# Self-Reported Stroke Risk Stratification
## Reasons for Geographic and Racial Differences in Stroke Study

George Howard, DrPH; Leslie A. McClure, PhD; Claudia S. Moy, PhD;
Virginia J. Howard, PhD; Suzanne E. Judd, PhD; Ya Yuan, MS; D. Leann Long, PhD;
Paul Muntner, PhD; Monika M. Safford, MD; Dawn O. Kleindorfer, MD

***Background and Purpose***—The standard for stroke risk stratification is the Framingham Stroke Risk Function (FSRF), an equation requiring an examination for blood pressure assessment, venipuncture for glucose assessment, and ECG to determine atrial fibrillation and heart disease. We assess a self-reported stroke risk function (SRSRF) to stratify stroke risk in comparison to the FSRF.

***Methods***—Participants from the REGARDS study (Reasons for Geographic and Racial Differences in Stroke) were evaluated at baseline and followed for incident stroke. The FSRF was calculated using directly assessed stroke risk factors. The SRSRF was calculated from 13 self-reported questions to exclude those with prevalent stroke and assess stroke risk. Proportional hazards analysis was used to assess incident stroke risk using the FSRF and SRSRF.

***Results***—Over an average 8.2-year follow-up, 939 of 23 983 participants had a stroke. The FSRF and SRSRF produced highly correlated risk scores ($r_{Spearman}$=0.852; 95% confidence interval, 0.849–0.856); however, the SRSRF had higher discrimination of stroke risk than the FSRF ($c_{SRSRF}$=0.7266; 95% confidence interval, 0.7076–0.7457; $c_{FSRF}$=0.7075; 95% confidence interval, 0.6877–0.7273; $P$=0.0038). The 10-year stroke risk in the highest decile of predicted risk was 11.1% for the FSRF and 13.4% for the SRSRF.

***Conclusions***—A simple self-reported questionnaire can be used to identify those at high risk for stroke better than the gold standard FSRF. This instrument can be used clinically to easily identify individuals at high risk for stroke and also scientifically to identify a subpopulation enriched for stroke risk. (***Stroke***. 2017;48:1737-1743. DOI: 10.1161/STROKEAHA.117.016757.)

**Key Words:** atrial fibrillation ◼ blood pressure ◼ electrocardiography ◼ risk factors ◼ stroke

The identification of individuals at high risk for stroke is important for the clinical management of patients, as well as for enabling efficient identification of high-risk cohorts for primary stroke prevention trials. The long-standing standard for stroke risk stratification in the United States is the Framingham Stroke Risk Function (FSRF)[1] that has recently been updated to reflect current risk factor levels and temporal changes in the association between risk factors and stroke risk.[2] However, the FSRF requires a blood draw to establish diabetes mellitus, a blood pressure measurement to obtain systolic blood pressure levels, and an ECG to establish atrial fibrillation.

Similarly, the standard for risk stratification for heart disease has been the Framingham Coronary Risk Score.[3] There have been two proposed simple approaches for risk stratification of heart disease: the first using self-reported information[4] and the other "blood-less risk function" (not requiring laboratory measures).[5] Both of these risk functions perform comparably to the Framingham Coronary Risk Score for risk prediction,[6] and the bloodless risk score has subsequently been validated in other populations.[7–9]

However, no simplified approach has been proposed for stroke risk stratification, and the goal of this analysis is to develop a simple self-reported stroke risk function (SRSRF). This would allow quick and easy identification of populations requiring more intensive clinical assessment and management for reduction of stroke risk. Also, successful development of an SRSRF would permit efficient telephone screening to identify the large high-risk cohort necessary to mount randomized trials for the primary prevention of stroke.

## Methods

The REGARDS study (Reasons for Geographic and Racial Differences in Stroke) randomly sampled individuals from a

commercially available list (Genesys, Inc), the same source used by other studies, including the Behavioral Risk Factor Surveillance System conducted by the Center for Disease Control and Prevention.[10] Using a combination of mail and telephone contact, 30 239 community-dwelling non-Hispanic black and white participants aged 45+years were recruited between 2003 and 2007. At baseline, a telephone interview was conducted to provide a cardiovascular risk profile, including assessment of demographics self-reported cardiovascular risk factors, and previous cardiovascular procedures. An in-person examination was conducted ≈2 to 3 weeks after the telephone interview and included direct assessment of blood pressure, venipuncture, and ECG. Additional details of the study design have been previously published.[11] The study was approved by the institutional review board at all participating institutions.

After baseline assessment, participants were contacted at 6-month intervals for ascertainment of potential stroke events. Medical records for suspected stroke events were retrieved and adjudicated by a panel of stroke clinicians.

The FSRF was calculated as described in a recently reported article[2] and with factors including:

– Demographic factors: age and sex (note that black race is not included in the FSRF);
– Stroke risk factors:
  – Systolic blood pressure, calculated as the average of 2 measurements taken a minute apart after the participant had rested for 5 minutes;
  – Self-reported use of antihypertensive medications;
  – Diabetes mellitus, defined as a fasting glucose of ≥126 mg/dL (or ≥200 mg/dL for those failing to fast) or self-reported use of medications for glucose control;
  – Current cigarette smoking, assessed by a positive response to both questions Have you ever smoked at least 100 cigarettes in your lifetime? and Do you smoke cigarettes now, even occasionally?;
  – Atrial fibrillation, defined by self-report (as below for the SRSRF) plus ECG evidence;
  – History of heart disease, defined by self-reported myocardial infarction (as below for the SRSRF) or self-reported revascularization (coronary artery bypass graft, angioplasty, or stenting), or ECG evidence of a previous myocardial infarction.

Factors established in the literature to be associated with stroke risk that are easily self-reported by participants were considered for inclusion in the SRSRF. These factors included:

– Demographic factors: age, race (black or white), and sex;
– Self-report of a physician diagnosis of cardiovascular risk factors that are included in the FSRF,[1] specifically hypertension, diabetes mellitus, atrial fibrillation, and heart disease. Each of these was assessed by the question Has a doctor or other health professional ever told you that you had ___?; that was separately asked for:
  – High blood pressure,
  – Diabetes mellitus or high blood sugar,
  – Atrial fibrillation, and
  – Myocardial infarction or heart attack;
– Current cigarette smoking, as defined in the FSRF above;
– Education[12] that was classified as less than high school, high school graduate, some college, or college graduate;
– General self-reported health,[13] assessed by the question In general, would you say that your health is excellent, very good, good, fair, or poor?;
– A history of stroke symptoms as assessed by the Questionnaire to Verify Stroke-Free Status.[14]

A questionnaire with these items is provided in the online-only Data Supplement.

For the purpose of this analysis, participants who self-reported a history of stroke or transient ischemic attack at baseline were excluded. In addition, all participants missing any data item considered for the SRSRF or FSRF were excluded from analysis.

## Statistical Analysis

Four risk factors were evaluated using direct measurements or laboratory values (plus self-report) in the FSRF but were assessed by self-report only in the SRSRF (hypertension, diabetes mellitus, atrial fibrillation, and heart disease). The agreement between these 4 risk factors evaluated incorporating direct measures and self-report only was assessed using the κ-statistic.

The estimated stroke risk from the FSRF incorporates age (ie, age is an integral component of the risk calculation) and is sex specific. With the FSRF calculated to incorporate both the effect of age and sex, the association between the calculated FSRF score and stroke risk was assessed in univariate analysis. The associations between both SRSRF factors and FSRF score with the risk of incident stroke were assessed using separate proportional hazards analyses. The discriminatory ability of the SRSRF factors and the FSRF was assessed using the c-statistic.[15] The c-statistic represents the likelihood that for 2 participants chosen at random, the participant with a higher predicted stroke risk will have a shorter time to their stroke event, so c-statistic values of 0.5000 indicate no discriminatory ability, whereas a value of 1.0000 indicates perfect discrimination. Because differences between alternative models are relatively small, we report c-statistics to 4 significant digits. The calibration of both the proportional hazards models for both SRSRF multivariable model and the FSRF score was assessed by comparing the observed with the predicted probability of stroke events within deciles of predicted stroke risk. A test of whether the c-statistics from the SRSRF multivariable model and the FSRF score differed was constructed as a Wald statistic, where the SE of the difference in the c-statistic estimated by bootstrap methods with 500 replications.

Among the 30 239 REGARDS participants, 3100 (10%) self-reported stroke or transient ischemic attack (TIA) at baseline, reducing the cohort to 27 139 participants. Of these, 2789 (10%) were missing information for ≥1 factors in the SRSRF or FSRF, and follow-up data were not available on 367 (2%) of the participants, collectively reducing the data set to 23 983 participants (88% of those stroke/TIA free at baseline).

## Results

Over an average follow-up of 8.2 years, 939 of the 23 983 (4%) participants had an incident stroke. A description of this study population for those with and without a stroke during follow-up is provided in Table 1. Those who had a stroke during follow-up were older, more likely black and male, and have a poorer risk factor profile for both the SRSRF and FSRF factors.

## Discrimination

The hazard ratio for a 10-point difference in the FSRF was 1.75 (95% confidence interval [CI], 1.66–1.84), with a c-statistic of 0.7075 (95% CI, 0.6877–0.7273).

In the univariate model, age was strongly associated with stroke risk (hazard ratio=1.92 for a 10-year difference; 95% CI, 1.79–2.06), with a c-statistic of 0.6765 (95% CI, 0.6548–0.6983). The addition of sex to the model with age increased the c-statistic to 0.6790 (95% CI, 0.6576–0.7004), and the addition of race and age-by-race interaction (factors included in the SRSRF but not the FSRF) to the model increased the c-statistic to 0.6942 (95% CI, 0.6740–0.7144).

Table 2 provides the hazard ratio and c-statistic for the addition of each of the factors considered for the SRSRF. The single factor with the largest impact on discrimination was the general health question, where compared with the full demographic model, there was a 0.0133 increase in the c-statistic, from 0.6942 for the demographic model to 0.7075 (95% CI, 0.6879–0.7271).

**Table 1.    Description of the Study Population for Demographic Factors, Factors Included in the SRSRF, and FSRF**

| | | | No Stroke During Follow-Up (n=23 044) | Stroke During Follow-Up (n=939) |
|---|---|---|---|---|
| Demographic factors | Age | | 64±9 | 70±9 |
| | Black race | | 40 | 42 |
| | Male sex | | 45 | 53 |
| SRSRF factor | Factors included in the SRSRF | Self-reported hypertension | 55 | 67 |
| | | Self-reported diabetes mellitus | 21 | 30 |
| | | Self-reported atrial fibrillation | 7 | 12 |
| | | Self-reported MI | 7 | 14 |
| | | Cigarette smoking | 14 | 16 |
| | | Education | Less than HS | 11 | 14 |
| | | | HS grad | 25 | 31 |
| | | | Some college | 27 | 25 |
| | | | College grad | 36 | 29 |
| | | General self-reported health | Excellent | 17 | 14 |
| | | | Very good | 32 | 27 |
| | | | Good | 35 | 38 |
| | | | Fair | 13 | 17 |
| | | | Poor | 3 | 5 |
| | Self-reported stroke symptoms | | 17 | 21 |
| | Interview-based stroke risk function (mean±SD) | | 5.0±4.3 | 8.0±5.9 |
| FSRF factors | Factors included in FSRF | Systolic blood pressure | 127±16 | 132±17 |
| | | Use of antihypertensive medications | 49 | 60 |
| | | Diabetes mellitus (glucose+self-reported medication) | 20 | 27 |
| | | Atrial fibrillation (ECG or self-report) | 7 | 13 |
| | | Cigarette smoking | As above | As above |
| | | History of heart disease (self-reported MI or revascularization+ECG evidence of MI) | 15 | 27 |
| | FSRF, mean±SD | | 6.7±7.4 | 11.3±9.6 |

FSRF indicates Framingham Stroke Risk Function; HS, high school; MI, myocardial infarction; and SRSRF; Self-Reported Stroke Risk Function.

Although self-reported stroke symptoms were significant after adjustment for demographic factors (hazard ratio=1.34; 95% CI, 1.14–1.57), it became nonsignificant in the multivariable model ($P=0.16$) and was not considered in subsequent analysis. In multivariable analysis, the other SRSRF factors were each strongly associated with stroke risk, with a c-statistic of 0.7266 (95% CI, 0.7076–0.7457), which was significantly ($P=0.0038$) higher than for the FSRF. Information on the final multivariable coefficients and calculation of the risk score are provided in the online-only Data Supplement.

### Agreement Between SRSRF and FSRF

The agreement ($\kappa$) between self-reported risk factors and risk factors involving direct measurement was generally good to very good: hypertension =0.796 (95% CI, 0.788–0.804), diabetes mellitus =0.826 (95% CI, 0.817 – 0.834), atrial fibrillation =0.963 (95% CI, 0.956–0.969), and myocardial infarction =0.588 (95% CI, 0.572–0.604).

The observed agreement between the SRSRF and FSRF is shown in Figure 1, where the Spearman rank correlation between the measures was 0.852 (95% CI, 0.849–0.856).

### Calibration

For the FSRF (Figure 2A), the predicted 10-year stroke risk is substantially higher than the observed 10-year stroke risk at the higher quartiles; however, there is no evidence of discordance between observed and predicted values (Hosmer–Lemeshow $\chi^2=3.2$; $P=0.95$). Furthermore, the observed 10-year stroke risk increased monotonically from 0.7% in the lowest decile of observed risk to 11.1% in the highest decile of predicted stroke risk. There was less discordance at the higher deciles for the SRSRF, and again no evidence of a discordance of observed and

**Table 2. Associations and c-Statistics for Self-Reported Factors After Adjustment for Age, Race, Age-by-Race Interaction and Sex**

| | | After Adjustment for Age, Race, Age-by-Race Interaction and Sex | | Multivariable Model | |
| --- | --- | --- | --- | --- | --- |
| | | Hazard Ratio | c-Statistic | Hazard Ratio | c-Statistic |
| Self-reported hypertension | | 1.47 (1.28–1.69) | 0.6996 (0.6796–0.7195) | 1.28 (1.11–1.48) | 0.7266 (0.7076–0.7457) |
| Self-reported diabetes mellitus | | 1.60 (1.39–1.84) | 0.7003 (0.6805–0.7201) | 1.37 (1.18–1.59) | |
| Self-reported atrial fibrillation | | 1.61 (1.32–1.97) | 0.6979 (0.6778–0.7181) | 1.39 (1.13–1.71) | |
| Self-reported MI | | 1.83 (1.51–2.21) | 0.7027 (0.6827–0.7228) | 1.50 (1.24–1.82) | |
| Smoking | | 1.76 (1.47–2.10) | 0.7016 (0.6814–0.7217) | 1.68 (1.40–2.01) | |
| Education | Less than HS | 1.41 (1.14–1.75) | 0.6982 (0.6782–0.7182) | 1.15 (0.92–1.43) | |
| | HS grad | 1.54 (1.30–1.82) | | 1.37 (1.16–1.62) | |
| | Some college | 1.21 (1.02–1.44) | | 1.13 (0.95–1.35) | |
| | College grad | 1.00 (ref) | | 1.00 (ref) | |
| General self-reported health | Excellent | 1.00 (ref) | 0.7075 (0.6879–0.7271) | 1.00 (ref) | |
| | Very good | 1.03 (0.84–1.28) | | 0.93 (0.75–1.15) | |
| | Good | 1.38 (1.13–1.69) | | 1.10 (0.90–1.36) | |
| | Fair | 1.74 (1.37–2.20) | | 1.21 (0.94–1.55) | |
| | Poor | 3.12 (2.21–4.42) | | 1.89 (1.32–2.73) | |
| Self-reported stroke symptoms | | 1.34 (1.14–1.57) | 0.6971 (0.6770–0.7172) | Not significant ($P=0.16$) and removed from model | |

The c-statistic for this demographic model was 0.6938 (95% CI, 0.6737–0.7138). The multivariable model provides the hazard ratios (with 95% CI) and the c-statistic for the joint effects of all factors. CI indicates confidence interval; HS, high school; and MI, myocardial infarction.

predicted values (Hosmer–Lemeshow $\chi^2=0.08$; $P=0.99$). The observed 10-year stroke risk increased monotonically from 0.7% in the lowest decile of predicted risk to 13.4% in the highest decile of predicted stroke risk, corresponding to a 1.4% annual stroke risk.

### Description of Population by SRSRF Decile

Table 3 provides a description of the self-reported characteristics of the population within each decile of SRSRF estimated stroke risk. At a higher predicted stroke risk, participants were older, more likely to be male, and had a monotonically increasing prevalence of all risk factors. The proportion of black participants increased from decile 1 to 7 from 7% to 54% but then decreased to 39% in the highest risk decile. Participants in the highest risk decile had an average age of 77 years and clearly had a constellation of risk factors and low socioeconomic status; however, there was still a broad age distribution in this risk stratum with 10% of the participants ≤53 years of age, 25% with age ≤58, and half with age ≤64.

### Post Hoc Analysis

In post hoc analysis, we performed stratified analysis by race showing the c-statistic for the SRSRF was nonsignificantly higher than the FSRF within each race. Specifically, in whites the SRSRF c-statistic was 0.7568 (95% CI, 0.7323–0.7813) compared with the FSRF of 0.7440 (95% CI, 0.7189–0.7691; $P=0.065$); whereas in blacks, the corresponding c-statistics were 0.6757 (95% CI, 0.6443–0.7071) versus 0.6601 (95% CI, 0.6290–0.6911; $P=0.10$).

### Discussion

These data suggest that a simple survey, constructed to include 2 questions that exclude those with prevalent stroke or transient ischemic attacks plus 11 questions to assess stroke risk, can identify a stroke-free population at higher risk for incident stroke with a reliability better than the gold standard FSRF. In addition to providing better prediction, this approach avoids the need to see someone in person for direct measurement of blood pressure, venipuncture for glucose levels, and performing an ECG for assessment of atrial fibrillation. In the REGARDS population, the annual stroke risk of those in the top decile was 1.4%—a remarkably high stroke risk group for an asymptomatic population, and one that calls for primary prevention trials to identify potential interventions to reduce the burden of stroke in this easily identified population. Importantly, this survey can be (and actually was) conducted over the telephone using a script that can be executed by nonmedical staff, making the screening of thousands of potential participants for a stroke primary prevention trial both feasible and efficient.

In the process of performing this analysis, the critical role of the impact of race became apparent. Blacks between the ages of 45 and 64 have 2 to 3 times the risk of stroke of their white counterparts.[16] With few blacks in Framingham, the FSRF could not model the impact of this important predictor, and as such several the high-risk REGARDS participants are blacks with low FSRF risk factor scores. That the FSRF fails to capture the risk differences between blacks and whites, which was modeled in the SRSRF, was the reason for the superior predictive value of the SRSRF. However, in the post
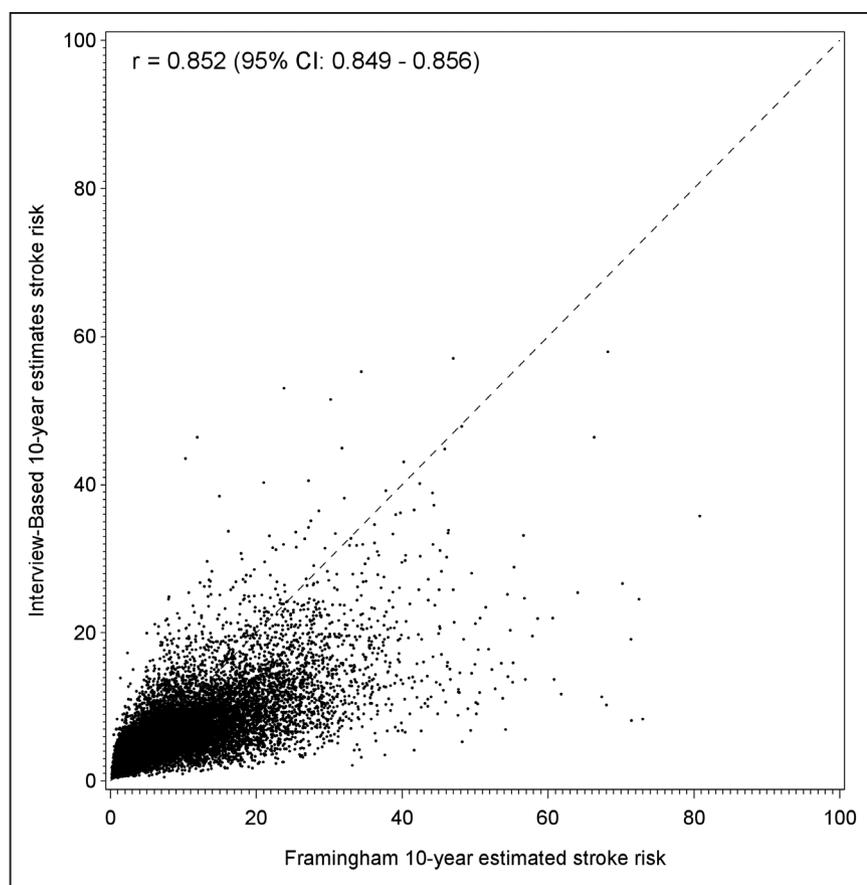
**Figure 1.** Scattergram of the 10-year stroke risk from the Framingham Stroke Risk Function vs the estimated 10-year stroke risk from the Self-Reported Stroke Risk Function. CI indicates confidence interval.

hoc stratified analysis by race, the SRSRF was only marginally better than the FSRF, an achievement which is in itself is remarkable performance for an index that requires no direct contact with the participant. Importantly, in the identification



**Figure 2.** Predicted and observed stroke risks by decile of expected risk from the Framingham Stroke Risk Function (**A**) and from the Self-Reported Stroke Risk Function (**B**).

of populations at high risk, the approach is not to select high-risk blacks or high-risk whites but rather high-risk participants. This post hoc analysis also revealed a better ability to discriminate (ie, higher c-statistic) stroke risk in whites than blacks. This not only supports the use of the SRSRF as a predictor of risk but also underscores the need for a subsequent article from REGARDS (or other populations with substantial representation of black) that directly incorporates race into the predictive model that considers directly measured risk factors.

We have previously shown that stroke symptoms are a powerful predictor for subsequent stroke risk[14]; however, those participants reporting stroke symptoms also tended to report lower self-reported general health and have a higher prevalence of risk factors, and as such stroke symptoms added little to the multivariable prediction of stroke risk.

We anticipated a substantial correlation between the SRSRF and FSRF because they share 3 identically defined factors: age, sex, and cigarette smoking. In addition, the relatively high agreement for the 4 factors measured by self-report in the SRSRF and a combination of self-report and direct assessment in the FSRF (hypertension, diabetes mellitus, atrial fibrillation, and heart disease) also contribute to the correlation between the 2. However, we did not expect the association to be strong, because (1) we did not anticipate the agreement between the self-reported and direct measurement of hypertension, diabetes mellitus, atrial fibrillation, and heart disease to be as high as we observed, and (2) we assumed the inclusion of education and self-reported health in the SRSRF (not in the FSRF) would reduce agreement. The high correlation between these
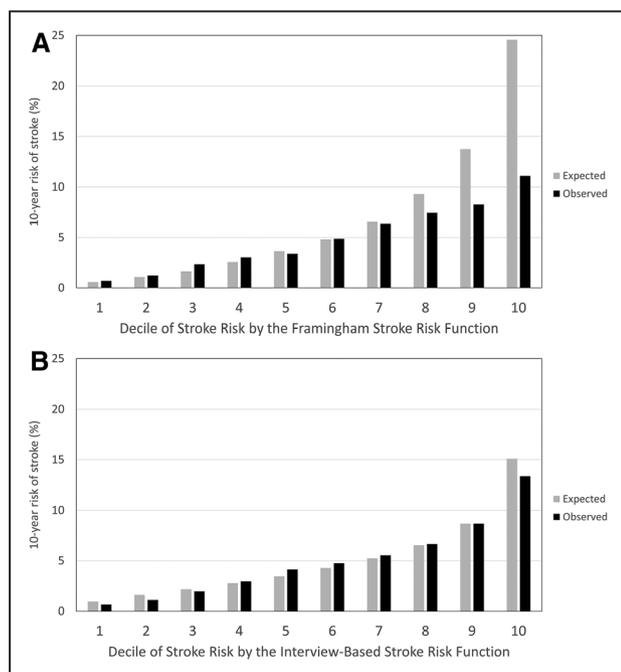
**Table 3.   Description of Self-Reported Characteristics Within Decile of SRSRF Predicted Risk**

|  | Decile of SRSRF Predicted Stroke Risk | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Age, mean±SD | 52±4 | 57±4 | 59±5 | 61±5 | 63±6 | 65±6 | 67±6 | 70±7 | 73±7 | 77±7 |
| Male sex, % | 28 | 37 | 39 | 44 | 43 | 47 | 44 | 51 | 54 | 65 |
| Black race, % | 7 | 26 | 35 | 42 | 47 | 52 | 54 | 52 | 47 | 39 |
| Self-reported hypertension, % | 16 | 33 | 44 | 50 | 57 | 63 | 69 | 72 | 75 | 79 |
| Self-reported diabetes mellitus, % | 3 | 5 | 10 | 13 | 17 | 23 | 27 | 33 | 37 | 46 |
| Self-reported atrial fibrillation, % | 1 | 2 | 4 | 4 | 6 | 5 | 7 | 9 | 11 | 22 |
| Self-reported MI, % | 0 | 1 | 1 | 2 | 3 | 4 | 6 | 9 | 14 | 31 |
| Current smoking, % | 5 | 7 | 11 | 12 | 14 | 17 | 16 | 18 | 20 | 21 |
| High school or less education, % | 13 | 18 | 24 | 30 | 38 | 39 | 47 | 50 | 54 | 58 |
| Poor/fair self-reported health, % | 3 | 6 | 8 | 10 | 15 | 16 | 19 | 22 | 25 | 35 |

MI indicates myocardial infarction; and SRSRF, Self-Reported Stroke Risk Function.

risk scores (0.852) implies that nearly three fourths (the $r^2$ would be $0.852^2=0.726$) of the information in the FSRF can be captured by the simple 11 question self-reported survey.

The discrimination of the FSRF remains impressive with a 15.8× (11.1% / 0.7%) higher observed stroke risk in the top than bottom decile of estimated stroke risk, but the SRSRF had a more outstanding discrimination with a 21.6× (15.1% / 0.7%) higher observed stroke risk in the top relative to the bottom decile of stroke risk. The SRSRF having been developed and assessed in the same data set (whereas the FSRF was developed in a different data set than this assessment) could contribute to both the higher discrimination and better calibration; however, with nearly 1000 stroke events in this analysis, the inflation of the discrimination should be minor. The discrimination and calibration of the SRSRF should be confirmed by others in independent populations.

We acknowledge other stroke risk prediction models including those developed in the ARIC study (Atherosclerosis Risk in Communities),[17] the CHS (Cardiovascular Health Study),[18] QStroke risk function developed from the QResearch database, including data from 451 general practices in England and Wales,[19] and the health-behavior-based SPoRT risk function.[20] Other stroke risk functions have been developed for subgroups of the population, such as those with atrial fibrillation.[21–23] However, in this work, we focused on the comparison of the SRSRF to the FSRF both because of (1) the preeminence of the FSRF, (2) recognition that many of the general risk functions contain factors not assessed in REGARDS (eg, the timed walk in CHS or rheumatoid arthritis in QSTROKE), (3) limitations in the age range of the ARIC and CHS models (45–65 and 65+ at baseline, respectively) making them less comparable to REGARDS (ages 45+ at baseline), and (4) the inclusion of a broad general population in the Framingham cohort.

We were also concerned that the individuals in the highest decile of SRSRF-predicted risk would be remarkably nonrepresentative, for example, with extremely old ages or predominately black, or with a unexpected constellations of risk factors. However, the median age in the highest decile was only 77 years, only 39% were black, and the prevalence

of risk factors ranged from 22% with atrial fibrillation to 79% with hypertension. As such, a primary prevention trial done in such a population would provide generalizable information to a broad population.

In a closely related topic, we note that the proportion of blacks declines above the seventh decile of SRSRF risk. This is likely because blacks have a lower prevalence of atrial fibrillation and heart disease, two of the more powerful risk factors for stroke, and to have a high probability of stroke on the SRSRF a participant would have to be more likely to have atrial fibrillation or heart disease.

The major strength of this report is the use a large biracial cohort drawn from across the continental United States. This cohort provided a systematic direct measure of risk factors, adjudicated stroke outcomes, and had a sample size sufficient to give rise to a large number of stroke events resulting in stable estimates. The major limitation is that REGARDS included only black and white participants, so it is not clear whether these findings can be generalized to other race-ethnic groups. REGARDS may also not be completely representative of the general population or specifically to the population from other countries.

In conclusion, a simple self-reported questionnaire can be used to stratify stroke risk in a general adult population better than the gold standard FSRF. Much of this superior performance arises from the explicit incorporation of race as a predictor, and when this is taken into account, the performances of the SRSRF and FSRF are similar. However, if the goal is to find high-risk patients/subjects, then the SRSRF has the advantage of explicitly including race. This instrument can be used clinically to easily identify individuals at high risk for stroke and also scientifically as part of a primary stroke prevention trial by identifying a subpopulation enriched for stroke risk.

## Acknowledgments

## Sources of Funding

## Disclosures

None.

## References

1. Wolf PA, D'Agostino RB, Belanger AJ, Kannel WB. Probability of stroke: a risk profile from the Framingham Study. *Stroke*. 1991;22:312–318.
2. Dufouil C, Beiser A, McLure LA, Wolf PA, Tzourio C, Howard VJ, et al. Revised Framingham Stroke Risk Profile to reflect temporal trends. *Circulation*. 2017;135:1145–1159. doi: 10.1161/CIRCULATIONAHA.115.021275.
3. D'Agostino RB Sr, Grundy S, Sullivan LM, Wilson P; CHD Risk Prediction Group. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *JAMA*. 2001;286:180–187.
4. Mainous AG 3rd, Koopman RJ, Diaz VA, Everett CJ, Wilson PW, Tilley BC. A coronary heart disease risk score based on patient-reported information. *Am J Cardiol*. 2007;99:1236–1241. doi: 10.1016/j.amjcard.2006.12.035.
5. Gaziano TA, Young CR, Fitzmaurice G, Atwood S, Gaziano JM. Laboratory-based versus non-laboratory-based method for assessment of cardiovascular disease risk: the NHANES I Follow-up Study cohort. *Lancet*. 2008;371:923–931. doi: 10.1016/S0140-6736(08)60418-3.
6. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97:1837–1847.
7. Pandya A, Weinstein MC, Gaziano TA. A comparative assessment of non-laboratory-based versus commonly used laboratory-based cardiovascular disease risk scores in the NHANES III population. *PLoS One*. 2011;6:e20416. doi: 10.1371/journal.pone.0020416.
8. Pandya A, Weinstein MC, Salomon JA, Cutler D, Gaziano TA. Who needs laboratories and who needs statins?: comparative and cost-effectiveness analyses of non-laboratory-based, laboratory-based, and staged primary cardiovascular disease screening guidelines. *Circ Cardiovasc Qual Outcomes*. 2014;7:25–32. doi: 10.1161/CIRCOUTCOMES.113.000397.
9. Gaziano TA, Pandya A, Steyn K, Levitt N, Mollentze W, Joubert G, et al. Comparative assessment of absolute cardiovascular disease risk characterization from non-laboratory-based risk assessment in South African populations. *BMC Med*. 2013;11:170. doi: 10.1186/1741-7015-11-170.
10. Schneider KL, Clark MA, Rakowski W, Lapane KL. Evaluating the impact of non-response bias in the Behavioral Risk Factor Surveillance System (BRFSS). *J Epidemiol Community Health*. 2012;66:290–295. doi: 10.1136/jech.2009.103861.
11. Howard VJ, Cushman M, Pulley L, Gomez CR, Go RC, Prineas RJ, et al. The reasons for geographic and racial differences in stroke study: objectives and design. *Neuroepidemiology*. 2005;25:135–143. doi: 10.1159/000086678.
12. Avendano M, Kawachi I, Van Lenthe F, Boshuizen HC, Mackenbach JP, Van den Bos GA, et al. Socioeconomic status and stroke incidence in the US elderly: the role of risk factors in the EPESE study. *Stroke*. 2006;37:1368–1373. doi: 10.1161/01.STR.0000221702.75002.66.
13. Benjamins MR, Hummer RA, Eberstein IW, Nam CB. Self-reported health and adult mortality risk: an analysis of cause-specific mortality. *Soc Sci Med*. 2004;59:1297–1306. doi: 10.1016/j.socscimed.2003.01.001.
14. Kleindorfer D, Judd S, Howard VJ, McClure L, Safford MM, Cushman M, et al. Self-reported stroke symptoms without a prior diagnosis of stroke or transient ischemic attack: a powerful new risk factor for stroke. *Stroke*. 2011;42:3122–3126. doi: 10.1161/STROKEAHA.110.612937.
15. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med*. 2004;23:2109–2123. doi: 10.1002/sim.1802.
16. Howard VJ, Kleindorfer DO, Judd SE, McClure LA, Safford MM, Rhodes JD, et al. Disparities in stroke incidence contributing to disparities in stroke mortality. *Ann Neurol*. 2011;69:619–627. doi: 10.1002/ana.22385.
17. Chambless LE, Heiss G, Shahar E, Earp MJ, Toole J. Prediction of ischemic stroke risk in the Atherosclerosis Risk in Communities Study. *Am J Epidemiol*. 2004;160:259–269. doi: 10.1093/aje/kwh189.
18. Manolio TA, Kronmal RA, Burke GL, O'Leary DH, Price TR. Short-term predictors of incident stroke in older adults. The Cardiovascular Health Study. *Stroke*. 1996;27:1479–1486.
19. Hippisley-Cox J, Coupland C, Brindle P. Derivation and validation of QStroke score for predicting risk of ischaemic stroke in primary care and comparison with other risk scores: a prospective open cohort study. *BMJ*. 2013;346:f2573.
20. Manuel DG, Tuna M, Perez R, Tanuseputro P, Hennessy D, Bennett C, et al. Predicting stroke risk based on health behaviours: development of the Stroke Population Risk Tool (SPoRT). *PLoS One*. 2015;10:e0143342. doi: 10.1371/journal.pone.0143342.
21. Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. *JAMA*. 2001;285:2864–2870.
22. European Heart Rhythm A, European Association for Cardio-Thoracic S, Camm AJ, Kirchhof P, Lip GY, Schotten U, et al. Guidelines for the management of atrial fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC). *Europace*. 2010;12:1360–1420.
23. Marcucci M, Lip GY, Nieuwlaat R, Pisters R, Crijns HJ, Iorio A. Stroke and bleeding risk co-distribution in real-world patients with atrial fibrillation: the Euro Heart Survey. *Am J Med*. 2014;127:979–986.e2. doi: 10.1016/j.amjmed.2014.05.003.

# Stroke

American Heart Association | American Stroke Association®

**Self-Reported Stroke Risk Stratification: Reasons for Geographic and Racial Differences in Stroke Study**

George Howard, Leslie A. McClure, Claudia S. Moy, Virginia J. Howard, Suzanne E. Judd, Ya Yuan, D. Leann Long, Paul Muntner, Monika M. Safford and Dawn O. Kleindorfer

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://stroke.ahajournals.org/content/48/7/1737

Data Supplement (unedited) at:
http://stroke.ahajournals.org/content/suppl/2017/05/19/STROKEAHA.117.016757.DC1

**SUPPLEMENTAL MATERIAL**

**Interview for assessing stroke risk**

1. The first set of questions asks about whether you have had a stroke or a mini-stroke.  Were you ever told by a physician that you had a stroke?
   1. Yes
   2. No
   8. Don't Know/Not Sure
   9. Refused

2. Were you ever told by a physician that you had a mini-stroke or TIA, also known as a transient ischemic attack?
   1. Yes
   2. No
   8. Don't Know/Not Sure
   9. Refused

3. What is your age?

4. What is your race
   1. White
   2. Black

5. What is your sex
   1. Male
   2. Female

6. Has a doctor or other health professional ever told you that you have high blood pressure?
   1. Yes
   2. No
   8. Don't Know/Not Sure
   9. Refused

7. Has a doctor or other health professional ever told you that you had diabetes or high blood sugar?
   1. Yes
   2. No
   8. Don't Know/Not Sure
   9. Refused

8. Has a doctor or  other health professional ever told you that you had a myocardial infarction or heart attack?
   1. Yes
   2. No
   8. Don't Know/Not Sure
   9. Refused

9. Has a doctor or  other health professional ever told you that you had a myocardial infarction or heart attack?
   1. Yes
   2. No
   8. Don't Know/Not Sure
   9. Refused

10. Have you smoked at least 100 cigarettes in your lifetime?
    1. Yes
    2. No
    8. Don't Know/Not Sure
    9. Refused

11. Do you smoke cigarettes now, even occasionally?
    1. Yes
    2. No
    8. Don't Know/Not Sure
    9. Refused

12. What is the highest grade or year of school you have completed?    (READ ONLY IF NECESSARY)
    11    Never attended or kindergarten only
    12    Eighth grade or less
    13    Some high school (9th - 11th grade)
    14    High school graduate or GED certificate
    15    Some technical school
    16    Technical school graduate
    17    Some college
    18    College graduate
    19    Postgraduate or professional degree
    88    Don't Know/Not Sure
    99    Refused

13. In general, would you say that your health is excellent, very good, good, fair or poor?
    1. Excellent
    2. Very good
    3. Good
    4. Fair
    5. Poor
    8. Don't Know/Not Sure
    9. Refused

**Calculation of Risk Level using the Self-Reported Stroke Risk Function (SRSRF)**

Using REGARDS data individuals can be placed into strata defined by the deciles of the SRSRF.  First, the multivariable (X'B) should be calculated from the factors in the SRSRF:

SRSRF = 0.078687*AGE
+ 2.13064 (if black)
-0.029208 *AGE (if black)
-0.28036 (if female)
+ 0.33039 (if self-reported atrial fibrillation)
+ 0.31412 (if self-reported diabetic)
+ 0.24689 (if self-reported hypertension)
+ 0.40563 (if self-reported MI)
+ 0.51806 (if current smoker)
+ 0.12186 (if education level = "some college")
+ 0.31384 (if education level = "high school graduate")
+ 0.13703 (if education level = "less than high school")
-0.07226 (if general self-reported health = "Very good")
+ 0.09785 (if general self-reported health = "Good")
+ 0.18702 (if general self-reported health = "Fair")
+0.63835 (if general self-reported health = "Poor")

Individual can then be placed into deciles of risk from the resulting score, and the associated predicted risk determined:

| Decile of risk | If the resulting score (X'B) is between: | | Then the predicted 10-year Stroke Risk is approximately (%) |
| :---: | :---: | :---: | :---: |
| | Low | High | |
| 1 | 3.19 | 4.47 | 0.7 |
| 2 | 4.47 | 4.84 | 1.1 |
| 3 | 4.84 | 5.11 | 2.0 |
| 4 | 5.11 | 5.34 | 3.0 |
| 5 | 5.34 | 5.55 | 4.1 |
| 6 | 5.55 | 5.76 | 4.7 |
| 7 | 5.76 | 5.97 | 5.5 |
| 8 | 5.97 | 6.23 | 6.7 |
| 9 | 6.23 | 6.56 | 8.7 |
| 10 | 6.56 | 8.65 | 13.4 |

Should an individual have a score less than 3.19, or greater than 8.68, then the results should be interpreted with caution; however, it is likely safe to say that the risk is similar to the lowest or highest decile of risk respectively.