SUPPLEMENTAL MATERIAL
**Data S1. Supplemental Methods**

**Measurement and Harmonization of Risk Factors**

We defined active malignancy as a malignancy diagnosed within the year prior to the stroke or a previously diagnosed malignancy not in remission. Among controls, active malignancy was defined as a malignancy diagnosed or treated within the last year based on the questionnaire.

Atrial fibrillation (AF) was defined as AF or an atrial flutter diagnosis, while cardiovascular disease (CVD) was defined as any of the following in the hospital discharge register before IS or the health examination date: angina pectoris, myocardial infarction, coronary revascularization, heart failure, or peripheral arterial disease.

Current smoking status among cases was defined as smoking ≥1 cigarettes per day within the year prior to a stroke. Among controls, we defined smoking through responses to the questionnaire as currently smoking or having quit smoking less than a year prior to participation in FINRISK.

Diabetes mellitus was defined as treated diabetes or a history of diabetes preceding IS or the health examination. Among cases, diabetes was defined according to the 1999 WHO criteria as a fasting plasma glucose ≥7 mmol/L (126 mg/dL) or a 2-hour oral glucose tolerance test of ≥11.1 mmol/L (200 mg/dL) using the available data from medical records. We distinguished between diabetes mellitus type 1 (T1D) and 2 (T2D) for cases based on medical records. Those with an onset of diabetes before the age of 40 and initiating insulin treatment within one year of diagnosis were classified as T1D. Among controls, diabetes was defined through a self-reported diagnosis collected from the FINRISK questionnaire, a diabetes diagnosis appearing in the electronic registers, a prescribed diabetes medication, or entitlement to diabetes medication reimbursement. T1D was distinguished from T2D through an initial purchase of insulin before age 35 or a T1D diagnosis code in the electronic registers.

We defined dichotomous components of dyslipidemia as high low-density lipoprotein cholesterol (LDL-C) of ≥3.0 mmol/L or 116 mg/dL, low high-density lipoprotein cholesterol (HDL-C) of <1.0 mmol/L or 39 mg/dL, and high triglycerides of ≥2.0 mmol/L or 177 mg/dL. An alternative combination variable dyslipidemia was defined as treated dyslipidemia, history of dyslipidemia, or any of the following: high LDL-C of ≥3.0 mmol/L or 116 mg/dL, low HDL-C of <1.0 mmol/L or 39 mg/dL, or high total cholesterol of ≥5.0 mmol/L or 193 mg/dL. Among cases to measure lipid levels, fasting blood samples were drawn on the first working day after a stroke and analyzed in the hospital’s laboratory. Among controls, blood samples were drawn after ≥4-hour fast at the health examination visit and analyzed in the laboratory of the National Institute of Health and Welfare. During the study period, both laboratories analyzed total cholesterol and triglycerides using enzymatic methods, first analyzing HDL-C cholesterol through the use of Dextran-MgCl2 precipitation and later using enzymatic methods. LDL-C was calculated using the Friedewald equation. If the Friedewald equation could not be used due to a high triglyceride value (>4.5mmol/L or 389 mg/dL), the binary LDL-C variable was coded as high.

We defined a family history of stroke as a history of any stroke or transient ischemic attack (TIA) in a first-degree relative among cases. Among controls, we defined a family history of stroke as any stroke occurring among a first-degree relative under the age of 75 using the questionnaire.
Hypertension was defined as taking an antihypertensive medication, a previous hypertension diagnosis, or current hypertension according to the 2003 World Health Organization (WHO) criteria as a systolic blood pressure (SBP) ≥140 mmHg or a diastolic blood pressure (DBP) ≥90 mmHg. A current hypertension diagnosis for cases was based on several hypertensive values taken on the ward after the admission; among controls, current hypertension was diagnosed during the health examination. For the alternative analysis of blood pressure as a continuous variable, the mean SBP and DBP measurements at admission and at 24 hours was used for cases (if only one measurement was available, we used either measurement). Among controls, BP was measured three times during the health examination, and the mean from the last two measurements was used.
Additional Analyses

To exclude multicollinearity, variance inflation factors (VIF) using a linear regression analysis and Cramer’s V measures of association were calculated. Variables were also individually deleted and the forced logistic regression model was repeated, whereby we observed no significant change in the regression coefficients.

All of the studied risk factors were statistically significant in univariate analysis, and thus all 11 dichotomous risk factors were included in the multivariable regression analyses. We used a backward stepwise binary multivariable regression analysis to calculate adjusted odds ratios and to determine which of the risk factors remained statistically significant after adjusting. To confirm our primary results, additional multivariable analyses were performed using a forward stepwise binary logistic regression analysis and a binary logistic regression analysis with all of the variables forced into the model. The same risk factors appeared statistically significant using these models and the strengths of the associations were practically unchanged from those achieved using the backward stepwise binary logistic regression model (data not shown).

Additionally, we ran alternative fully-adjusted multivariable analyses using (1) the prevalence of any diabetes mellitus instead of the specific diabetes types, (2) the combined variable dyslipidemia instead of the dichotomized lipid components, and (3) systolic blood pressure (SBP) and diastolic blood pressure (DBP) (per 10 mmHg) and lipid components [logarithmic transformation per 1 standard deviation (SD) increment] as continuous parameters instead of corresponding dichotomized variables. Model (3) with the continuous BP and lipid parameters was adjusted for antihypertensive and lipid-lowering medication. All continuous risk factor variables were confirmed as normally distributed.

To test the association between risk factors and sex and risk factors and age group, we have conducted chi-square test of homogeneity, where we observed significant association between sex and six of the risk factors (smoking, AF, hypertension, low HDL-C, high LDL-C, and high triglycerides) and significant association between age group (25-39 versus 40-49) and all of the 11 risk factors. To test the association between risk factors and stroke etiology, we have conducted likelihood ratio test, where all 11 risk factors appeared significantly associated with stroke etiology.
Supplemental Results

Missing Data Analysis

The amount of missing data in our study was low: values for any dichotomous variables were missing from 2.3% of subjects (43/4.5% cases and 11/0.8% controls). Data from lipid measurements were missing for 46/1.9% subjects (43/4.5% cases, 3/0.2% controls), active malignancy from 7 controls (0.5%), and hypertension from 1 (0.1%) control. Continuous blood pressure values were missing from 1.2% of subjects (28/2.9% cases, 1/0.1% controls) and lipid measurements from 1.9% of subjects (43/4.5% cases, 3/0.2% controls).

Comparing cases with complete data to those with incomplete data (Table S3), we observed that those with incomplete dichotomous variables were slightly more often female and younger. These differences were, however, not statistically significant. In addition, cases with missing dichotomous data were, on the average, diseased earlier (stroke occurrence from study start after a mean of 5 years versus 7.2 years), and had a lower Glasgow coma scale (GCS) score on admission (mean 13.5 versus 14.6 points). The differences between the groups regarding stroke year and GCS score reached statistical significance using the nonparametric Mann-Whitney U test.

When we compared the prevalence of all risk factors among subjects with and without missing data for dichotomous variables (Table S4), we observed that atrial fibrillation and active malignancy were slightly more prevalent and family history of stroke less prevalent among cases with missing values. However, when we compared the groups using Fisher’s Exact Test, no statistically significant associations appeared.
Alternative Dichotomized Risk Factors and Continuous Variables

The prevalence of any diabetes was 9.2% (88/961) among cases and 2.4% (33/1403) among controls. In the alternative multivariable models adjusted for age, sex, and all other risk factors, the multivariable adjusted OR for any type of diabetes was 3.46 (95% CI 2.24–5.34), while PAR% was 6.5% (95% CI 5.3–7.5). The prevalence of dyslipidemia reached 63.1% (603/956) among cases and 68.9% (964/1400) among controls. As such, dyslipidemia inversely associated with the risk of IS (OR 0.55, 95% CI 0.45–0.67).

Higher mean SBP was associated with an increased risk of IS (multivariable OR 1.46 per 10 mmHg, 95% CI 1.38–1.55), whereas we found no significant association for higher mean DBP (1.08, 95% CI 0.95–1.22). The associations for logarithmic lipid components were consistent with the dichotomous variables. That is, we found that higher HDL-C (OR 0.79 per 1 SD increment, 95% CI 0.71–0.87) and LDL-C (0.58, 95% CI 0.52–0.65) were inversely associated with the risk of IS, while we observed no significant association for triglycerides (0.99, 95% CI 0.88–1.11).
Sensitivity Analyses

To achieve comparable covariates, we have defined AF and CVD both for cases and controls as respective diagnoses in the hospital discharge register before the IS or the health examination date. The national hospital discharge register Care Register for Health Care contains information on hospital discharges since 1967, on operations performed since 1987, and on specialized outpatient health care provided since 1998. To assess the sensitivity and specificity of this national hospital discharge register regarding vascular diseases, data from controls were compared to separate register data (MACE-PAD) containing information for all major cardiovascular events, invasive procedures, heart failure, and peripheral vascular events based on hospital discharge diagnoses and specialized outpatient health care. Identical control subjects with a history of cardiovascular disease were identified in both registers. The data collected for cases were compared to risk factors collected from their medical records. As we expected, the sensitivity of the hospital discharge register was highest in the identification of coronary heart disease and atrial fibrillation (82% and 78%, respectively) and lowest regarding the identification of heart failure (34%) and peripheral arterial disease (33%).

To evaluate the possible effects of LDL-C coding choice, analyses were repeated using an alternative binary LDL-C classification for those with unanalyzable LDL-C due to high triglycerides of >4.5 mmol/L or 398 mg/dL (cases n=17/1.8%, controls n=28/2.0%). In the sensitivity analyses, unanalyzable binary LDL-C was alternatively classified as (1) for low LDL-C or (2) for missing value on LDL-C, and the multivariable analyses were repeated. The results of these additional analyses were virtually identical from the results of our primary analysis. We also tested whether the results regarding LDL-C would differ by applying an alternate threshold level for high LDL-C \( \geq 4.0 \) mmol/L. Using the higher threshold level, the binary high LDL-C variable remained inversely associated with IS.
Table S1. Frequencies of Cases and Controls According to Sex and Age Category.

| Age category | Women |        | Men  |        | Total |
|--------------|-------|--------|------|--------|-------|
|              | Cases | Controls | Cases | Controls |       |
| 25−39 years  | 125   | 295     | 142  | 325     | 887   |
| 40−49 years  | 233   | 315     | 461  | 468     | 1477  |
| Total        | 358   | 610     | 603  | 793     | 2364  |
Table S2. Proportion of Cases According to Modified Trial of Org 10172 in Acute Stroke Treatment (TOAST) Classification.

| Etiologic subgroup      | All cases (n=961) | Cases with complete data (n=918) |
|-------------------------|-------------------|----------------------------------|
|                         | No. (%)           | No. (%)                          |
| LAA                     | 76 (7.9)          | 73 (8.0)                         |
| CE high-risk            | 95 (9.9)          | 89 (9.7)                         |
| SVO                     | 138 (14.4)        | 134 (14.6)                       |
| Dissection              | 153 (15.9)        | 150 (16.3)                       |
| Other determined etiologies | 96 (10.0)      | 90 (9.8)                         |
| ESUS                    | 190 (19.8)        | 187 (20.4)                       |
| UE non-ESUS             | 213 (22.2)        | 195 (21.2)                       |

LAA indicates large artery atherosclerosis; CE high-risk, cardioembolism from high-risk source; SVO, small vessel occlusion; ESUS: embolic stroke of an undetermined source; UE non-ESUS, undetermined etiology non-ESUS.
Table S3. Comparison of Cases with Complete and Incomplete Data.

|                  | All cases (n=961) | Dichotomous variables | Continuous variables |
|------------------|-------------------|-----------------------|----------------------|
|                  | Cases with complete data (n=918)* | Cases with incomplete data (n=43)* | Cases with complete data (n=877)† | Cases with incomplete data (n=84)† |
| Women No. (%)    | 358 (37.2)        | 339 (36.9)            | 19 (44.2)            | 322 (36.7)        |
|                  |                   | 36 (42.9)             |                      |                     |
| Age‡, median (Range), y | 44.0 (24)       | 44.0 (24)             | 43.0 (22)            | 44.0 (24)         |
| NIHSS on admission‡ median (Range) | 3.0 (35)         | 3.0 (35)              | 3.0 (22)             | 3.0 (22)           |
|                  |                   |                      | 3.0 (35)             |                     |
| Years from study start to stroke‡ median (Range) | 7.0 (13)         | 7.0 (13)§             | 4.0 (13)§            | 7.0 (13)§         |
|                  |                   |                      | 5.0 (13)§            |                     |
| GCS on admission‡ median (Range) | 15.0 (12)        | 15.0 (12)§            | 15.0 (12)§           | 15.0 (12)§        |

NIHSS indicates National Institutes of Health Stroke Scale; GCS, Glasgow coma scale. *Unanalyzable low-density lipoprotein cholesterol (LDL-C) value due to high triglycerides is coded as a high LDL-C and is not considered as a missing dichotomous value. †Unanalyzable LDL-C value due to high triglycerides is considered as a missing continuous value. ‡Variable is not normally distributed. §Statistically significant differences according to nonparametric Mann-Whitney U Test between cases with complete and incomplete set of variables.
Table S4. Comparison of Risk Factors in Subjects with Complete and Incomplete Dichotomous Data.

| Risk Factor                        | Cases with complete data* (n=918) | Cases with incomplete data* (n=43) | Controls with complete data* (n=1392) | Controls with incomplete data* (n=11) |
|------------------------------------|-----------------------------------|-----------------------------------|--------------------------------------|-------------------------------------|
| Atrial fibrillation                | 20/918 (2.2)                      | 3/43 (7.0)                        | 2/1392 (0.1)                        | 0/11 (0.0)                         |
| Active malignancy                 | 12/918 (1.4)                      | 2/43 (4.7)                        | 6/1392 (0.4)                        | 0/4 (0.0)                          |
| Cardiovascular disease†           | 43/918 (4.7)                      | 3/43 (7.0)                        | 5/1392 (0.4)                        | 0/11 (0.0)                         |
| Current smoking status            | 412/918 (44.9)                    | 15/43 (34.9)                      | 432/1392 (31.0)                     | 3/11 (27.3)                        |
| Family history of stroke          | 123/918 (13.4)                    | 3/43 (7.0)                        | 128/1392 (9.2)                      | 0/11 (0.0)                         |
| High LDL-C‡                       | 472/918 (51.4)                    | NA                                | 846/1392 (60.8)                     | 4/8 (50.0)                         |
| High triglycerides‡               | 209/918 (22.8)                    | NA                                | 215/1392 (15.4)                     | 2/8 (25.0)                         |
| Hypertension                      | 371/918 (40.4)                    | 20/43 (46.5)                      | 394/1392 (28.3)                     | 0/10 (0.0)                         |
| Low HDL-C‡                        | 147/918 (16.0)                    | NA                                | 112/1392 (8.0)                      | 2/8 (25.0)                         |
| Type 1 diabetes                   | 41/918 (4.5)                      | 3/43 (7.0)                        | 9/1392 (0.6)                        | 0/11 (0.0)                         |
| Type 2 diabetes                   | 40/918 (4.4)                      | 4/43 (9.3)                        | 24/1392 (1.7)                       | 0/11 (0.0)                         |

NA indicates not applicable. *Unanalyzable LDL-C value due to high triglycerides is coded as a high LDL-C and not considered as a missing dichotomous value. †The presence of coronary heart disease, heart failure, or peripheral arterial disease. ‡High low-density lipoprotein cholesterol (LDL-C) ≥3.0 mmol/L (116 mg/dL), low high-density lipoprotein cholesterol (HDL-C) <1.0 mmol/L (39 mg/dL), and high triglycerides ≥2.0 mmol/L (177mg/dL).
Supplemental References:

1. Putaala J, Liebkind R, Gordin D, Thorn LM, Haapaniemi E, Forsblom C, Groop PH, Kaste M, Tatlisumak T. Diabetes mellitus and ischemic stroke in the young: clinical features and long-term prognosis. *Neurology*. 2011;76:1831-1837.

2. Sund R. Quality of the Finnish Hospital Discharge Register: a systematic review. *Scand J Public Health*. 2012;40:505-515.