medicare vastly expands, which could play a role in reducing health disparities.[29] A causal structure that could potentially be a source of survival bias, distinct from the causal structures we investigated, is that due to heterogeneity in biological sensitivity, that is, black Americans who survive to old age represent a subset of the population whose health is not sensitive to the harmful effects of racially patterned adversity.[44] This would be consistent with selection bias in the absence of collider stratification.[35,37] Another potential noncausal explanation is that the observed age patterning of racial differences in survival and stroke risk reflect cohort, rather than age effects.[45]

Eliminating racial disparities in health is a national public health priority[46]; the goal of this paper was to help understand age patterns in racial inequalities in stroke incidence. The findings of this article are not dependent on the specific mechanisms linking race and stroke, although identifying such mechanisms would be invaluable for achieving health equity across the lifecourse. An important result of the present study is that there may be a racial disparity among older adults–that is, the underlying social disadvantage that contributed to higher stroke incidence among blacks than whites earlier in the lifecourse may persist—even if there is no observed racial *difference* in stroke incidence, or even *lower* stroke incidence among older blacks compared with older whites. In our simulations, blacks who survived stroke-free to older ages were nonetheless disadvantaged with respect to stroke risk due to their race. In other words, these hypothetical people who survived to old age would have had *even lower* stroke risk had they been white.

Simulation studies depend on the particular causal structures and input parameters selected by investigators. Our data-generating process was guided by real data for mortality and stroke incidence. We were able to show that with relatively modest effects of a common cause of stroke risk and mortality risk (*U*), observed stroke rates were lower among blacks than whites who survived to old age stroke-free. For input parameters relevant to other outcomes, the magnitude of bias would likely vary. An important consideration is that our data-generating model includes a built-in selection bias due to examining racial inequalities in stroke at age $j + 1$ conditional on stroke at age j.[47] Although this did not introduce considerable bias in our hypothetical simulation study (demonstrated by scenario with no anticipated bias), the magnitude of bias could be substantial for outcomes with higher cumulative incidence. Thus, it is important for not only the causal structures, but also the parameterization of the causal structures, to be relevant for the specific exposure–outcome association of interest.

Using simulations guided by real data for mortality and stroke incidence, we found that the observed age attenuation of racial inequalities in stroke incidence could be easily explained by selective survival, even if black race has a constant adverse effect on stroke risk through late life.

ACKNOWLEDGMENTS

We thank Dr. George Howard (University of Alabama at Birmingham) for providing updated stroke incidence rates from the REGARDS Study in December 2016.

REFERENCES

1. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, Das SR, de Ferranti S, Després J-P, Fullerton HJ. Heart disease and stroke statistics-2016 update: A report from the American Heart Association. *Circulation*. 2016;133:e38–e360.
2. Howard VJ, Kleindorfer DO, Judd SE, et al. Disparities in stroke incidence contributing to disparities in stroke mortality. *Ann Neurol*. 2011;69:619–627.
3. Kissela B, Schneider A, Kleindorfer D, et al. Stroke in a biracial population: the excess burden of stroke among blacks. *Stroke*. 2004;35:426–431.
4. Kleindorfer D, Broderick J, Khoury J, et al. The unchanging incidence and case-fatality of stroke in the 1990s: a population-based study. *Stroke*. 2006;37:2473–2478.
5. Giles WH, Kittner SJ, Hebel JR, Losonczy KG, Sherwin RW. Determinants of black-white differences in the risk of cerebral infarction. The National Health and Nutrition Examination Survey Epidemiologic Follow-up Study. *Arch Intern Med*. 1995;155:1319–1324.
6. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15:615–625.
7. Hernán MA, Robins JM. Selection bias. *Causal Inference* Boca Raton: Chapman & Hall/CRC; 2018.
8. Vaupel JW, Yashin AI. Heterogeneity's ruses: some surprising effects of selection on population dynamics. *Am Stat*. 1985;39:176–185.
9. Eberstein IW, Nam CB, Heyman KM. Causes of death and mortality crossovers by race. *Biodemography Soc Biol*. 2008;54:214–228.
10. Hummer RA, Rogers RG, Masters RK, Saint Onge JM. Mortality patterns in late life. In Uhlenberg P. (eds) *International Handbook of Population Aging* Springer, Dordrecht; 2009:521–542.
11. Jackson JS, Hudson D, Kershaw K, Mezuk B, Rafferty J, Tuttle KK. Discrimination, chronic stress, and mortality among Black Americans: A life course framework. In Rogers R., Crimmins E. (eds) *International Handbook of Adult Mortality*. Springer, Dordrecht; 2011:311–328.
12. Mayeda ER, Tchetgen Tchetgen EJ, Power MC, et al. A simulation platform for quantifying survival bias: an application to research on determinants of cognitive decline. *Am J Epidemiol*. 2016;184:378–387.
13. Liu W, Brookhart MA, Schneeweiss S, Mi X, Setoguchi S. Implications of M bias in epidemiologic studies: a simulation study. *Am J Epidemiol*. 2012;176:938–948.
14. Glymour MM, Vittinghoff E. Commentary: selection bias as an explanation for the obesity paradox: just because it's possible doesn't mean it's plausible. *Epidemiology*. 2014;25:4–6.
15. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003;14:300–306.
16. Sperrin M, Candlish J, Badrick E, Renehan A, Buchan I. Collider bias is only a partial explanation for the obesity paradox. *Epidemiology*. 2016;27:525–530.
17. Viallon V, Dufournet M. Re: collider bias is only a partial explanation for the obesity paradox. *Epidemiology*. 2017;28:e43–e45.
18. Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. *Int J Epidemiol*. 2018;47:226–235.
19. Stensrud MJ, Valberg M, Røysland K, Aalen OO. Exploring selection bias by causal frailty models: the magnitude matters. *Epidemiology*. 2017;28:379–386.
20. King NB, Harper S, Young ME. Use of relative and absolute effect measures in reporting health inequalities: structured review. *BMJ*. 2012;345:e5774.
21. Poole C. Some thoughts on consequential epidemiology and causal architecture. *Epidemiology*. 2016;28:6–11.
22. Harper S, Lynch J. Health Inequalities: Measurement and Decomposition. In: Oakes J, Kaufman J, eds. *Methods in Social Epidemiology, 2nd ed*. San Francisco: Jossey-Bass; 2016.
23. Arias E. United States life tables 2006. *Natl Vital Stat Rep*. 2010;58:1–40.
24. VanderWeele TJ, Robinson WR. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology*. 2014;25:473–484.
25. Glymour MM, Spiegelman D. Evaluating public health interventions: 5. Causal inference in public health research-do sex, race, and biological factors cause health outcomes? *Am J Public Health*. 2017;107:81–85.
26. Williams DR. Race and health: basic questions, emerging directions. *Ann Epidemiol*. 1997;7:322–333.
27. Krieger N. The science and epidemiology of racism and health: racial/ethnic categories, biological expressions of racism, and the embodiment of inequality—an ecosocial perspective. In: Whitmarsh I, Jones DS, eds. *What's the Use of Race?: Modern Governance and the Biology of Difference*. Cambridge, MA, London, England: The MIT Press; 2010.
28. Kaufman JS. Epidemiologic analysis of racial/ethnic disparities: some fundamental issues and a cautionary example. *Soc Sci Med*. 2008;66:1659–1669.
29. Feng W, Nietert PJ, Adams RJ. Influence of age on racial disparities in stroke admission rates, hospital charges, and outcomes in South Carolina. *Stroke*. 2009;40:3096–3101.
30. Banack HR, Kaufman JS. Does selection bias explain the obesity paradox among individuals with cardiovascular disease? *Ann Epidemiol*. 2015;25:342–349.
31. Banack HR, Kaufman JS. From bad to worse: collider stratification amplifies confounding bias in the "obesity paradox". *Eur J Epidemiol*. 2015;30:1111–1114.
32. Mayeda ER, Glymour MM. The obesity paradox in survival after cancer diagnosis: tools for evaluation of potential bias. *Cancer Epidemiol Biomarkers Prev*. 2017;26:17–20.
33. Vansteelandt S. Asking too much of epidemiologic studies: the problem of collider bias and the obesity paradox. *Epidemiology*. 2017;28:e47–e49.
34. Howe LD, Tilling K, Galobardes B, Lawlor DA. Loss to follow-up in cohort studies: bias in estimates of socioeconomic inequalities. *Epidemiology*. 2013;24:1–9.
35. Howe CJ, Cole SR, Lau B, Napravnik S, Eron JJ Jr. Selection bias due to loss to follow up in cohort studies. *Epidemiology*. 2016;27:91–97.
36. Greenland S. Response and follow-up bias in cohort studies. *Am J Epidemiol*. 1977;106:184–187.
37. Hernán MA. Invited commentary: selection bias without colliders. *Am J Epidemiol*. 2017;185:1048–1050.
38. Poole C. On the origin of risk relativism. *Epidemiology*. 2010;21:3–9.
39. Kaufman JS. Toward a more disproportionate epidemiology. *Epidemiology*. 2010;21:1–2.
40. Kaplan GA, Haan MN, Wallace RB. Understanding changing risk factor associations with increasing age in adults. *Annu Rev Public Health*. 1999;20:89–108.
41. Howard G, Goff DC Jr. A call for caution in the interpretation of the observed smaller relative importance of risk factors in the elderly. *Ann Epidemiol*. 1998;8:411–414.
42. Aune D, Sen A, Prasad M, et al. BMI and all cause mortality: systematic review and non-linear dose-response meta-analysis of 230 cohort studies with 3.74 million deaths among 30.3 million participants. *BMJ*. 2016;353:i2156.
43. Manson JE, Bassuk SS, Hu FB, Stampfer MJ, Colditz GA, Willett WC. Estimating the number of deaths due to obesity: can the divergent findings be reconciled? *J Womens Health (Larchmt)*. 2007;16:168–176.
44. Boyce WT, Ellis BJ. Biological sensitivity to context: I. An evolutionary-developmental theory of the origins and functions of stress reactivity. *Dev Psychopathol*. 2005;17:271–301.
45. Masters RK. Uncrossing the U.S black-white mortality crossover: the role of cohort forces in life course mortality risk. *Demography*. 2012;49:773–796.
46. US Department of Health and Human Services. Office of Disease Prevention and Health Promotion. Healthy People 2020. Available at: https://www.healthypeople.gov/2020/About-Healthy-People. Accessed January 8, 2017.
47. Hernán MA. The hazards of hazard ratios. *Epidemiology*. 2010;21:13–15.