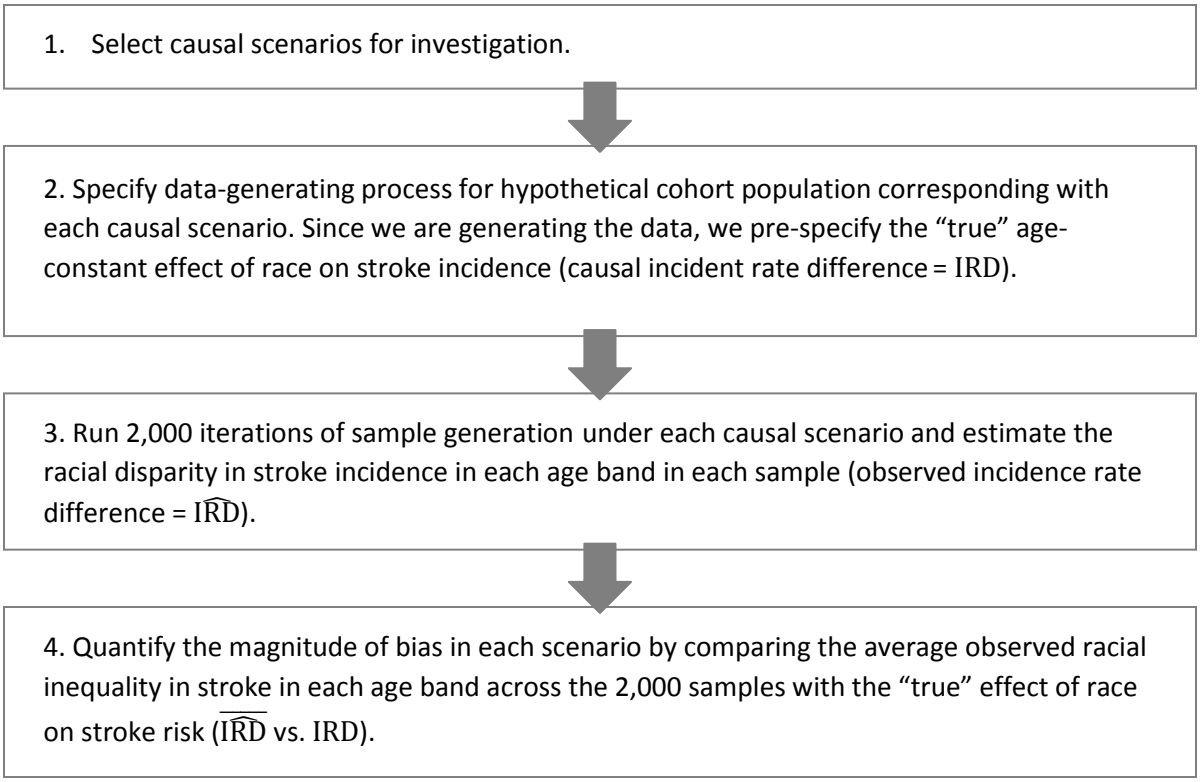
Web appendix

Contents

eAppendix 1. Outline of simulation study steps



eAppendix 2. Assessment of fit of multiplicative versus additive model for effect of race on stroke incidence

To determine whether we should generate the effect of race on stroke incidence as an additive or multiplicative process in our simulations, we assessed whether the incidence rate ratio (IRR) or incidence rate difference (IRD) for the effect of race was a better fit for REGARDS results using updated results provided by Dr. George Howard via personal communication in December 2016 as an update to the numbers published in 2011.[1] Our calculations are displayed in eTable 1. First, we calculated the IRR and IRD for stroke incidence for blacks versus whites age 45-54 years in REGARDS. Second, we calculated the *expected* stroke incidence rate per 10,000 person-years for blacks in each age band (*k*) if the IRR and IRD observed in the 45-54 year age band held constant for other age bands as:

$$\text{expected age band } k \text{ rate in blacks assuming constant IRR}$$
$$= \text{observed age band } k \text{ rate in whites} * \text{observed age band } 45-54 \text{ IRR}$$
$$\text{expected age band } k \text{ rate in blacks assuming constant IRD}$$
$$= \text{observed age band } k \text{ rate in whites} + \text{observed age band } 45-54 \text{ IRD}$$

where the observed age band $45-54$ IRR $= \frac{28.1}{8.4} = 3.36$ and observed age band $45-54$ IRD $=$ $28.1/10{,}000 - 8.4/10{,}000 = 19.7/10{,}000 \approx 20/10{,}000$ person $-$ years.

Third, we calculated the error assuming a constant IRR and IRD as:

$$\text{age band } k \text{ error} = \text{observed age band } k \text{ rate in blacks} - \text{expected age band } k \text{ rate in blacks}$$

Finally, we calculated the sum of squared errors across age band.

Based on the smaller sum of squared errors for the additive compared with the multiplicative model, we concluded that an additive model fit the observed data better, and we generated the effect of race on stroke incidence assuming an age-constant black-white IRD.

eTable 1. Assessment of fit of multiplicative versus additive model for effect of race on stroke incidence

Age band	Observed stroke incidence rate per 10,000 person-years		Assuming a multiplicative model (IRR=3.36 at all ages)			Assuming an additive model (IRD=20/10,000 person-years at all ages)		
	White	Black	Expected rate per 10,000 person-years in blacks if constant IRR	Error with constant IRR	Squared errors with constant IRR	Expected rate per 10,000 person-years in blacks if constant IRD	Error with constant IRD	Squared errors with constant IRD
45-54	8.4	28.1	28.1	0.0	0.0	28.4	-0.2	0.1
55-64	27.2	38.8	91.5	-52.7	2,773.4	47.2	-8.4	70.0
65-74	57.9	71.0	194.7	-123.7	15,302.0	77.9	-6.9	47.3
75-84	112.1	109.4	376.9	-267.5	71,568.2	132.1	-22.7	516.2
85+	153.9	109.0	517.5	-408.5	166,840.4	173.9	-64.9	4,209.6
					Sum of squared errors with constant IRR: 256,484			**Sum of squared errors with constant IRD:** 4,843

eAppendix 3. Details on generation of mortality and stroke data for the primary simulation study

Starting from birth in our hypothetical cohort population for age intervals 0-1, 1-5, and 5-years intervals thereafter, we generated a time to death value for each individual who remained alive up to the beginning of age interval *j*. If the randomly generated time to death at the age of the beginning of the interval *j* exceeded the individual's age of the end of the interval, the individual was considered alive at the beginning of the next age interval, and a new survival time was generated for this next age interval. The process is repeated until the individual's survival time falls within a given age interval or age 94 years, whichever comes first. In our hypothetical study, no strokes occurred prior to age 45 years. Starting at age 45 years, we generated a time to first incident stroke value for each individual who remained alive and stroke-free up to the beginning of age interval *j* as a function of race and *U*. For each individual *i* at the *jth* age interval, time to death and time to stroke variables are generated as random variables drawn from exponential survival distributions based on the hazard functions specified in equations 1 (for mortality) and 2 (for stroke):

$$\mathrm{h}\left(\mathrm{death\_t_{ij}}\middle|\mathrm{x}\right) = \mathrm{death\_}\lambda_{0\mathrm{j}}\exp\left(\gamma_1\mathrm{race_i} + \gamma_2 U_\mathrm{i} + \gamma_3\mathrm{race_i} * U_\mathrm{i} + \gamma_4 * \mathrm{stroke\_history_{ij}}\right) \quad (1)$$

$$\mathrm{h}\left(\mathrm{stroke\_t_{ij}}\middle|\mathrm{x}\right) = \mathrm{stroke\_}\lambda_{0\mathrm{j}_\mathrm{race}}\exp(\beta_1 U_\mathrm{i}), \quad (2)$$

Where $\mathrm{stroke\_}\lambda_{0\mathrm{j}_\mathrm{black}} = \mathrm{stroke\_}\lambda_{0\mathrm{j}_\mathrm{white}} + 20/10{,}000\ \mathrm{person-years}$. Any individual whose time to stroke was longer than his/her time to death was assumed to have died stroke-free. To ensure the pool of people alive and at risk for first incident stroke remained realistically large at older ages, we generated the data such that 25% of strokes were fatal and history of stroke was associated with a twofold increased risk of death (HR=2.0) (input parameter for the effect of stroke on mortality were approximated based on Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association[2]).

eAppendix 4. Methods for supplemental simulation study in which *U* partially mediates effects of race

Objective

In the primary simulation study, *U* was independent of race at birth, but a closely related phenomenon would occur if *U* were influenced by race, *i.e.*, data-generating models where *U* is a partial mediator of the effect of race on stroke incidence (in all scenarios) and mortality (in scenarios consistent with collider-stratification bias). This supplemental simulation study examines this extension of the causal scenarios investigated in the primary simulation study.

Methods

To generate causal structures where *U* is influenced by race, we modified the data-generating process for *U* and the effects of *U* on stroke incidence compared with procedures used in our primary simulation study so that (1) race influences *U,* (2) age-specific stroke incidence rates in whites in the hypothetical cohort reflect age-specific incidence rates in REGARDS for whites (as in the primary simulation study), and (3) the total effect of race on stroke incidence is 20 excess strokes per 10,000 person-years for all age bands (as in the primary simulation study). In the primary simulation study, the effect of race on stroke incidence was additive and the effect of *U* on stroke incidence was multiplicative. In this supplemental simulation study where *U* partially mediates effects of race, we generated the effect of *U* on stroke as additive so that both the direct and indirect (via *U*) effects of race on stroke incidence were additive. To accommodate this, we generated *U* as a variable drawn from a gamma distribution to constrain *U* to positive values. Because the incidence of stroke increases dramatically with age and the total effect of race on stroke incidence is age-constant in our simulations, we generated (1) time invariant *U* (U_i) that is independent of race and (2) time-varying *U* (U_{ij}) that increases with age at the same rate stroke rates increase for whites and is influenced by U_i (time-invariant *U*) and race. Under this data-generating structure, U_i (time-invariant *U*) could represent genetic predisposition for vascular disease and U_{ij} (time-varying *U*) could represent vascular disease. To summarize:

$U_{\mathrm{i}} \sim \mathrm{Gamma}(1, 0.5)$ and $U_{\mathrm{ij}} = \alpha_{0\mathrm{j}} U_{\mathrm{i}} + \alpha_1 \mathrm{race}_{\mathrm{i}}$,

$$\text{where } \alpha_{0\mathrm{j}} = \begin{cases} 1 & \text{if j} < age\ band\ 45-49 \\ \dfrac{\mathrm{stroke}_{\lambda_{\mathrm{jwhite}}}}{\mathrm{stroke}_{\lambda_{45-49\mathrm{white}}}} & \text{if j} \geq \text{age band } 45-49 \end{cases}$$

$$\alpha_1 = 1$$

Time to death is generated based on the mortality hazard function applied in the primary analysis (formula 1 in Appendix 3). Time to stroke is generated based on the hazard function described below (equation 3):

$$\mathrm{h}\left(\mathrm{stroke\_t}_{\mathrm{ij}} \middle| \mathrm{x}\right) = \mathrm{stroke}_{\lambda_{0\mathrm{j}}} + \beta_1 \mathrm{race}_{\mathrm{i}} + \beta_2 U_{\mathrm{ij}} \quad (3)$$

$$\text{where } \beta_1 = 0.0015,\ \beta_2 = 0.0005$$

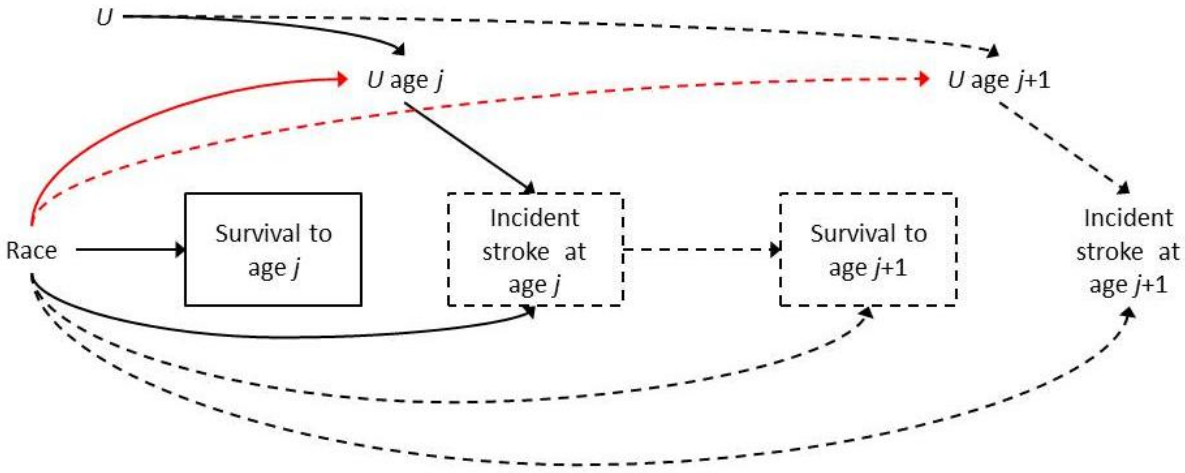
In this data-generating model, the total effect of race on stroke incidence $= \beta_1 + \alpha_1 \beta_2 = 0.0015 + 1 * 0.0005 = 0.002 = 20/10{,}000$ person-years, the age-constant effect of race on stroke incidence in our hypothetical cohort study.

eFigure 1 represents the casual scenarios under investigation in this supplemental analysis, with the path from race to U_{ij} highlighted in red. As in the primary simulation study, we begin with a base scenario with no anticipated survivor bias (Scenario A*). In this scenario, U_{ij} directly influences stroke risk, but has no direct effect on mortality (additional stroke hazard for 1-unit higher U_{ij}=0.0005; 25% of total effect of race on stroke incidence). Scenarios B* and C* are causal structures consistent with survivor bias.[3,4] In Scenario B*, U_{ij} directly influences stroke risk and mortality risk for both blacks and whites (additional stroke hazard for 1-unit higher U_{ij}=0.0005; hazard ratio (HR) for 1-unit higher U_{ij} on mortality=1.5). In Scenario C*, U_{ij} directly influences both stroke risk and mortality risk for blacks (additional stroke hazard for 1-unit higher U_{ij}=0.0005; HR for 1-unit higher U_{ij} on mortality=1.5), but has no direct effect on mortality for whites (additional stroke hazard for 1-unit higher U_{ij}=0.0005; HR for 1-unit higher U_{ij} on mortality=1.0).

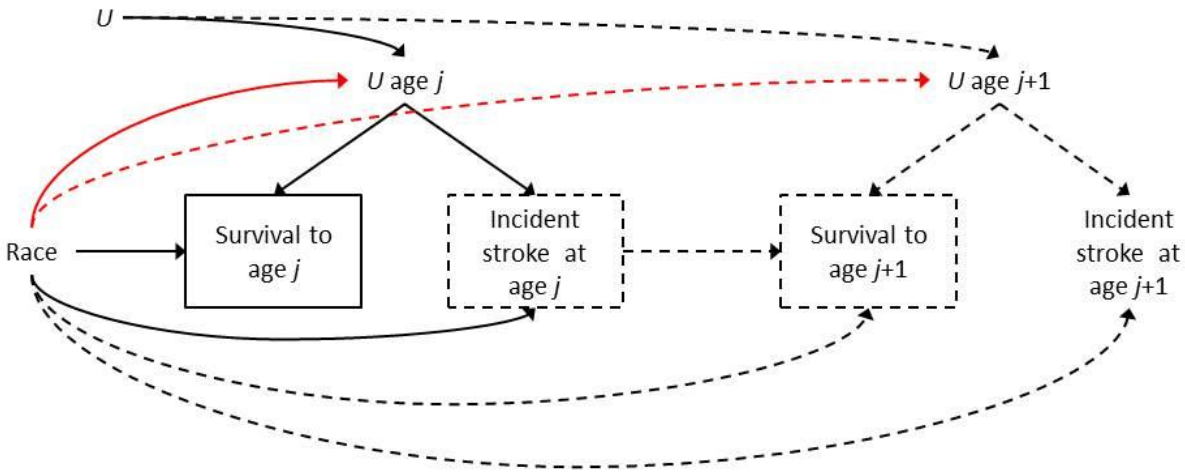
The simulation code, in Stata SE version 13.1 (StataCorp LP, College Station, Texas), is available online (https://github.com/ermayeda/stroke_inequalities_simulation).

eFigure 1. Causal scenarios under investigation in supplementary simulation study. In all scenarios, race influences mortality (following U.S. life table estimates from birth) and stroke risk (assuming a constant effect of black race on stroke risk from age 45-94). *U* represents a time-invariant factor, or set of factors, that influences risk of stroke and mortality and is independent of race at birth. *U* age *j* and *U age j*+1 represent time-varying *U*, which is influenced by time-invariant *U* and race. History of stroke influences mortality, 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 these variables are only relevant after age *j*: 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. eFigure 1A refers to Scenario A*, where time-varying *U* influences stroke, but has no direct effect on mortality. eFigure 1B refers to Scenarios B* and C*. In Scenario B*, time-varying *U* influences stroke risk and directly affects mortality risk. In Scenario C*, time-varying *U* influences stroke risk for both blacks and whites; time-varying *U* only directly affects mortality risk for blacks.

A.



B.

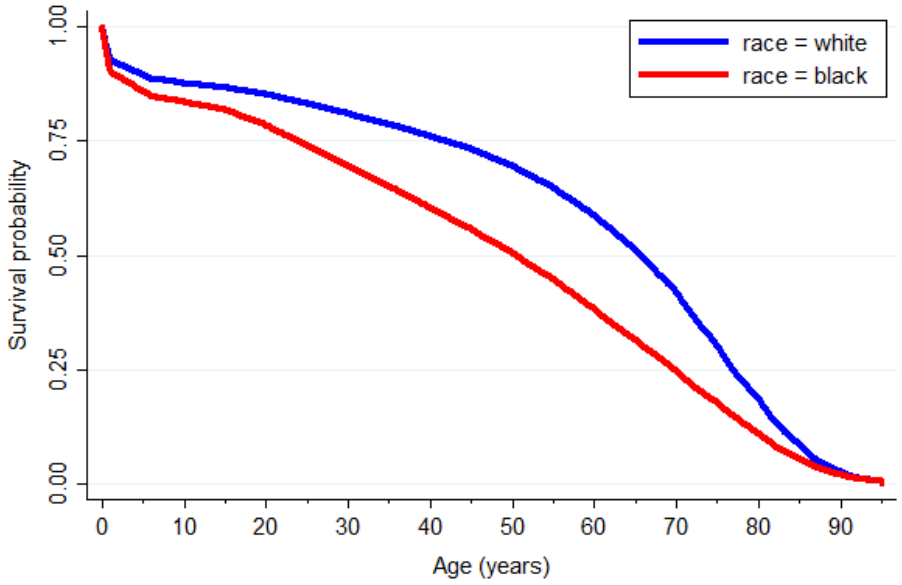

U
U age j
U age j+1
Race
Survival to age j
Incident stroke at age j
Survival to age j+1
Incident stroke at age j+1

eAppendix 5. Additional simulation results from the primary simulation study

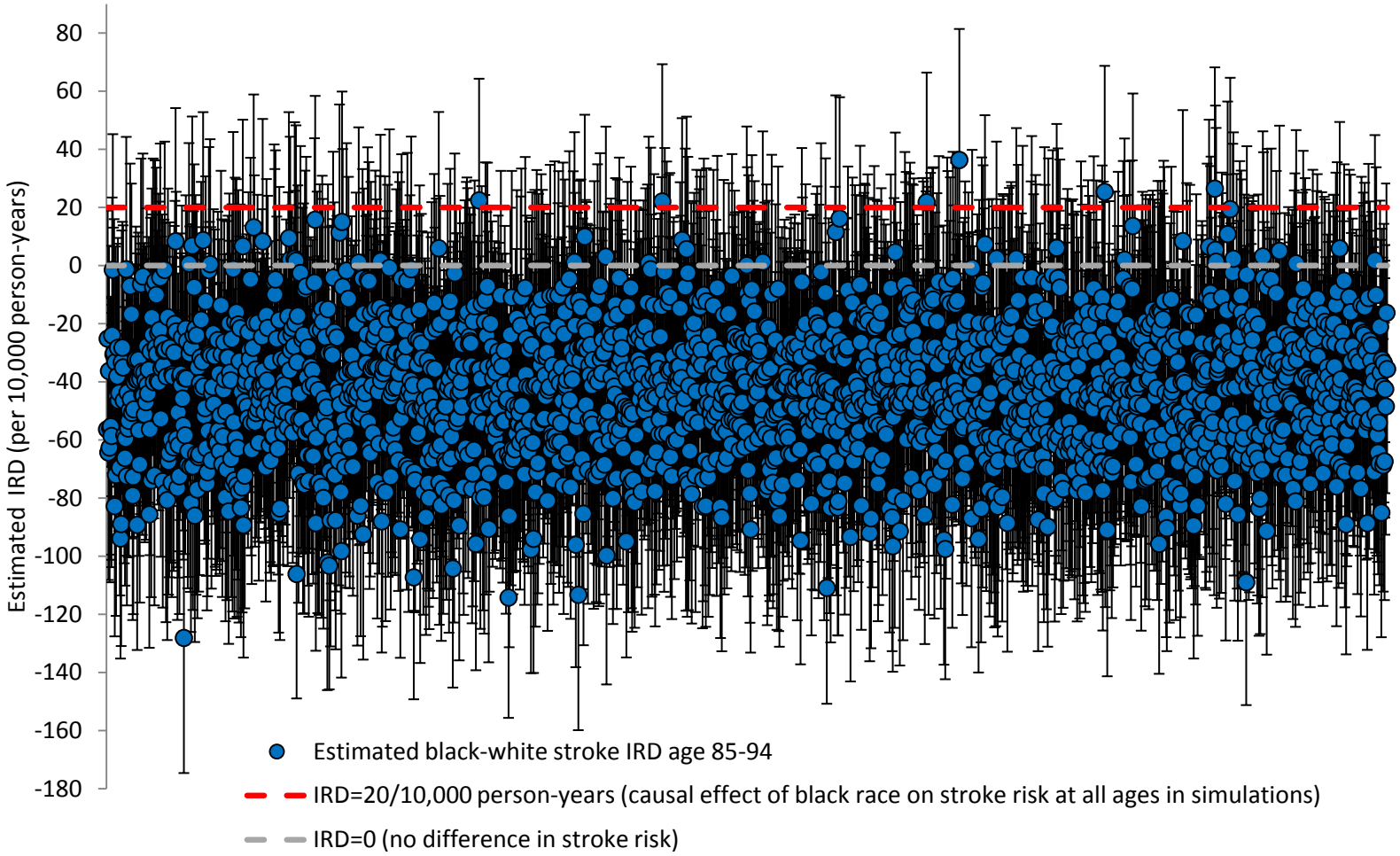
eTable 2. Proportion of birth cohort alive by age and race among simulated study participants from 2,000 simulated samples from the primary simulation study.

	45 years	50 years	55 years	60 years	65 years	70 years	75 years	80 years	85 years	90 years	95 years
Black											
U.S. life tables[5]	0.564	0.514	0.458	0.394	0.327	0.256	0.180	0.114	0.058	0.023	0.007
Scenario A	0.557	0.505	0.447	0.384	0.318	0.247	0.174	0.108	0.054	0.021	0.006
Scenario B	0.544	0.496	0.444	0.387	0.328	0.260	0.189	0.118	0.060	0.024	0.007
Scenario C	0.557	0.506	0.448	0.384	0.318	0.246	0.174	0.109	0.054	0.022	0.006
White											
U.S. life tables[5]	0.737	0.703	0.659	0.600	0.524	0.427	0.311	0.191	0.091	0.030	0.006
Scenario A	0.731	0.694	0.649	0.589	0.514	0.419	0.303	0.185	0.086	0.028	0.006
Scenario B	0.720	0.685	0.644	0.592	0.523	0.435	0.323	0.197	0.093	0.031	0.007
Scenario C	0.731	0.694	0.649	0.588	0.513	0.418	0.303	0.185	0.086	0.027	0.006

eFigure 2. Simulated survival curves from birth to age 95 based on US life tables for 1919-1921 birth cohort from one simulated sample from the primary simulation study.



eFigure 3. Observed black-white stroke incidence rate difference (IRD) and 95% confidence interval in 85-94 age band in Scenario C across 2,000 simulated samples from the primary simulation study. Note that the width of the 95% confidence intervals is influenced by the sample size of people at risk for stroke in the 85-94 age band in our simulations.



eTable 3. Black-white stroke incidence rate ratio (IRR) results from the REGARDS cohort and from simulation scenarios from the primary simulation study (2,000 simulated samples).

	Age 45-54 years				Age 55-64 years				Age 65-74 years				Age 75-84 years				Age 85-94 years			
	IRR[a]	ln(IRR)[b] (SE ln(IRR)[c]	% bias[d]	RMSE[e]	IRR[a]	ln(IRR)[b] (SE ln(IRR)[c]	% bias[d]	RMSE[e]	IRR[a]	ln(IRR)[b] (SE ln(IRR)[c]	% bias[d]	RMSE[e]	IRR[a]	ln(IRR)[b] (SE ln(IRR)[c]	% bias[d]	RMSE[e]	IRR[a]	ln(IRR)[b] (SE ln(IRR)[c]	% bias[d]	RMSE[e]
REGARDS[f]	3.36	1.21 (0.33)	-	-	1.43	0.36 (0.12)	-	-	1.23	0.20 (0.09)	-	-	0.98	-0.02 (0.12)	-	-	0.71	-0.34 (0.35)	-	-
Truth in simulations	3.39	-	-	-	1.73	-	-	-	1.35	-	-	-	1.18	-	-	-	1.13	-	-	-
Scenario A[g]	3.51	1.25 (0.11)	5%	0.11	1.78	0.57 (0.08)	6%	0.08	1.33	0.29 (0.07)	-4%	0.07	1.18	0.16 (0.08)	3%	0.08	1.15	0.13 (0.16)	12%	0.16
Scenario B[h]	3.24	1.17 (0.11)	-6%	0.12	1.63	0.49 (0.08)	-14%	0.10	1.21	0.19 (0.07)	-39%	0.13	1.07	0.07 (0.08)	-59%	0.12	1.07	0.06 (0.15)	-45%	0.16
Scenario C[i]	3.14	1.14 (0.11)	-11%	0.14	1.52	0.42 (0.08)	-29%	0.16	1.08	0.08 (0.07)	-76%	0.23	0.88	-0.14 (0.09)	-169%	0.31	0.73	-0.33 (0.18)	-308%	0.49

See eTable 4 for average stroke incidence rate estimates for blacks and whites in each scenario.

[a] IRR = stroke incidence rate ratio for blacks versus whites in each age band. For REGARDS, the IRR is the observed age-specific IRR in REGARDS. For truth in simulations, IRR=3.39 in the 45-54 age band and decreases with age. For simulation scenario results, the IRR is the average of the observed age-specific IRR across 2,000 simulated samples for each simulation scenario.

[b] ln(IRR) = log stroke incidence rate ratio for blacks versus whites in each age band. For REGARDS, the ln(IRR) is the observed age-specific ln(IRR) in REGARDS. For simulation scenario results, the ln(IRR) is the average of the observed age-specific ln(IRR) across 2,000 simulated samples for each simulation scenario.

[c] For REGARDS, the SE is the SE for REGARDS age-specific ln(IRR) estimates. For simulations, the SE is the empirical SE, calculated as standard deviation of the observed age-specific ln(IRR) across 2,000 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).

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

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

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

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

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

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

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

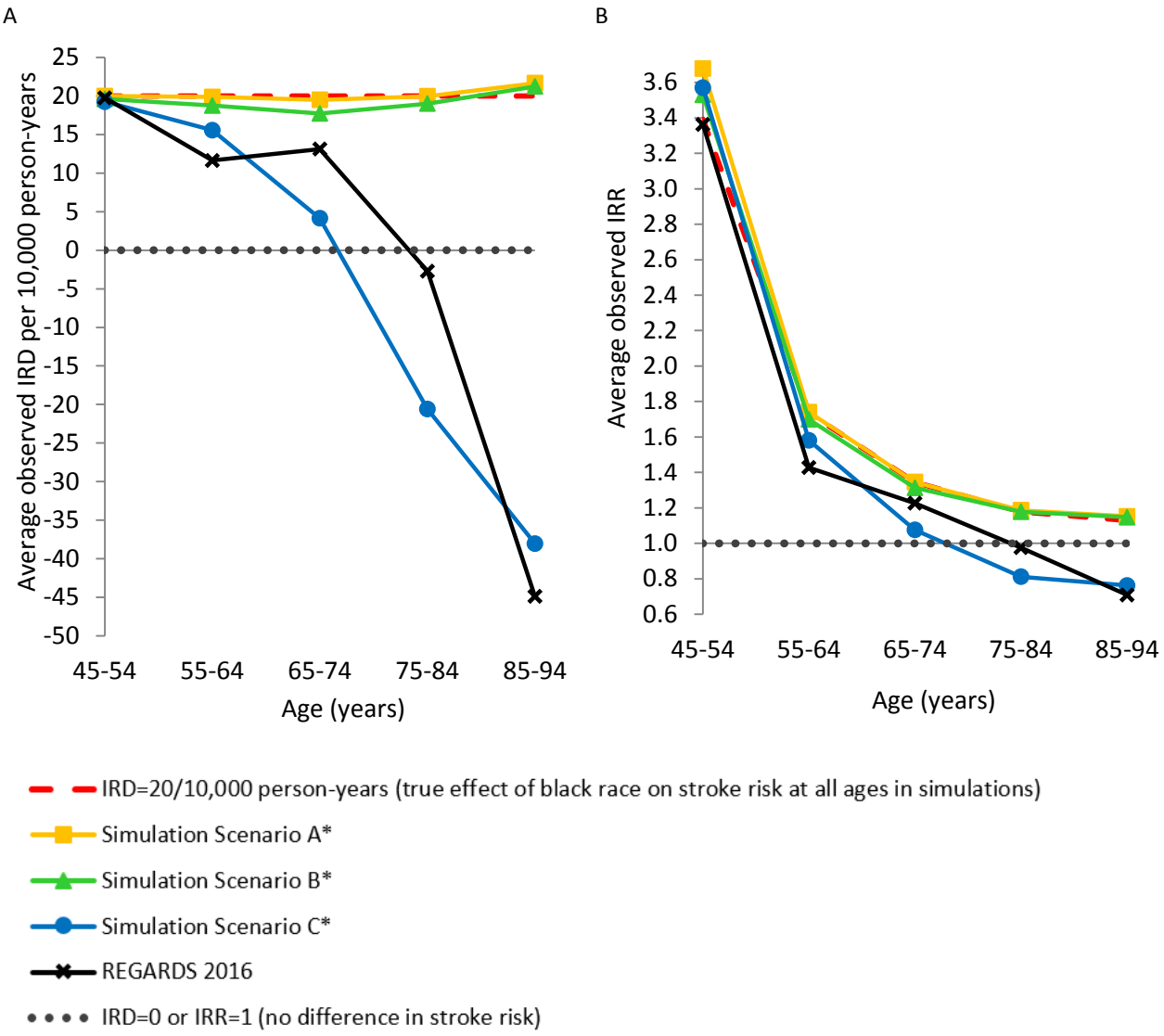
eTable 4. Average stroke incidence rates (per 10,000 person-years) by age band simulated study participants from 2,000 simulated samples from the primary simulation study.

	45-54 years	55-64 years	65-74 years	75-84 years	85-94 years
Black					
REGARDS[1]	28.1	38.8	71.0	109.4	109.0
Scenario A	30.3	48.5	81.6	130.0	173.4
Scenario B	27.4	43.0	72.5	123.1	166.6
Scenario C	27.0	41.4	66.4	96.2	111.5
White					
REGARDS[1]	8.4	27.2	57.9	112.1	153.9
Scenario A	8.7	27.3	61.3	110.2	152.9
Scenario B	8.5	26.4	60.0	114.9	156.9
Scenario C	8.7	27.3	61.4	110.1	154.5

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

Appendix 6. Results for supplemental simulation study in which *U* partially mediates effects of race

eFigure4. 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 2,000 simulated samples per scenario for the supplemental simulation study. For comparison, results observed in the REGARDS cohort are also displayed. The red dashed line shows the true effect of black race on stroke risk at all ages is IRD=20/10,000 person-years in the supplemental simulation study.



eTable 5. Black-white stroke incidence rate difference (IRD) results from the REGARDS cohort and from simulation scenarios from the supplemental simulation study (2,000 simulated samples). In Scenario A* (time-varying *U* influences stroke, but has no direct effect on mortality), there was minimal bias: the observed black-white stroke IRD was approximately 20/10,000 person-years across all age bands, the causal effect of race on stroke incidence specified in the simulation. Bias in Scenario B* (time-varying *U* increases stroke risk and mortality risk) emerged in early old age but was minimal in the oldest groups. In Scenario C* (time-varying *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 the REGARDS Study estimates, similar to Scenario C our primary simulations where race and *U* were independent.

	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[d]	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]	20.0 (1.8)	0%	1.8	19.9 (2.9)	-1%	2.9	19.5 (4.8)	-2%	4.8	20 (9.7)	0%	9.7	21.7 (25.7)	8%	25.7
Scenario B*[g]	19.6 (1.8)	-2%	1.9	18.8 (2.9)	-6%	3.1	17.7 (4.9)	-11%	5.4	19 (10.0)	-5%	10.1	21.2 (28.3)	6%	28.3
Scenario C*[h]	19.3 (1.8)	-4%	2.0	15.6 (2.9)	-22%	5.3	4.1 (4.7)	-79%	16.5	-20.6 (8.7)	-203%	41.5	-38.0 (22.6)	-290%	62.3

See eTable 8 for average 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 2,000 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 age-specific IRD estimates across the 2,000 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 eTable 7).

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

[d] RMSE = root mean square error 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.

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

[e] Scenario A*: U_{ij} increases stroke risk, but has no direct effect on mortality risk.

[f] Scenario B*: U_{ij} increases stroke risk and mortality risk.

[g] Scenario C*: U_{ij} 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.

eTable 6. Black-white stroke incidence rate ratio (IRR) results from the REGARDS cohort and from simulation scenarios from the supplemental simulation study (2,000 simulated samples).

	Age 45-54 years				Age 55-64 years				Age 65-74 years				Age 75-84 years				Age 85-94 years			
	IRR[a]	ln(IRR)[b] (SE ln(IRR)[c]	% bias[d]	RMSE[e]	IRR[a]	ln(IRR)[b] (SE ln(IRR)[c]	% bias[d]	RMSE[e]	IRR[a]	ln(IRR)[b] (SE ln(IRR)[c]	% bias[d]	RMSE[e]	IRR[a]	ln(IRR)[b] (SE ln(IRR)[c]	% bias[d]	RMSE[e]	IRR[a]	ln(IRR)[b] (SE ln(IRR)[c]	% bias[d]	RMSE[e]
REGARDS[f]	3.36	1.21 (0.33)	-	-	1.43	0.36 (0.12)	-	-	1.23	0.20 (0.09)	-	-	0.98	-0.02 (0.12)	-	-	0.71	-0.34 (0.35)	-	-
Truth in simulations	3.39	-	-	-	1.73	-	-	-	1.35	-	-	-	1.18	-	-	-	1.13	-	-	-
Scenario A*[g]	3.68	1.30 (0.12)	12%	0.14	1.74	0.55 (0.08)	1%	0.08	1.35	0.29 (0.07)	0%	0.07	1.19	0.17 (0.08)	5%	0.08	1.15	0.13 (0.16)	17%	0.16
Scenario B*[h]	3.53	1.26 (0.11)	6%	0.12	1.70	0.53 (0.08)	-5%	0.08	1.31	0.27 (0.07)	-9%	0.08	1.18	0.16 (0.08)	0%	0.08	1.15	0.12 (0.17)	13%	0.17
Scenario C*[i]	3.57	1.27 (0.12)	7%	0.12	1.58	0.46 (0.08)	-21%	0.13	1.08	0.07 (0.08)	-78%	0.24	0.81	-0.21 (0.09)	-205%	0.39	0.76	-0.29 (0.18)	-283%	0.45

See eTable 8 for average incidence rate estimates for blacks and whites in each scenario.

[a] IRR = stroke incidence rate ratio for blacks versus whites in each age band. For REGARDS, the IRR is the observed age-specific IRR in REGARDS. For truth in simulations, IRR=3.68 in the 45-54 age band and decreases with age. For simulation scenario results, the IRR is the average of the observed age-specific IRR across 2,000 simulated samples for each simulation scenario.

[b] ln(IRR) = log stroke incidence rate ratio for blacks versus whites in each age band. For REGARDS, the ln(IRR) is the observed age-specific ln(IRR) in REGARDS. For simulation scenario results, the ln(IRR) is the average of the observed age-specific ln(IRR) across 2,000 simulated samples for each simulation scenario.

[c] For REGARDS, the SE is the SE for REGARDS age-specific ln(IRR) estimates. For simulations, the SE is the empirical SE, calculated as standard deviation of the observed age-specific ln(IRR) across 2,000 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 eTable 7).

[d] Percentage bias in the simulation estimates is defined as deviation from the truth,
$100 \times (\overline{\mathrm{IRR}}_{\text{age-specific observed}} - \mathrm{IRR}_{\text{age-specific causal}}) / (\mathrm{IRR}_{\text{age-specific causal}})$.

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

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

[g] Scenario A*: U_{ij} increases stroke risk, but has no direct effect on mortality risk.

[h] Scenario B*: U_{ij} increases stroke risk and mortality risk.

[i] Scenario C*: U_{ij} increases stroke risk for both blacks and whites, but only directly affects mortality for blacks.

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

eTable 7. Proportion of birth cohort alive by age and race among simulated study participants from 2,000 simulated samples from the supplemental simulation study.

	45 years	50 years	55 years	60 years	65 years	70 years	75 years	80 years	85 years	90 years	95 years
Black											
U.S. life tables[5]	0.564	0.514	0.458	0.394	0.327	0.256	0.180	0.114	0.058	0.023	0.007
Scenario A*	0.557	0.505	0.447	0.384	0.319	0.248	0.175	0.109	0.055	0.021	0.006
Scenario B*	0.566	0.516	0.460	0.393	0.325	0.251	0.171	0.104	0.050	0.018	0.005
Scenario C*	0.557	0.506	0.450	0.384	0.318	0.248	0.176	0.111	0.057	0.023	0.007
White											
U.S. life tables[5]	0.737	0.703	0.659	0.600	0.524	0.427	0.311	0.191	0.091	0.030	0.006
Scenario A*	0.731	0.694	0.649	0.589	0.514	0.419	0.304	0.186	0.087	0.028	0.006
Scenario B*	0.739	0.704	0.659	0.598	0.520	0.420	0.298	0.180	0.081	0.025	0.005
Scenario C*	0.731	0.694	0.649	0.589	0.514	0.419	0.304	0.186	0.087	0.028	0.006

eTable 8. Average stroke incidence rates (per 10,000 person-years) by age band simulated study participants from 2,000 simulated samples from the supplemental simulation study.

	45-54 years	55-64 years	65-74 years	75-84 years	85-94 years
Black					
REGARDS[1]	28.1	38.8	71.0	109.4	109.0
Scenario A*	27.6	46.9	76.4	128.4	176.5
Scenario B*	27.5	45.8	74.9	127.7	180.2
Scenario C*	26.8	42.6	60.9	87.7	116.5
White					
REGARDS[1]	8.4	27.2	57.9	112.1	153.9
Scenario A*	7.6	27.0	56.9	108.5	154.9
Scenario B*	7.9	27.0	57.1	108.7	159.0
Scenario C*	7.6	27.0	56.8	108.3	154.6

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

References

1. Howard VJ, Kleindorfer DO, Judd SE, McClure LA, Safford MM, Rhodes JD, Cushman M, Moy CS, Soliman EZ, Kissela BM. Disparities in stroke incidence contributing to disparities in stroke mortality. *Ann Neurol.* 2011;69:619-627.
2. 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:e38-360.
3. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology.* 2004;15:615-25.
4. Hernán MA, Robins JM. Selection Bias. *Causal Inference* Boca Raton: Chapman & Hall/CRC, forthcoming, 2017.
5. Arias E. United States life tables 2006. *Natl Vital Stat Rep.* 2010;58:1-40.