

## Risk Factors for Early-Onset Ischemic Stroke: A Case-Control Study

Reetta Kivioja, MD; Arto Pietilä, MSc; Nicolas Martinez-Majander, MD; Daniel Gordin, MD, PhD; Aki S. Havulinna, DSc (tech); Veikko Salomaa, MD, PhD; Karoliina Aarnio, MD; Sami Curtze, MD, PhD; Jaana Leiviskä, PhD; Jorge Rodríguez-Pardo, MD; Ida Surakka, PhD; Markku Kaste, MD, PhD; Turgut Tatlisumak, MD, PhD; Jukka Putaala, MD, PhD

**Background**—Recent studies have shown an increasing prevalence of vascular risk factors in young adults with ischemic stroke (IS). However, the strength of the association between all vascular risk factors and early-onset IS has not been fully established.

**Methods and Results**—We compared 961 patients with a first-ever IS at 25 to 49 years to 1403 frequency-matched stroke-free controls from a population-based cohort study (FINRISK). Assessed risk factors included an active malignancy, atrial fibrillation, cardiovascular disease, current smoking status, a family history of stroke, high low-density lipoprotein cholesterol, high triglycerides, low high-density lipoprotein cholesterol, hypertension, and type 1 and type 2 diabetes mellitus. We performed subgroup analyses based on age, sex, and IS etiology. In a fully adjusted multivariable logistic regression analysis, significant risk factors for IS consisted of atrial fibrillation (odds ratio [OR], 10.43; 95% confidence interval [CI], 2.33–46.77), cardiovascular disease (OR, 8.01; 95% CI, 3.09–20.78), type 1 diabetes mellitus (OR, 6.72; 95% CI, 3.15–14.33), type 2 diabetes mellitus (OR, 2.31; 95% CI, 1.35–3.95), low high-density lipoprotein cholesterol (OR, 1.81; 95% CI, 1.37–2.40), current smoking status (OR, 1.81; 95% CI, 1.50–2.17), hypertension (OR, 1.43; 95% CI, 1.17–1.75), and a family history of stroke (OR, 1.37; 95% CI, 1.04–1.82). High low-density lipoprotein cholesterol exhibited an inverse association with IS. In the subgroup analyses, the most consistent associations appeared for current smoking status and type 1 diabetes mellitus.

**Conclusions**—Our study establishes the associations between 11 vascular risk factors and early-onset IS, among which atrial fibrillation, cardiovascular disease, and both type 1 and 2 diabetes mellitus in particular showed strong associations. (*J Am Heart Assoc.* 2018;7:e009774. DOI: 10.1161/JAHA.118.009774.)

**Key Words:** brain infarction • middle-aged • risk factors • stroke • young adult

Ischemic stroke (IS) causes a remarkable loss of quality-adjusted life years. The socioeconomic burden of IS among young adults is particularly severe because affected patients tend to expect a long life after becoming ill. The cutoff age most often used to define early-onset IS or IS in young adults stands at less than 50 years.<sup>1</sup>

Recent studies have reported an increasing incidence of IS among young adults, contrary to decreasing IS incidence among older populations.<sup>2–6</sup> This observation may result from vascular risk factors becoming more prevalent among young individuals in general and accumulating in certain high-risk individuals who subsequently experience IS.<sup>4–9</sup> Determining

From the Clinical Neurosciences, University of Helsinki and Department of Neurology, Helsinki University Hospital, Helsinki, Finland (R.K., N.M.-M., K.A., S.C., M.K., T.T., J.P.); National Institute for Health and Welfare, Helsinki, Finland (A.P., A.S.H., V.S., I.S.); Abdominal Center Nephrology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland (D.G.); Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland (D.G.); Joslin Diabetes Center, Harvard Medical School, Boston, MA (D.G.); Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland (A.S.H., I.S.); Department of Clinical Chemistry, University of Helsinki and Helsinki University Hospital, HUSLAB, Helsinki, Finland (J.L.); Department of Neurology, La Paz University Hospital, Madrid, Spain (J.R.-P.); Department of Clinical Neurosciences, Institute of Neuroscience and Physiology, Sahlgrenska Academy at the University of Gothenburg, Sweden (T.T.); Department of Neurology, Sahlgrenska University Hospital, Gothenburg, Sweden (T.T.).

Accompanying Data S1, Tables S1 through S4 are available at <https://www.ahajournals.org/doi/suppl/10.1161/JAHA.118.009774>

A preliminary version of this work was presented at the European Stroke Organisation Conference, May 10 to 12, 2016, in Barcelona, Spain.

**Correspondence to:** Reetta Kivioja, MD, Department of Neurology, Helsinki University Hospital, Haartmaninkatu 4, FI-00029, Helsinki, Finland. E-mail: reetta.kivioja@helsinki.fi

Received May 10, 2018; accepted September 6, 2018.

© 2018 The Authors. Published on behalf of the American Heart Association, Inc., by Wiley. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

## Clinical Perspective

### What Is New?

- Atrial fibrillation, cardiovascular disease, and both type 1 and type 2 diabetes mellitus emerge as strong risk factors for early-onset IS.
- Risk profiles among the age, sex, and etiological-specific subgroups differ; however, smoking status and type 1 diabetes mellitus exhibit rather consistent associations.

### What Are the Clinical Implications?

- Thorough screening for cardiovascular risk factors and cessation of smoking appear to be important for the prevention of ischemic strokes in young population.

the strength of the association between well-documented vascular risk factors and early-onset IS would justify more intensive risk factor screening among this age group and provide the tools for individually tailored primary prevention. Yet, thus far, only a few case-control studies have assessed vascular risk factors in large, nonselected young IS patient populations, with a limited number of covariates and limited data on IS subtypes.<sup>10–15</sup>

Therefore, we aimed to determine the strength of the association among 11 vascular risk factors and early-onset IS for the Finnish population and stratify the analysis according to meaningful subgroups defined by age, sex, and stroke etiology.

## Methods

We compared patients with first-ever IS at age 25 to 49 years enrolled in the Helsinki Young Stroke Registry to frequency-matched (by sex and 5-year age bands) stroke-free controls from a population-based cohort study (the National FINRISK Study, hereafter FINRISK). The study was carried out in the Department of Neurology at the Helsinki University Hospital and approved by the Coordinating Ethics Committee of the Helsinki and Uusimaa Hospital District. Written consent from the participants of this retrospective register-based study was not required. It is not within our rights to make the data and study materials available to other researchers for purposes of reproducing the results or replicating the procedure.

## Case Population

The detailed methods from Helsinki Young Stroke Registry were published previously.<sup>1</sup> In short, the registry includes all consecutive patients aged 15 to 49 admitted to Helsinki University Hospital between January 1994 and May 2007 with

a discharge diagnosis of first-ever IS. All patients underwent a range of blood tests, a chest radiograph, an ECG, and brain imaging at admission. Stroke risk factors were obtained from medical, laboratory, and imaging records, a nationwide electronic hospital discharge register (Care Register for Health Care, hereafter hospital discharge register), and a register of reimbursed prescribed medications from the Social Insurance Institution of Finland.<sup>15,16</sup> Stroke subtypes were classified according to the modified Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria<sup>17,18</sup> and grouped as follows to allow for meaningful analysis: (1) large-artery atherosclerosis, (2) cardioembolism from a high-risk source (cardioembolism high-risk), (3) small-vessel occlusion, (4) dissections, (5) other determined etiologies, (6) embolic stroke of undetermined source (ESUS), and (7) undetermined non-ESUS. Patients fulfilling ESUS criteria<sup>19</sup> were identified among those originally classified as cryptogenic (TOAST 5b) or cardioembolism from a low-risk source (cardioembolism low-risk, TOAST 2).<sup>20</sup> Remaining patients in TOAST categories 5a–c were classified as undetermined non-ESUS.

## Control Population

FINRISK is a large population-based Finnish health examination survey on the risk factors of noncommunicable diseases coordinated by the National Institute for Health and Welfare.<sup>21</sup> At present, surveys are carried out every 5 years using representative random population samples from different parts of Finland. Its target population includes individuals aged between 25 and 74, stratified to contain at least 250 subjects from each sex in 10-year age groups (25–34, 35–44, 45–54, 55–64, and 65–74) from each geographic region.

For this study, a stroke-free control population, frequency matched for age (5-year bands) and sex, was randomly selected from the southern Finland FINRISK survey participants from years 1997, 2002, and 2007, that is, from the same time period and the same geographic region as the case population. Only one control per case for the oldest age groups (men 40–49 and women 45–49) was available, whereas in younger age groups we were able to include from 2 to 4 controls per case. The stroke risk factors among controls were collected from the FINRISK study questionnaires, a standardized health examination, blood samples, and electronic health registers (hospital discharge register, Drug Reimbursement Register, and a registry of prescribed medications from the Social Insurance Institution of Finland).<sup>15,16</sup>

## Risk Factors

We examined 11 well-established vascular risk factors, for which we were able to create comparable covariates: active malignancy, atrial fibrillation (AF), cardiovascular disease

(CVD), current smoking status, a family history of stroke, high low-density lipoprotein cholesterol (LDL-C), low high-density lipoprotein cholesterol (HDL-C), high triglycerides, hypertension, type 1 (T1D) and type 2 (T2D) diabetes mellitus. Detailed definitions of the risk factors appear in Data S1.

## Statistical Analyses

Statistical analyses were performed using SPSS for Windows version 22.0 (Armonk, NY). We considered a 2-sided  $P < 0.05$  as statistically significant.

First, we performed a comparison between groups using the chi-square and Fisher's exact tests, and calculated univariate odds ratios (ORs) for the 11 dichotomized risk factors. Second, sex, age, lipid-lowering treatment, and each risk factor with a significant association in the univariate comparisons were entered into a binary multivariable logistic regression model, for which the adjusted ORs and 95% confidence intervals were calculated. We used a backward stepwise logistic regression analysis with a statistical variable removal level of  $P < 0.10$ . We calculated the population attributable risk percentages from the adjusted OR and 95% confidence interval values using the following formula: prevalence of exposed cases  $\times [(OR - 1) / OR] \times 100$ .<sup>22</sup>

Apart from the entire study population, we performed subgroup analyses by sex, age group (25–39 years and 40–49 years), and IS etiology. In the etiology-specific analyses, the entire control group served as controls.

Additional description of statistical analyses is provided in Data S1.

## Results

Among the 1008 patients in Helsinki Young Stroke Registry, 4 were excluded as stroke mimics during the follow-up period, and another 43 were under 25 years old, for whom controls were not available. Thus, the study population in the univariate analysis consisted of 961 cases who fell between the age of 25 and 49 (358 women and 603 men) and 1403 controls (610 women and 793 men). The age distribution of cases and controls appears in Table S1, while Table S2 provides the distribution of IS etiologies. Due to missing data for dichotomous variables, 43 cases (4.5%) and 11 controls (0.8%) were excluded from the multivariable analyses. Hence, our fully adjusted multivariable analyses included data from 918 cases and 1392 controls. For a detailed description of the missing data and comparison of subjects with complete and incomplete data, please see the Data S1 and Tables S3 and S4.

## Dichotomized Risk Factors

Among the entire study population, all studied risk factors were more prevalent among cases compared with controls, with the exception of high LDL-C. In the multivariable logistic regression analysis, significant risk factors for IS beginning with the highest OR consisted of AF, CVD, T1D, T2D, low HDL-

**Table 1.** Univariate Analysis and Multivariable Logistic Regression Analysis of Risk Factors for Early-Onset Ischemic Stroke

	Prevalence				
	Case (n=961) n/N (%)	Control (n=1403) n/N (%)	Unadjusted OR (95% CI)	Adjusted OR (95% CI)*	PAR% (95% CI)*
Atrial fibrillation	23/961 (2.4)	2/1403 (0.1)	17.18 (4.04 to 73.03)	10.43 (2.33 to 46.77)	2.2% (1.4 to 2.3)
Cardiovascular disease <sup>†</sup>	46/961 (4.8)	5/1403 (0.4)	14.06 (5.56 to 35.51)	8.01 (3.09 to 20.78)	4.2% (3.2 to 4.6)
Type 1 diabetes mellitus	44/961 (4.6)	9/1403 (0.6)	7.43 (3.61 to 15.30)	6.72 (3.15 to 14.33)	3.9% (3.1 to 4.3)
Active malignancy	15/961 (1.6)	6/1396 (0.4)	3.67 (1.42 to 9.50)	2.73 (0.99 to 7.50)	1.0% (0.0 to 1.4)
Type 2 diabetes mellitus	44/961 (4.6)	24/1403 (1.7)	2.76 (1.67 to 4.57)	2.31 (1.35 to 3.95)	2.6% (1.2 to 3.4)
Low HDL-C <sup>‡</sup>	148/921 (16.1)	114/1400 (8.1)	2.16 (1.67 to 2.80)	1.81 (1.37 to 2.40)	7.2% (4.4 to 9.4)
Current smoking status	427/961 (44.4)	435/1403 (31.0)	1.78 (1.50 to 2.11)	1.81 (1.50 to 2.17)	19.9% (14.8 to 23.9)
Hypertension	391/961 (40.7)	394/1402 (28.1)	1.76 (1.48 to 2.09)	1.43 (1.17 to 1.75)	12.2% (5.9 to 17.4)
Family history of stroke	126/961 (13.1)	128/1403 (9.1)	1.50 (1.16 to 1.95)	1.37 (1.04 to 1.82)	3.5% (0.5 to 5.9)
High triglycerides <sup>‡</sup>	212/926 (22.9)	217/1400 (15.5)	1.62 (1.31 to 2.00)	1.19 (0.94 to 1.53)	3.7% (−1.5 to 7.9)
High LDL-C <sup>‡</sup>	473/921 (51.4)	850/1400 (60.7)	0.68 (0.58 to 0.81)	0.51 (0.42 to 0.62)	−49.4% (−80.0 to −31.5)

CI indicates confidence interval; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; n/N, number of subjects divided by total of subjects excluding those with missing data; OR, odds ratio; PAR%, population attributable risk percentage.

\*Age, sex, lipid-lowering treatment, and each risk factor in the table were entered into the multivariate model. Multivariable analysis included 918 cases and 1392 controls.

<sup>†</sup>The presence of coronary heart disease, heart failure, or peripheral arterial disease.

<sup>‡</sup>High LDL-C  $\geq 3.0$  mmol/L (116 mg/dL), low HDL-C  $< 1.0$  mmol/L (39 mg/dL), and high triglycerides  $\geq 2.0$  mmol/L (177 mg/dL).

**Table 2.** Multivariable Logistic Regression Analysis of Risk Factors for Early-Onset Ischemic Stroke by Sex

	Women			Men		
	Case (n=358) n/N (%)	Control (n=610) n/N (%)	Adjusted OR (95% CI)*	Case (n=603) n/N (%)	Control (n=793) n/N (%)	Adjusted OR (95% CI)*
Atrial fibrillation	3/358 (0.8)	2/610 (0.3)	1.36 (0.18–20.40)	20/603 (3.3)	0/793 (0.0)	NA <sup>†</sup>
Active malignancy	8/358 (2.2)	1/605 (0.2)	10.88 (1.29–91.56)	7/603 (1.2)	5/791 (0.6)	1.27 (0.36–4.48)
Cardiovascular disease <sup>‡</sup>	11/358 (3.1)	4/610 (0.7)	2.08 (0.59–7.31)	35/603 (5.8)	1/793 (0.1)	28.06 (3.77–209.05)
Current smoking status	136/358 (38.0)	192/610 (31.5)	1.58 (1.17–2.13)	291/603 (48.3)	243/793 (30.6)	1.95 (1.54–2.48)
Family history of stroke	48/358 (13.4)	55/610 (9.0)	1.38 (0.88–2.16)	78/603 (12.9)	73/793 (9.2)	1.49 (1.04–2.15)
High LDL-C <sup>§</sup>	134/341 (39.3)	346/593 (58.3)	0.33 (0.24–0.44)	339/580 (58.4)	489/792 (61.7)	0.67 (0.52–0.86)
High triglycerides <sup>§</sup>	55/341 (16.1)	86/608 (14.1)	1.09 (0.71–1.67)	157/585 (26.8)	131/792 (16.5)	1.24 (0.91–1.67)
Hypertension	116/358 (32.4)	147/609 (24.1)	1.34 (0.96–1.87)	275/603 (45.6)	247/793 (31.1)	1.40 (1.08–1.80)
Low HDL-C <sup>§</sup>	33/340 (9.7)	54/608 (8.9)	1.20 (0.73–1.96)	115/581 (19.8)	60/792 (7.6)	2.36 (1.66–3.37)
Type 1 diabetes mellitus	15/358 (4.2)	3/610 (0.5)	8.64 (2.35–31.75)	29/603 (4.8)	6/793 (0.8)	5.83 (2.28–14.88)
Type 2 diabetes mellitus	13/358 (3.6)	11/610 (1.8)	2.28 (0.96–5.41)	31/603 (5.1)	13/793 (1.6)	2.19 (1.08–4.44)

CI indicates confidence interval; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; n/N, number of subjects divided by total of subjects excluding those with missing data; NA, not applicable; OR, odds ratio.

\*Age, lipid-lowering treatment, and each available risk factor in the table were entered into the model. Multivariable analyses included 339 female cases and 602 controls, and 579 male cases and 790 controls.

<sup>†</sup>Atrial fibrillation could not be analyzed in men because of the lack of male controls with atrial fibrillation.

<sup>‡</sup>The presence of coronary heart disease, heart failure, or peripheral arterial disease.

<sup>§</sup>High LDL-C  $\geq 3.0$  mmol/L (116 mg/dL), low HDL-C  $< 1.0$  mmol/L (39 mg/dL), and high triglycerides  $\geq 2.0$  mmol/L (177 mg/dL).

C, current smoking status, hypertension, and a family history of stroke. High LDL-C, however, was inversely associated with IS (Table 1).

### Population Attributable Risks

For the entire young IS population, we found the highest population attributable risk percentages for current smoking status, followed by hypertension, low HDL-C, CVD, T1D, a family history of stroke, T2D, and AF (Table 1).

### Sex-Specific Analysis

In the sex-specific multivariable analysis (Table 2), T1D and current smoking status were significantly associated with IS in both sexes. However, CVD, low HDL-C, T2D, a family history of stroke, and hypertension emerged as significant risk factors only among men, whereby CVD carried the strongest association. Active malignancy emerged as a strong risk factor only for women. Furthermore, high LDL-C was inversely associated with IS among both sexes, although the association was stronger among women. We could not analyze the strength of the association between AF and IS among men because there were no male controls with AF.

### Age-Specific Analysis

In the age-specific analyses (Table 3), T1D, current smoking status, and low HDL-C emerged as significant risk factors for IS among individuals aged 25 to 39 and 40 to 49 years. A family history of stroke was significantly associated with IS only in the younger group of individuals aged 25 to 39 years. In addition, AF, CVD, T2D, and hypertension emerged as significant risk factors among subjects aged  $\geq 40$  years only. High LDL-C was inversely associated with IS among both age groups. We could not analyze the strength of the association between active malignancy and IS among the younger age group because there were no younger controls with active malignancy.

### Etiological Subgroups

For large-artery atherosclerosis, significant associations emerged for CVD, T1D, current smoking status, hypertension, T2D, and low HDL-C. For cardioembolism high-risk, AF, CVD, T1D, and low HDL-C carried significant associations. For small-vessel occlusion, T1D, CVD, T2D, hypertension, and current smoking status emerged as risk factors for IS. In these 3 subgroups, high LDL-C showed no statistically significant association (Figure 1).

For dissection, T1D, low HDL-C, and a family history of stroke emerged as significant. For other rare determined

**Table 3.** Multivariable Logistic Regression Analysis of Risk Factors for Early-Onset Ischemic Stroke by Age Group

	Age 25–39 Years			Age 40–49 Years		
	Case (n=267) n/N (%)	Control (n=620) n/N (%)	Adjusted OR (95% CI)*	Case (n=694) n/N (%)	Control (n=783) n/N (%)	Adjusted OR (95% CI)*
Atrial fibrillation	1/267 (0.4)	1/620 (0.2)	2.62 (0.15–46.34)	22/694 (3.2)	1/783 (0.1)	16.01 (2.08–123.03)
Active malignancy	1/267 (0.4)	0/617 (0.0)	NA <sup>§</sup>	14/694 (2.0)	6/778 (0.4)	2.48 (0.88–7.02)
Cardiovascular disease <sup>†</sup>	4/267 (1.5)	2/620 (0.3)	2.57 (0.43–15.37)	42/694 (6.1)	3/783 (0.4)	11.07 (3.34–36.71)
Current smoking status	112/267 (41.9)	187/620 (30.2)	1.72 (1.26–2.35)	315/694 (45.4)	248/783 (31.3)	1.86 (1.48–2.34)
Family history of stroke	28/267 (10.5)	32/620 (5.2)	1.82 (1.03–3.24)	98/694 (14.1)	96/783 (12.3)	1.27 (0.92–1.76)
High LDL-C <sup>‡</sup>	102/253 (40.3)	311/617 (50.4)	0.54 (0.39–0.74)	371/668 (55.5)	539/783 (68.8)	0.48 (0.37–0.60)
High triglycerides <sup>‡</sup>	42/255 (16.5)	65/617 (10.5)	1.50 (0.94–2.38)	170/671 (25.3)	152/783 (19.4)	1.12 (0.84–1.49)
Hypertension	52/267 (19.5)	102/620 (16.5)	0.98 (0.64–1.49)	339/694 (37.3)	292/782 (48.8)	1.58 (1.26–1.99)
Low HDL-C <sup>‡</sup>	35/253 (13.8)	40/617 (6.5)	2.00 (1.20–3.33)	113/668 (16.9)	74/783 (9.5)	1.66 (1.18–2.32)
Type 1 diabetes mellitus	10/267 (3.7)	3/620 (0.5)	8.78 (2.30–33.58)	34/694 (4.9)	6/783 (0.8)	6.45 (2.58–16.14)
Type 2 diabetes mellitus	5/267 (1.9)	4/620 (0.6)	1.94 (0.48–7.78)	39/694 (5.6)	20/783 (2.6)	2.38 (1.33–4.28)

HDL-C indicates high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol.

\*Age, sex, lipid-lowering treatment, and each available risk factor in the table were entered into the model. Multivariable analyses included 253 cases and 615 controls between the ages of 25–39, and 665 cases and 777 controls between the ages of 40–49.

<sup>†</sup>The presence of coronary heart disease, heart failure, or peripheral arterial disease.

<sup>‡</sup>High LDL-C  $\geq 3.0$  mmol/L (116 mg/dL), low HDL-C  $< 1.0$  mmol/L (39 mg/dL), and high triglycerides  $\geq 2.0$  mmol/L (177 mg/dL).

<sup>§</sup>Active malignancy could not be analyzed in the younger age group because of the lack of controls with an active malignancy.

etiologies, active malignancy showed the strongest association, followed by current smoking status, a family history of stroke, and high triglycerides. For ESUS, only current smoking status exhibited a statistically significant association. For undetermined non-ESUS, low HDL-C and current smoking status emerged as risk factors for IS. In these latter 4 etiological groups, high LDL-C was inversely associated with IS (Figure 2).

Because of the absence of risk factor–positive cases, AF could not be analyzed in subgroups other than cardioembolism high-risk, active malignancy could not be analyzed in the small-vessel occlusion or dissection subgroups, CVD could not be analyzed in the dissection subgroup, and T1D could not be analyzed in the ESUS subgroup.

### Additional and Sensitivity Analyses

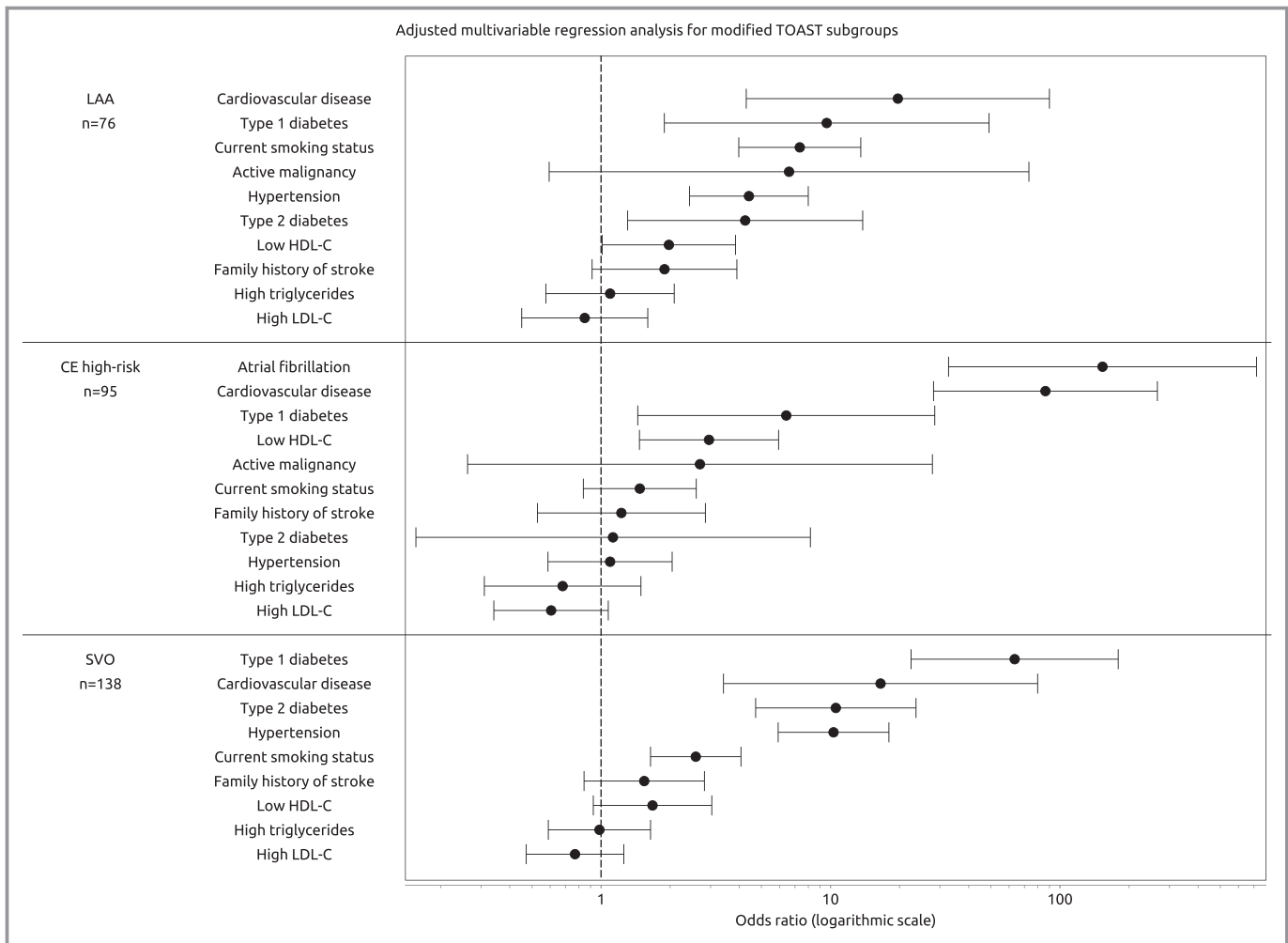
We performed additional analyses, specifically examining the alternative risk factor covariates, hospital discharge register, and the classification of high LDL-C. The results from these sensitivity analyses did not contradict our main findings. For a detailed description, please see Data S1.

### Discussion

Our study established the associations of 11 risk factors for early-onset ischemic stroke. In particular, our results showed

stronger associations than expected for AF, CVD, and both types of diabetes mellitus. Given these findings, a new hypothesis emerges suggesting that these comorbidities may represent more aggressive forms of disease predisposing individuals to a severe vascular end point early in life. This phenomenon has been characterized as “losing the relative protection of youth” in cases of early-onset type 2 diabetes mellitus.<sup>23</sup> As expected, we found marked differences in risk profiles among the age, sex, and etiological-specific subgroups, which may help clinicians to establish personalized risk factor screening strategies and aid in focusing future research.

Compared with results among a young American white IS population,<sup>10</sup> our study showed a similar population attributable risk percentage and association for current smoking status. Moreover, another 2 recent studies on European population<sup>11</sup> and on young American men<sup>24</sup> also showed a rather similar strengths of association for current smoking. Our study confirms that the association between current smoking and early-onset ischemic stroke is present across both sexes, both age groups, and in most etiological subgroups. Various tobacco control policy measures have been implemented in Finland since the 1980s,<sup>25</sup> and the rate of daily smoking has been decreasing, especially among men. Given that smoking is a modifiable risk factor, it appears essential to further support tobacco abstinence to prevent ischemic strokes among young adults.



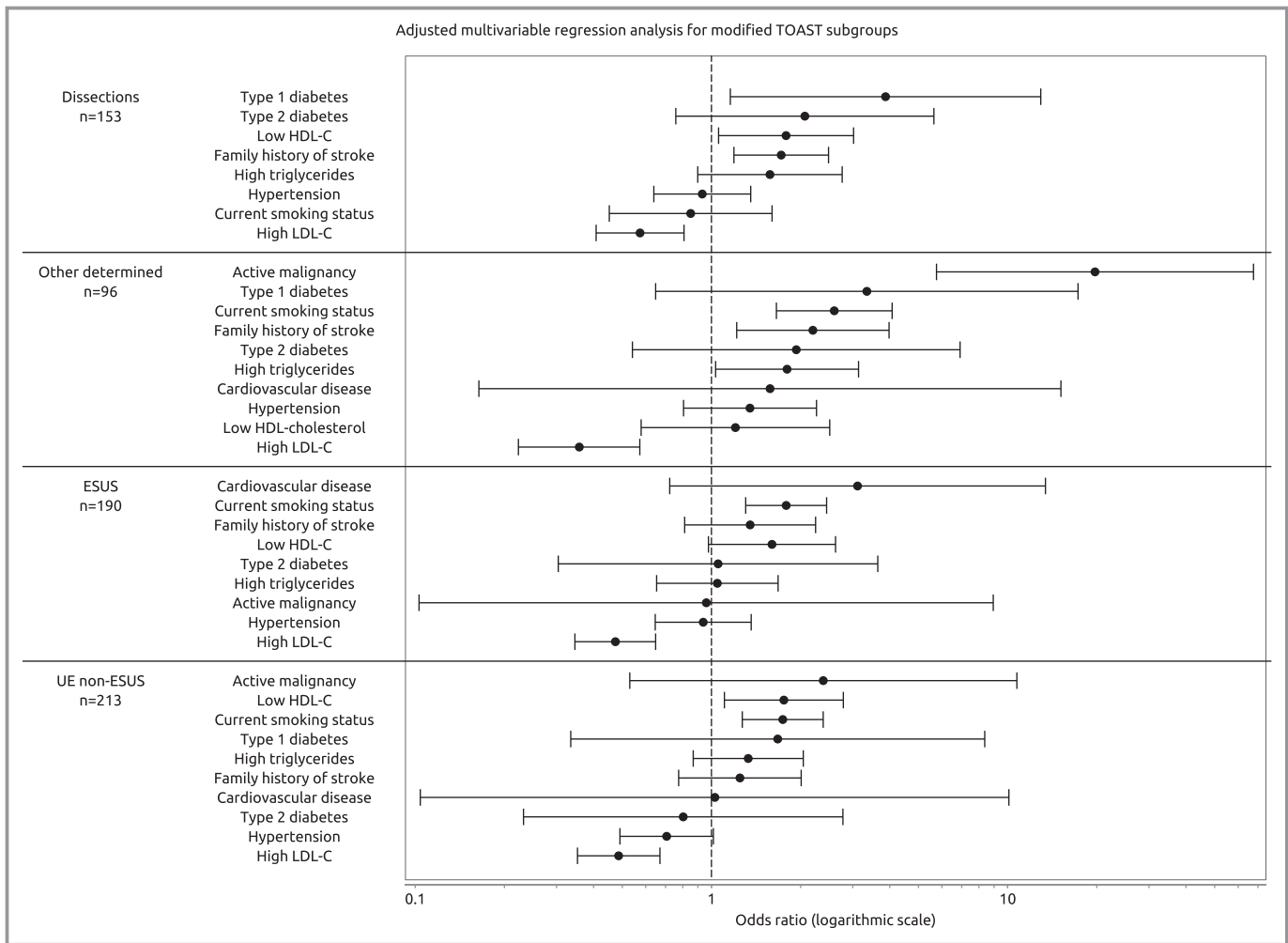
**Figure 1.** Multivariable logistic regression analysis of the risk factors for early-onset ischemic stroke in the following etiologies: large artery atherosclerosis (LAA), cardioembolism from high-risk source, and small vessel occlusion (SVO). Age, sex, lipid-lowering treatment, and each available risk factor were entered into the model. HDL-C indicates high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol.

The strength of the association with diabetes mellitus has varied widely across previous studies.<sup>10–14,26–28</sup> We analyzed T1D and T2D separately and found that both types of diabetes mellitus carried a high risk for IS at a young age, although T1D carried a stronger and more consistent association across demographic and etiological subgroups than did T2D.

Whether an association between dyslipidemia and young-onset IS exists has remained unclear. Several studies found an association with low HDL-C and early-onset IS or transient ischemic attack<sup>26,29,30</sup> but no association with LDL-C.<sup>29,30</sup> Similar to prior studies,<sup>26,29,30</sup> our study confirmed that low HDL-C represents a risk factor for early-onset IS. In addition to its reverse cholesterol transport action, HDL-C is associated with antithrombotic and anti-inflammatory actions.<sup>31,32</sup> Thus, it is reasonable to hypothesize that the entire spectrum of HDL-C functions may play a more important role in young adult IS. Interestingly, high LDL-C exhibited an inverse

association with IS only in subgroups of etiologies that frequently occur among young individuals—that is, dissections, other determined causes, ESUS, and undetermined non-ESUS—while the association with high LDL-C was absent among older-onset causes. Although our study cannot establish causality, at the very least this finding indicates that high LDL-C does not play a role in the IS pathogenic mechanisms that occur more frequently among younger patients. Furthermore, low cholesterol could merely represent a confounder reflecting an underlying susceptibility factor for early-onset IS.<sup>33</sup>

We demonstrated a family history of stroke as a significant risk factor for IS in the younger (aged 29–39) but not in the older (aged 40–49) age group. This finding emphasizes the hypothesis that genetic factors may have a stronger contribution to early-onset than older-onset IS.<sup>34</sup> Furthermore, we observed a similar association with family



**Figure 2.** Multivariable logistic regression analysis of the risk factors for early-onset ischemic stroke in the following etiologies: dissections, other determined etiologies, embolic stroke of an undetermined source (ESUS), and undetermined etiology non-ESUS (undetermined non-ESUS). Age, sex, lipid-lowering treatment, and each available risk factor were entered into the model. HDL-C indicates high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol.

history in those with dissection or other rare determined IS etiology. This suggests that inherited factors<sup>35</sup> might be in the play in these etiological subgroups, although less frequent family history of young-onset stroke also has been reported with dissection compared with strokes attributable to other causes.<sup>36</sup>

Accumulation of vascular risk factors with increasing age among young patients with IS was associated with increased risk of future atherothrombotic events.<sup>37</sup> In accordance with this, we found that more of the vascular risk factors were significantly associated with first IS in those who were aged 40 to 49 compared with those who were under 40.

Furthermore, our study adds to the knowledge of the connection between new cancer<sup>38</sup> and stroke in the young population, particularly in young women. Among our young female patients, 62.5% of the active malignancies were metastatic at stroke onset, and the most common primary

locations were breast cancer (37.5%) and ovarian cancer (25%).

The major strength of our study is that our cases consist of a large, consecutive, and homogenous young patient population with IS. The Helsinki Young Stroke Registry can be considered a population-based register given that it covers practically all hospitalized young patients with IS in the Helsinki University Hospital catchment area of 1.5 million people.<sup>1</sup> Our register is free of consent bias and therefore able to include a truly nonselective group of young patients with IS. Furthermore, IS etiology among cases was classified according to the modified TOAST criteria, allowing us to meaningfully analyze IS risk factors by etiological subgroups. Our controls were drawn from a large population-based cohort study from the same time period and geographic region. In addition, we analyzed 11 traditional vascular risk factors in a single study and were able to use a combination of reliable

data sources including nationwide electronic register data. Hence, we could include risk factors rarely analyzed in early-onset stroke, such as AF and CVD, and calculate specific risk estimates for both T1D and T2D.

The limitations of our study include its retrospective nature and several possible sources of biases inherent for any case-control study. For instance, although we carefully harmonized risk factors, some residual bias may remain. We included an unequal ratio of younger and older controls, which affected the prevalence of risk factors and the unadjusted ORs. Still, the prevalence of AF, CVD, and active malignancy among our controls remained low, and we could not analyze AF and active malignancy across all subgroups. Overall, there were too few events, especially in the subgroups, to estimate associations precisely. Consequently, the estimated ORs for infrequent risk factors (AF, CVD, T1D, and active malignancy) should be interpreted with caution. Additionally, the sensitivity of older hospital discharge register data remains suboptimal in identifying all subjects with AF or CVD.<sup>16</sup> As a limitation regarding current smoking status, we were unable to analyze the dose-response relationship between cigarette smoking and early-onset IS. Furthermore, we did not include any risk factors related to women's reproductive health or migraine with aura among our covariates. Finally, the inverse association for LDL-C identified in our study could result from a reverse causality, as lipids among cases were measured 12 to 72 hours after IS.<sup>39,40</sup> While the shorter fasting time required for controls also may have minimally affected the lipid levels,<sup>41</sup> several studies have established that plasma lipids change only modestly in response to habitual food intake.<sup>42</sup>

## Conclusions

AF, CVD, both primary types of diabetes mellitus, low HDL-C, current smoking status, hypertension, and a family history of stroke emerge as risk factors for IS among young patients. Additionally, current smoking status and T1D exhibited the most consistent associations across all demographic and etiological subgroups.

## Acknowledgments

We thank the staff of the Clinical Stroke Research Center at Meilahti Hospital in Helsinki, Finland, for their important contributions to the Helsinki Young Stroke Registry. We also thank Jani Pirinen, MD, for his constructive comments.

## Sources of Funding

Dr Kivioja received research grants from the Helsinki University Hospital Research Fund and participates in the

Doctoral Candidate Program funded by the Medical Faculty of the University of Helsinki. This study received funding from the Helsinki University Hospital Research Fund and the Academy of Finland (Dr Putaala). FINRISK has primarily received funding through the National Institute of Health and Welfare. Additional funding was obtained from the Academy of Finland and other domestic foundations. Dr Salomaa received support from the Finnish Foundation for Cardiovascular Research.

## Disclosures

Dr Tatlisumak received funding from the Helsinki University Hospital Research Fund, Sahlgrenska University Hospital, and the University of Gothenburg for research in young patients with stroke (all significant). Dr Putaala received a speaker's honorarium from Bayer, BMS-Pfizer, Boehringer-Ingelheim, and Abbott (all modest). Dr Putaala also received compensation to serve on the following advisory boards; Bayer, BMS-Pfizer, and Boehringer-Ingelheim (all modest). The remaining authors have no disclosures to report.

## References

- Putala J, Metso AJ, Metso TM, Konkola N, Kraemer Y, Haapaniemi E, Kaste M, Tatlisumak T. Analysis of 1008 consecutive patients aged 15 to 49 with first-ever ischemic stroke: the Helsinki Young Stroke Registry. *Stroke*. 2009;40:1195–1203.
- Rosengren A, Giang KW, Lappas G, Jern C, Torén K, Björck L. Twenty-four-year trends in the incidence of ischemic stroke in Sweden from 1987 to 2010. *Stroke*. 2013;44:2388–2393.
- Ramirez L, Kim-Tenser MA, Sanossian N, Cen S, Wen G, He S, Mack WJ, Towfighi A. Trends in acute ischemic stroke hospitalizations in the United States. *J Am Heart Assoc*. 2016;11:e003233. DOI: 10.1161/JAHA.116.003233.
- George MG, Tong X, Bowman BA. Prevalence of cardiovascular risk factors and strokes in younger adults. *JAMA Neurol*. 2017;74:695–703.
- George MG, Tong X, Kuklina EV, Labarthe DR. Trends in stroke hospitalizations and associated risk factors among children and young adults, 1995–2008. *Ann Neurol*. 2011;70:713–721.
- Kissela BM, Khouri JC, Alwell K, Moomaw CJ, Woo D, Adeoye O, Flaherty ML, Khatri P, Ferioli S, De Los Rios La Rosa F, Broderick JP, Kleindorfer DO. Age at stroke: temporal trends in stroke incidence in a large, biracial population. *Neurology*. 2012;79:1781–1787.
- von Sarnowski B, Putaala J, Grittner U, Gaertner B, Schminke U, Curtze S, Huber R, Tanislav C, Lichy C, Demarin V, Basic-Kes V, Ringelstein EB, Neumann-Haefelin T, Enzinger C, Fazekas F, Rothwell PM, Dichgans M, Jungehulsing GJ, Heuschmann PU, Kaps M, Norrving B, Rolfs A, Kessler C, Tatlisumak T, sifap1 Investigators. Lifestyle risk factors for ischemic stroke and transient ischemic attack in young adults in the Stroke in Young Fabry Patients study. *Stroke*. 2013;44:119–125.
- Putala J, Yesilot N, Waje-Andreassen U, Pitkaniemi J, Vassilopoulou S, Nardi K, Odier C, Hofgart G, Engelter S, Burow A, Mihalka L, Kloss M, Ferrari J, Lemmens R, Coban O, Haapaniemi E, Maaijwee N, Rutten-Jacobs L, Bersano A, Cereda C, Baron P, Borellini L, Valcarengi C, Thomassen L, Grau AJ, Palm F, Urbanek C, Tuncay R, Durukan-Tolvanen A, van Dijk EJ, de Leeuw FE, Thijs V, Greisenegger S, Vemmos K, Lichy C, Berczki D, Csiba L, Michel P, Leys D, Spengos K, Naess H, Bahar SZ, Tatlisumak T. Demographic and geographic vascular risk factor differences in European young adults with ischemic stroke: the 15 cities young stroke study. *Stroke*. 2012;43:2624–2630.
- Andersson C, Vasan RS. Epidemiology of cardiovascular disease in young individuals. *Nat Rev Cardiol*. 2018;15:230–240.
- Rohr J, Kittner S, Feeser B, Hebel JR, Whyte MG, Weinstein A, Kanarak N, Buchholz D, Earley C, Johnson C, Macko R, Price T, Sloan M, Stern B, Wityk R, Wozniak M, Sherwin R. Traditional risk factors and ischemic stroke in young



- adults: the Baltimore-Washington Cooperative Young Stroke Study. *Arch Neurol*. 1996;53:603–607.
11. Aigner A, Grittner U, Rolfs A, Norrvig B, Siegerink B, Busch MA. Contribution of established stroke risk factors to the burden of stroke in young adults. *Stroke*. 2017;48:1744–1751.
  12. Naess H, Nyland HI, Thomassen L, Aarseth J, Myhr KM. Etiology of and risk factors for cerebral infarction in young adults in western Norway: a population-based case-control study. *Eur J Neurol*. 2004;11:25–30.
  13. You RX, McNeil JJ, O'Malley HM, Davis SM, Thrift AG, Donnan GA. Risk factors for stroke due to cerebral infarction in young adults. *Stroke*. 1997;28:1913–1918.
  14. Hillbom M, Numminen H, Juvela S. Recent heavy drinking of alcohol and embolic stroke. *Stroke*. 1999;30:2307–2312.
  15. Sund R. Quality of the Finnish Hospital Discharge Register: a systematic review. *Scand J Public Health*. 2012;40:505–515.
  16. Mähönen M, Jula A, Harald K, Antikainen R, Tuomilehto J, Zeller T, Blankenberg S, Salomaa V. The validity of heart failure diagnoses obtained from administrative registers. *Eur J Prev Cardiol*. 2013;20:254–259.
  17. Adams HP Jr, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE III. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke*. 1993;24:35–41.
  18. Ay H, Furie KL, Singhal A, Smith WS, Sorensen AG, Koroshetz WJ. An evidence-based causative classification system for acute ischemic stroke. *Ann Neurol*. 2005;58:688–697.
  19. Hart RG, Diener HC, Coutts SB, Easton JD, Granger CB, O'Donnell MJ, Sacco RL, Connolly SJ; Cryptogenic Stroke/ESUS International Working Group. Embolic strokes of undetermined source: the case for a new clinical construct. *Lancet Neurol*. 2014;13:429–438.
  20. Martinez-Majander N, Aarnio K, Pirinen J, Lumikari T, Nieminen T, Lehto M, Sinisalo J, Kaste M, Tatlisumak T, Putaala J. Embolic strokes of undetermined source in young adults: baseline characteristics and long-term outcome. *Eur J Neurol*. 2018;25:535–541.
  21. Borodulin K, Vartiainen E, Peltonen M, Jousilahti P, Juolevi A, Laatikainen T, Männistö S, Salomaa V, Sundvall J, Puska P. Forty-year trends in cardiovascular risk factors in Finland. *Eur J Public Health*. 2015;25:539–546.
  22. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. *Am J Public Health*. 1998;88:15–19.
  23. Hillier TA, Pedula KL. Complications in young adults with early-onset type 2 diabetes: losing the relative protection of youth. *Diabetes Care*. 2003;26:2999–3005.
  24. Markidan J, Cole JW, Cronin CA, Merino JG, Phipps MS, Wozniak MA, Kittner SJ. Smoking and risk of ischemic stroke in young men. *Stroke*. 2018;49:1276–1278.
  25. Helakorpi SA, Martelin TP, Torppa JO, Patja KM, Kiiskinen UA, Vartiainen EA, Uutela AK. Did the tobacco control act amendment in 1995 affect daily smoking in Finland? Effects of a restrictive workplace smoking policy. *J Public Health*. 2008;30:407–414.
  26. Lipska K, Sylaja PN, Sarma PS, Thankappan KR, Kutty VR, Vasani RS, Radhakrishnan K. Risk factors for acute ischaemic stroke in young adults in South India. *J Neurol Neurosurg Psychiatry*. 2007;78:959–963.
  27. Haapaniemi H, Hillbom M, Juvela S. Lifestyle-associated risk factors for acute brain infarction among persons of working age. *Stroke*. 1997;28:26–30.
  28. Love BB, Biller J, Jones MP, Adams HP Jr, Bruno A. Cigarette smoking: a risk factor for cerebral infarction in young adults. *Arch Neurol*. 1990;47:693–698.
  29. Albuquer JF, Ferrieres J, Ruidavets JB, Guiraud-Chaumeil B, Perret BP, Chollet F. Serum lipids in young patients with ischaemic stroke: a case-control study. *J Neurol Neurosurg Psychiatry*. 2000;69:29–33.
  30. Marini C, Carolei A, Roberts RS, Prencipe M, Gandolfo C, Inzitari D, Landi G, De Zanche L, Scoditti U, Fieschi C. Focal cerebral ischemia in young adults: a collaborative case-control study. *Neuroepidemiology*. 1993;12:70–81.
  31. Van der Stoep M, Korporaal SJ, Van Eck M. High-density lipoprotein as a modulator of platelet and coagulation responses. *Cardiovasc Res*. 2014;103:362–371.
  32. Rye KERRY-ANNE, Barter PHILIP J. Cardioprotective functions of HDLs. *J Lipid Res*. 2014;55:168–179.
  33. Debette S, Metso T, Pezzini A, Abboud S, Metso A, Leys D, Bersano A, Louillet F, Caso V, Lamy C, Medeiros E, Samson Y, Grond-Ginsbach C, Engelter ST, Thijs V, Beretta S, Béjot Y, Sessa M, Lorenza Muesan M, Amouyel P, Castellano M, Arveiler D, Tatlisumak T, Dallongeville J; Cervical Artery Dissection and Ischemic Stroke Patients (CADISP) Group. Association of vascular risk factors with cervical artery dissection and ischemic stroke in young adults. *Circulation*. 2011;123:1537–1544.
  34. MacClellan LR, Mitchell BD, Cole JW, Wozniak MA, Stern BJ, Giles WH, Brown DW, Sparks MJ, Kittner SJ. Familial aggregation of ischemic stroke in young women: the Stroke Prevention in Young Women Study. *Genet Epidemiol*. 2006;30:602–608.
  35. Debette S, Markus HS. The genetics of cervical artery dissection: a systematic review. *Stroke*. 2009;40:e459–e466.
  36. Kloss M, Grond-Ginsbach C, Pezzini A, Metso TM, Metso AJ, Debette S, Leys D, Dallongeville J, Caso V, Thijs V, Bersano A, Touzé E, Bonati LH, Tatlisumak T, Arnold ML, Lyrer PA, Engelter ST; Cervical Artery Dissection and Ischemic Stroke Patients (CADISP) Study Group. Stroke in first-degree relatives of patients with cervical artery dissection. *Eur J Neurol*. 2014;21:1102–1107.
  37. Putaala J, Haapaniemi E, Kaste M, Tatlisumak T. How does number of risk factors affect prognosis in young patients with ischemic stroke? *Stroke*. 2012;43:356–361.
  38. Navi BB, Reiner AS, Kamel H, Iadecola C, Elkind MS, Panageas KS, DeAngelis LM. Association between incident cancer and subsequent stroke. *Ann Neurol*. 2015;77:291–300.
  39. Butterworth RJ, Marshall WJ, Bath PMW. Changes in serum lipid measurements following acute ischaemic stroke. *Cerebrovasc Dis*. 1997;7:10–13.
  40. Kargman DE, Tuck C, Berglund L, Lin IF, Mukherjee RS, Thompson EV, Jones J, Boden-Albala B, Paik MC, Sacco RL. Lipid and lipoprotein levels remain stable in acute ischemic stroke: the Northern Manhattan Stroke Study. *Atherosclerosis*. 1998;139:391–399.
  41. Sundvall J, Laatikainen T, Hakala S, Leiviskä J, Alfthan G. Systematic error of serum triglyceride measurements during three decades and the effect of fasting on serum triglycerides in population studies. *Clin Chim Acta*. 2008;397:55–59.
  42. Nordestgaard BG, Langsted A, Mora S, Kolovou G, Baum H, Bruckert E, Watts GF, Sypniewska G, Wiklund O, Borén J, Chapman MJ, Cobbaert C, Descamps OS, von Eckardstein A, Kamstrup PR, Pulkki K, Kronenberg F, Remaley AT, Rifai N, Ros E, Langlois M; European Atherosclerosis Society (EAS) and the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Joint Consensus Initiative. Fasting Is Not Routinely Required for Determination of a Lipid Profile: Clinical and Laboratory Implications Including Flagging at Desirable Concentration Cutpoints—A Joint Consensus Statement from the European Atherosclerosis Society and European Federation of Clinical Chemistry and Laboratory Medicine. *Clin Chem*. 2016;62:930–946.

# **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  $\geq 1$  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  $\geq 7$  mmol/L (126 mg/dL) or a 2-hour oral glucose tolerance test of  $\geq 11.1$  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.<sup>1</sup> 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  $\geq 3.0$  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  $\geq 2.0$  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  $\geq 3.0$  mmol/L or 116 mg/dL, low HDL-C of  $< 1.0$  mmol/L or 39 mg/dL, or high total cholesterol of  $\geq 5.0$  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  $\geq 4$ -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-MgCl<sub>2</sub> 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 ( $> 4.5$  mmol/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)  $\geq 140$  mmHg or a diastolic blood pressure (DBP)  $\geq 90$  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<sup>2</sup> 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.**

	<b>Women</b>		<b>Men</b>		
<b>Age category</b>	<b>Cases</b>	<b>Controls</b>	<b>Cases</b>	<b>Controls</b>	<b>Total</b>
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.**

	<b>All cases (n=961)</b>	<b>Cases with complete data (n=918)</b>
<b>Etiologic subgroup</b>	<b>No. (%)</b>	<b>No. (%)</b>
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)	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)§	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.**

	<b>Cases with complete data* (n=918)</b>	<b>Cases with incomplete data* (n=43)</b>	<b>Controls with complete data* (n=1392)</b>	<b>Controls with incomplete data* (n=11)</b>
	<b>No. (%)</b>	<b>No. (%)</b>	<b>No. (%)</b>	<b>No. (%)</b>
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)  $\geq 3.0$  mmol/L (116 mg/dL), low high-density lipoprotein cholesterol (HDL-C)  $< 1.0$  mmol/L (39 mg/dL), and high triglycerides  $\geq 2.0$  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.