**Appendix 1: Baseline Characteristics of Participants Eligible and Included into the study vs. Participants Eligible but Excluded due to Missing Memory Score or Risk Factors**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Non-Hispanics aged 50+ with no history of stroke | |  |
|  | Those included in the analysis | Those with missing memory score or risk factors | P-value\* |
| Number of population (%)  Gender | 17341(100%) | 203(100%) |  |
| Male | 7388(42.6%) | 90(44.3%) | <0.001 |
| Female | 9953(57.4%) | 113(55.7%) |  |
| Race |  |  | <0.001 |
| White | 14439(83.3%) | 142(70.0%) |  |
| Non-white | 2902(16.7%) | 61(30.0%) |  |
| Age (SD) | 66.4 (10.3) | 64.9(11.4) | 0.006 |
| Marital status |  |  |  |
| Married/partnered | 11779(67.9%) | 135(66.5%) | 0.67 |
| Widowed/separated/divorced | 5054(29.1%) | 47(23.2%) | 0.06 |
| Never married | 508(2.9%) | 5(2.5%) | 0.69 |
| Years of education (SD) | 12.3(3.0) | 11.3(4.1) | 0.007 |
| Mother’s education years (SD) | 9.4(3.0) | 9.8(3.4) | 0.20 |
| Wealth in $1000 (25th, 75th) | 334.2(47.4, 347.0) | 317.2(30.0, 348.5) | 0.04 |
| Height in meters (SD) | 1.69(0.1) | 1.71(0.1) | 0.007 |
| Place of birth |  |  |  |
| Northeast | 3618(20.9%) | 29(14.3%) | 0.02 |
| Midwest | 5170(29.8%) | 26(12.8%) | <0.001 |
| South | 6315(36.4%) | 75(37.0%) | 0.88 |
| West | 1266(7.3%) | 12(5.9%) | 0.45 |
| Non US | 972(5.6%) | 14(6.9%) | 0.42 |

Descriptive data are not weighted.

Values in the table are counts (percentage) for categorical variables and means (SD) for continuous variables. SD is standard deviation. 25th and 75th are the first and the third quartiles.

\*Null hypothesis for P-value: Baseline characteristics are the same for eligible participants included in the study and excluded due to missing memory score or risk factors.

**Appendix 2: Full Parameterization of the Model includes all Stroke Survivors, Stroke Decedents and Stroke-free Cohort Members who were ≤70 Years Old at Baseline**

*Average Memory Score=*

*β0 (representing average memory score for individuals 65 years old in the reference category who remained stroke-free throughout follow-up)*

*+ β1\*Indicator for whether ever survived a stroke (1 if yes, 0 for individuals who had a stroke but did not survive or individuals who remained stroke-free)*

*+β2\* Years before stroke onset among stroke survivors (# years, or 0 for individuals who have already survived a stroke or individuals who had a stroke but did not survive or individuals who remained stroke-free)*

*+β3\* Indicator variable for whether already survived a stroke (1 if yes, or 0 for individuals who have not yet had a stroke or individuals who remained stroke-free)*

*+β4\* Years after survived stroke (# years, or 0 for individuals who have not yet had a stroke or individuals who remained stroke-free)*

*+β5\*Indicator for whether ever had a stroke and did not survive (1 if yes, or 0 if survived a stroke or remained stroke free)*

*+β6\*Years before stroke which was not survived (# years or 0 if survived a stroke or remained stroke free)*

*+β7\*Age at stroke centered at 65 years (age in years, or 0 if remained stroke free)*

*+β8\*Age at each interview centered at 65 years (age for people who remained stroke free; 0 if had a stroke)*

*+βi\*Other baseline covariates*

**Appendix 3: Baseline Characteristics of Participants who Stayed until the End of Follow-up and Censored before the Final Follow-up**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Those who stayed until the end of follow-up | Those censored due to dropout before the final follow-up | P-value\* |
| Number of population (%) | 10903(100%) | 6438(100%) |  |
| Last available memory score (SD) | 1.8(1.4) | 0.8(2.1) | <0.001 |
| Gender |  |  | <0.001 |
| Male | 4432(40.7%) | 2956(45.9%) |  |
| Female | 6471(59.3%) | 3482(54.1%) |  |
| Race |  |  | 0.18 |
| White | 9110(83.6%) | 5329(82.8%) |  |
| Non-white | 1793(16.4%) | 1109(17.2%) |  |
| Age (SD) | 63.3(8.3) | 71.7(11.2) | <0.001 |
| Marital status |  |  |  |
| Married/partnered | 8026(73.6%) | 3753(58.3%) | <0.001 |
| Widowed/separated/divorced | 2600(23.9%) | 2454(38.1%) | <0.001 |
| Never married | 277(2.5%) | 231(3.6%) | <0.001 |
| Years of education (SD) | 12.7(2.8) | 11.7(3.2) | <0.001 |
| Mother’s education years (SD) | 9.7(3.1) | 7.5(2.6) | <0.001 |
| Wealth in $1000 (25th, 75th) | 355.2(59.0, 383.0) | 298.6(30.0, 283.0) | <0.001 |
| Height in meters (SD) | 1.69(0.1) | 1.69(0.1) | 0.31 |
| Place of birth |  |  |  |
| Northeast | 2250(20.6%) | 1368(21.3%) | 0.34 |
| Midwest | 3340(30.6%) | 1830(28.4%) | 0.002 |
| South | 3888(35.7%) | 2427(37.7%) | 0.007 |
| West | 827(7.6%) | 439(6.8%) | 0.06 |
| Non US | 598(5.5%) | 374(5.8%) | 0.37 |
| Descriptive data are not weighted.  Values in the table are counts (percentage) for categorical variables and means (SD) for continuous variables. SD is standard deviation. 25th and 75th are the first and the third quartiles.  \*Null hypothesis for P-value: Baseline characteristics are the same for participants stayed until the end of follow-up and censored before the final follow-up. | | | |

**Appendix 4: Sensitivity analysis for bias due to differential stroke probability**

In this appendix, we estimate the range of bias that might plausibly be induced in analyses restricted to stroke survivors. This selection reflects both differences in probability that an individual *has a stroke* and differences in probability the stroke is survived, but in our bias estimates, we consider this as a joint process and make estimates based on assumptions about the probability that someone both had and survived a stroke. We consider the example of age, which is a strong determinant of stroke risk. Let us assume the existence of a single unmeasured confounder of stroke and memory decline, such as a gene that causes fast progression of cerebrovascular disease. Estimating the bias introduced by stratifying on stroke status requires assumptions about the prevalence of the gene among young stroke survivors versus the prevalence of the gene among old stroke survivors. The magnitude of selection bias is then estimated as γ\*δ, where γ is the constant effect of the underlying gene on memory decline conditioning on age stratum and other covariates in the model among stroke survivors, and δ is the prevalence difference of the gene comparing stroke survivors in the older stratum to the younger stratum conditioning on other covariates.[53](#_ENREF_56) The difference in the prevalence of the gene between young and old stroke survivors is a function of the prevalence of the gene in the pre-stroke population and the effect of the gene on stroke risk in young and old individuals, conditioning on other covariates. For example, let us assume that 20% of the young stroke-free population has the “high stroke risk” gene which doubles their risk of stroke, and 2% of the young population will develop nonfatal stroke; therefore, around 19% of the older stroke-free population will carry the risk allele. We also assume that although this gene doubles the risk of nonfatal stroke in the young, it only increases the risk by 50% in the old population. Under these assumptions, 33% of young stroke survivors will carry the gene but only 26% of old stroke survivors will be carriers. Thus, the prevalence difference () is 7%. If this gene accelerates memory decline by 0.1 SD (which would approximately double rate of decline associated with age), the magnitude of bias introduced by the gene is 0.1\*0.07=0.007. The table below shows a range of bias estimates for the young/old comparison. The range we consider most plausible is highlighted in gray. Compared with the magnitude of the annual memory decline for all time periods in regards to stroke, this sensitivity analysis suggests that the magnitude of bias is trivial.

Appendix Table: Estimates of the Range of Selection Bias among Stroke Survivors, Conditioning on all the Other Covariates#

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Effect of the gene on rate of memory decline conditioning on age stratum | RR for the effect of gene on nonfatal stroke among people in the younger stratum | RR for the effect of gene on nonfatal stroke among people in the older stratum | Prevalence of the gene in the young population at baseline | | | |
| 0.05 | 0.10 | 0.20 | 0.50 |
| 0.05 | 2.0 | 1.0 | 0.002 | 0.004 | 0.007 | 0.009 |
| 0.05 | 2.0 | 1.5 | 0.001 | 0.002 | 0.003 | 0.004 |
| 0.10 | 2.0 | 1.0 | 0.005 | 0.009 | 0.014 | 0.018 |
| 0.10 | 2.0 | 1.5 | 0.002 | 0.004 | 0.007 | 0.008 |
| 0.20 | 2.0 | 1.0 | 0.009 | 0.017 | 0.028 | 0.036 |
| 0.20 | 2.0 | 1.5 | 0.005 | 0.009 | 0.014 | 0.016 |

# Covariates include race, sex, marital status, household wealth level, years of education, mother’s years of education, height and birth place