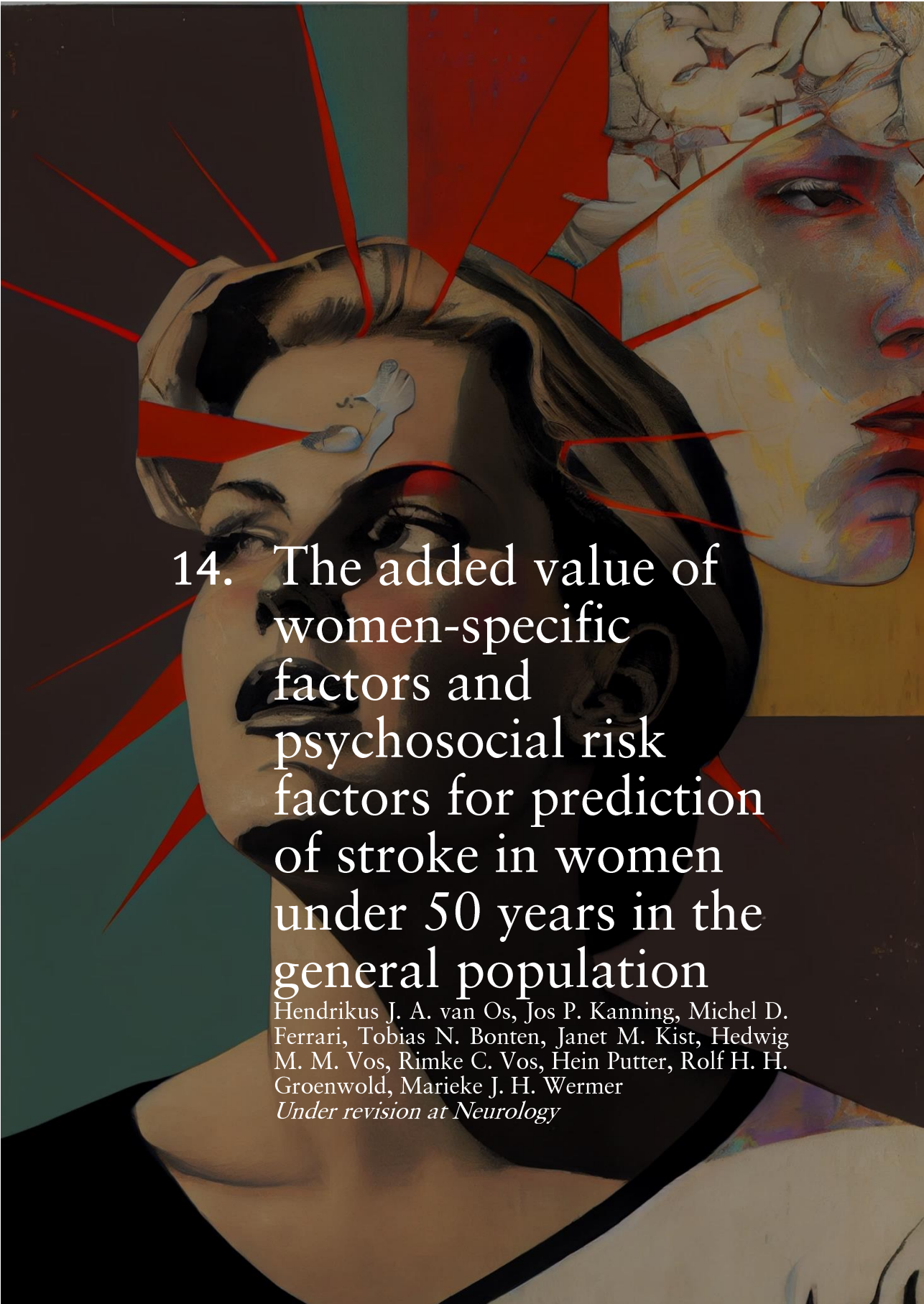
Hacking stroke in women: towards aetiology-driven precision prevention

Os, H.J.A. van



14. The added value of women-specific factors and psychosocial risk factors for prediction of stroke in women under 50 years in the general population

Hendrikus J. A. van Os, Jos P. Kanning, Michel D. Ferrari, Tobias N. Bonten, Janet M. Kist, Hedwig M. M. Vos, Rimke C. Vos, Hein Putter, Rolf H. H. Groenwold, Marieke J. H. Wermer

Under revision at Neurology

Abstract

Background: Female-specific and psychosocial factors may be important in the prediction of stroke, but are not included in prediction models that are currently used. We investigated the added predictive value of these factors in women under 50.

Methods: We used data from the STIZON, population-based, primary care database of women aged 20–49 years without a history of cardiovascular disease. Analyses were stratified by 10-year age intervals at cohort entry. Cox proportional hazards models to predict stroke risk were developed, including traditional cardiovascular factors, and compared with models that additionally included female-specific and psychosocial factors. We compared the risk models using the C-statistic and the slope of the calibration curve at a follow-up of 10 years. We developed an age-specific stroke risk prediction tool that may help communicating the risk of stroke in clinical practice.

Results: We included 409,026 women with a total of 3,990,185 person years of follow-up. Stroke occurred in 2,751 women (incidence rate 6.9 [95%CI:6.6–7.2] per 10,000 person years). Models with only traditional cardiovascular factors performed poorly to moderately in all age groups: 20–29 years: C-statistic: 0.617 (95%CI:0.592–0.639); 30–39 years: C-statistic: 0.615 (95%CI:0.596–0.634); 40–49 years: C-statistic: 0.585 (95%CI:0.573–0.597). After adding the female-specific and psychosocial risk factors to the reference models, the model discrimination increased moderately, especially in the age groups 30–39 (ΔC-statistic: 0.023) and 40–49 years (ΔC-statistic: 0.029) compared to the reference models, respectively.

Conclusion: The addition of female-specific and psychosocial risk factors improves the discriminatory performance of prediction models for stroke in women under 50.

Introduction

Stroke is one of the leading causes of death and disability globally.[1] A decision to start preventive treatment depends first of all on the absolute risk of cardiovascular disease, including stroke and myocardial infarction, over a period of ten years. The current European guidelines recommend the use of the Systematic COronary Risk Evaluation 2 (SCORE2) for estimating cardiovascular risk in the general population.[2,3] This prediction model only includes traditional cardiovascular factors such as age, diabetes, hypertension, cholesterol, and smoking. However, there is increasing evidence that female-specific risk factors for stroke and other cardiovascular diseases, such as migraine, hormonal disorders and preeclampsia, are also important. In a systematic review of cardiovascular risk models in the general population, only two of 160 (1.3%) studies had used female-specific factors.[4] Both studies, and an additional one published a year later, concluded that inclusion of female-specific risk factors did not result in substantial improvement of model discrimination and reclassification.[5,6] However, the primary outcome measure of these studies was a combination of several major cardiovascular events, including myocardial infarction. Female-specific factors, however, primarily increase the risk of stroke.[7,8] Moreover, these studies included mainly postmenopausal women, while female-specific factors such as migraine and oral contraceptives increase the risk of stroke especially at reproductive age.[9,10] Psychosocial factors, such as low socioeconomic status and depression, have also been found to increase the risk of stroke in women to a greater extent in women compared with men.[11-14] However, their added value has hardly been assessed in prediction models for stroke.[4]
The aim of this study was therefore to assess the added value of female-specific and psychosocial factors, compared to traditional cardiovascular factors alone, in predicting the risk of stroke in women under 50 years.

Methods

Data source

We used data from the STIZON database, which directly retrieves data from electronic health records (EHRs) of a large number of primary care providers throughout the Netherlands and covers around 20% of the Dutch population.[15] From the STIZON dataset we only selected women from general practice centers which were situated in catchment areas of hospitals participating in the STIZON network. This allowed for linkage of hospital ICD-9 and ICD-10 diagnoses to primary care data. The STIZON dataset contains ICPC diagnosis codes for clinical entities and medication prescriptions according to the Anatomical Therapeutic Chemical (ATC) Classification System from primary care pharmacies.[16,17] ICD-9 and ICD-10 codes were present for all in-hospital diagnoses that occurred during

follow-up, while ICPC diagnosis codes were in principle available since birth. Inclusion criteria were female sex (as determined by registration in the primary care EHR), age of 20 to 49 years, and subscription to a STIZON general practice center between January 1st 2007 and December 31st 2020 for minimally one year. The first year of subscription was necessary as we defined this as a one year run-in period to assess predictor values such as prescribed medication or assessments of vital parameters. Exclusion criteria were a history of cardiovascular disease prior to baseline, including myocardial infarction, stroke, angina pectoris, peripheral artery disease, heart failure, and transient ischemic attack. Follow-up time started at the end of the one year run-in period, which was on January 1st 2008 or on the first general practice center subscription to the STIZON network after this date. Women were censored at the earliest date of the diagnosis of a major adverse cardiovascular event, death, deregistration with any practice connected to the STIZON network, or last upload of computerised data to the STIZON database (December 31st 2020). Medical ethical approval was not required for the utilization of data in this study.

Outcome definition

The primary outcome of our study was fatal and non-fatal stroke, defined as the presence of an ICD-9, ICD-10 or ICPC code for overall stroke or ischemic or hemorrhagic stroke subtypes specifically (Appendix I).

Traditional cardiovascular factors

Data on the following traditional cardiovascular factors were included for this study: age, smoking (defined as current or former tobacco use), and either an ICD-9, ICD-10 or ICPC diagnosis code or condition-specific ATC medication prescription code for hyperlipidemia, hypertension, and diabetes mellitus (Appendix I). We did not use biomarker measurements such as serum cholesterol, blood pressure and blood glucose because these measurements were missing in the vast majority (> 80%) of women in our research population. The measurement data are likely not missing at random, and in combination with the large extent of missingness, imputation would probably lead to biased imputations.[18] For binary factors such as smoking, in case of absence of the registration of smoking status in the EHR, it was not possible to distinguish between actual or unknown smoking status. Therefore, we considered the absence of the registration of smoking status as the absence of smoking, implying the imputation of zero. Risk factor information was assessed at the start of follow-up, and at the end of the one-year run-in period.

Female-specific factors

The following female-specific factors for stroke were included based on previous literature: migraine, gestational diabetes, preeclampsia, preterm birth (0 vs. ≥ 1),

miscarriage (0 vs. ≥ 1), stillbirth (0 vs. ≥ 1), menstrual irregularity or primary ovarian insufficiency, female infertility (unspecified), hysterectomy in medical history, poor fetal growth or small for gestational age of a women's child, complications during birth (postpartum hemorrhage, intrapartum hemorrhage, umbilical cord complications), hormonal replacement therapy, and combined hormonal contraceptive use.[19,20,21] A female-specific factor was considered present when either an ICD-10, ICD-9 or ICPC diagnosis code, or condition specific ATC medication prescription code was present. The female-specific factors menstrual irregularity and primary ovarian insufficiency were clustered into menstrual irregularity of any cause, since primary ovarian insufficiency is a main cause for menstrual irregularity.[22] The definition of female-specific factors based on these codes can be found in Appendix I.

Psychosocial factors

Based on the literature we selected the following psychosocial factors.[11-14] Socioeconomic status score was derived from the first four postal code digits, using data from the Netherlands Institute for Social Research (SCP) as a standardized measure based on income, education and occupation of the inhabitants.[23] An history of depression or psychotic disorders was defined by either an ICD-10, ICD-9 or ICPC diagnosis code, or ATC code for antidepressant or antipsychotic drug prescriptions.

Statistical analysis

We developed multivariable Cox proportional hazard (PH) regression models for prediction of the risk of stroke. Because previous literature showed significant age-dependent effects of female-specific factors on risk of stroke,[24,25] we stratified all analyses by three 10-year intervals between the ages of 20–49 at baseline (20–29, 30–39, and 40–49 years) to study potential age-dependent effects of female-specific and psychosocial factors. Women from each age interval at baseline could have follow-up extending the 10-year interval, and could therefore potentially contribute to more than one interval. We assessed the potential added value of female-specific factors with respect to the prediction of risk of stroke using a step-wise approach. First, female-specific factors with a prevalence of less than 0.5% in the overall research cohort were excluded. Second, we assessed both the univariate association of each female-specific factor with risk of stroke, and the association between female-specific factors and risk of stroke independent of traditional cardiovascular factors by developing different models with one female-specific factor together with the five traditional cardiovascular factors. For all three age-based strata, we reported both the hazard ratio (HR), adjusted HR (aHR), and 95% confidence interval (CI) of each female-specific factor, and model discrimination and change in model

discrimination that resulted from including each female-specific factor separately. Third, all female-specific and psychosocial factors that occurred in more than 0.5% of the overall research cohort were included in final models per age stratum (Table 2).

We compared model performance using the selection of traditional cardiovascular, female-specific, and psychosocial factors from step three, compared with reference models with traditional cardiovascular factors alone. Model performance was assessed via both model discrimination (C-statistic) and calibration (calibration curve slope, assessed at 10-years of follow-up). Further, we expressed change in C-statistic between reference models and models including female-specific, and psychosocial factors as difference with the reference model relative to the full scale, which follows from the equation below.

$$\frac{C\text{–}statistic\ (new\ model)\ -\ C\text{–}statistic\ (reference\ model)\ -\ 0.5\ (C\text{–}statistic\ base\ value)}{C\text{–}statistic\ (reference\ model)}$$

Performance metrics were internally validated using 100 bootstraps and corrected for optimism using a previously validated method (Harrell's bias correction).[26] We derived empirical confidence intervals by repeating the bootstrap procedure 50 times. We did not take non-cardiovascular death into account as a competing risk because we assessed a cohort of young women at a maximum of 49 years at baseline. In this population, the cumulative incidence of non-cardiovascular death will be very small compared with the entire population, limiting the competing risk effect on the estimation of the risk of stroke. Because our cohort consists of relatively young women, the absolute 10-year risk of stroke will be predominantly under 1%, which is the lower bound of the moderate risk category according to the European Society of Cardiology Prevention guideline for cardiovascular disease.[2] Consequently, no meaningful absolute risk cut-off is available to use for the assessment of model performance using for example the categorical net reclassification index.[27]

To facilitate the interpretation of the absolute 10-year risk predictions of (non-)fatal stroke from our models, we have developed a novel tool based on the previously published cardiovascular risk age tool.[28] The principle of this tool is that i) as a reference, for each age, the absolute risk of stroke is calculated for women without any traditional cardiovascular, female-specific and psychosocial risk factors; ii) For women at a certain age and one or more risk factors, the absolute 10-year risk is compared with the reference to find the corresponding 'stroke risk age', which may be substantially higher than the actual age. We will present two clinical vignettes to illustrate the clinical utility of our stroke risk age tool.

Results

We included 409,026 women, aged 20–49 years, with no history of cardiovascular disease at baseline with a total of 3,990,185 person-years of follow-up. Stroke occurred in 2,751 women over a median of 11 years. The overall incidence rate of stroke was 6.9 (95% CI: 6.6–7.2) per 10,000 person-years, increasing exponentially in the three age groups (Table 1).

Table 1. Incidence rate of stroke per age group

Age group (years at baseline)	Patients (n)	Total follow-up (years)	Events (n)	Incidence rate per 10,000 person years (95% CI)
20–29	128,885	1,145,403	254	2.2 (1.9–2.5)
30–39	136,708	1,340,917	705	5.3 (4.9–5.6)
40–49	143,433	1,503,865	1,792	12 (11–13)
Total	409,026	3,990,185	2,751	6.9 (6.6–7.2)

The prevalence of traditional cardiovascular factors at baseline increased significantly by age group. Hypertension was the most common traditional cardiovascular risk factor (12% in women aged 40–49 years at baseline) and complications during childbirth the most frequent female-specific risk factor (11% in women aged 30–39 years at baseline). Female-specific factors that occurred in less than 0.5% of the entire population were PCOS, gestational diabetes, and history of hysterectomy (Table 2).

The female-specific and psychosocial factors that were independently associated with stroke as traditional cardiovascular factors were, in women aged 20–29 years: irregular menstruation for any cause and complications during childbirth, and hormonal replacement therapy; In women aged 30–39 years: migraine, preeclampsia, complications during childbirth, combined hormonal contraceptive use, socioeconomic status score, and depression; and in women aged 40–49 years: combined hormonal contraceptive use, socioeconomic status score, depression, and psychotic disorder (Table 3).

Model performance of models including only traditional cardiovascular factors was poor to moderate in all age groups: 20–29 years: C-statistic: 0.617 (95% CI: 0.592–0.639); 30–39 years: C-statistic: 0.615 (95% CI: 0.596–0.634); 40–49 years: C-statistic of 0.585 (95% CI: 0.0.573–0.597). The slopes of the calibration curves of the reference models in the three age groups were good: 20–29 years: 0.949 (95% CI: 0.894 - 0.978); 30–39 years: 0.977 (95% CI: 0.952–1.000); 40–49 years: 0.984 (95% CI: 0.962–1.000; Table 4). The addition of female-specific risk factors to the

reference models led to a moderate improvement of model discrimination, especially in the 40–49 year age group (Δ C-statistic: 0.018 compared to reference model, 18.8% difference with the reference model relative to full scale). The addition of psychosocial factors social status score and history of depression further increased the discriminatory performance (Δ C-statistic: 0.019 and 0.029, respectively, compared with reference models, 16.8% and 34.5% difference, respectively, with the reference model compared with the full scale C-statistic, Table 4). The absolute 10-year risks of stroke predicted by the models combining traditional cardiovascular, female-specific and psychosocial factors were generally low, but increased substantially in all age strata (Figure 1).

Figure 2 shows calibration curves of the three models containing traditional cardiovascular, female-specific, and psychosocial risk factors.

Finally, we present two illustrative clinical vignettes based on the prediction models from this study. First, a 33-year-old woman with a history of migraine, who smokes and uses combined hormonal contraceptives, has a mean predicted absolute 10-year risk of stroke of 0.7% (95% CI: 0.4%–1.1%) according to our model. According to our stroke risk age tool, this risk is comparable to that of a 43-year-old woman without any predefined risk factors other than age. Second, a 40-years-old woman with a history of depression and hypertension using combined hormonal contraceptives has a mean predicted absolute 10-year risk of stroke of 1.1% (95% CI: 0.8%–1.4%) in our model, which is similar to the risk of stroke of a 48-year-old woman without any predefined risk factors according to the stroke risk age tool (Figure 3).

Table 2. Baseline characteristics for women in three age groups between 20 and 49 years at baseline with and without stroke

Groups	Baseline characteristic	20–29 years* Stroke (n = 254)	20–29 years* No stroke (n = 128,631)	30–39 years Stroke (n = 705)	30–39 years No stroke (n = 136,003)	40–49 years Stroke (n = 1,792)	40–49 years No stroke (n = 141,641)
	Age, mean (± SD)	25.6 (2.6)	24.6 (2.9)	36.4 (3.0)	35.1 (3.2)	45.5 (3.1)	44.9 (3.1)
Cardiovascular risk factors, *n*	Smoking (ever)	<10 (<3.9)	1345 (1.0)	22 (3.1)	2000 (1.5)	87 (4.9)	3304 (2.3)
	Hyperlipidemia	<10 (<3.9)	346 (0.3)	22 (3.1)	1022 (0.8)	108 (6.0)	3666 (2.6)
	Hypertension	13 (5.1)	3997 (3.1)	73 (10.4)	5893 (4.3)	355 (19.8)	15044 (10.6)
	Diabetes mellitus	<10 (<3.9)	556 (0.4)	15 (2.1)	1176 (0.9)	67 (3.7)	2478 (1.7)
Women-pecific risk factors, *n (%)*	Migraine	<10 (<3.9)	3678 (2.9)	47 (6.7)	5116 (3.8)	106 (5.9)	7316 (5.2)
	Gestational diabetes	<10 (<3.9)	216 (0.2)	<10 (1.4)	734 (0.5)	<10 (0.6)	285 (0.2)
	Preeclampsia	<10 (<3.9)	556 (0.4)	20 (2.8)	1933 (1.4)	<10 (0.6)	645 (0.5)
	Preterm birth ≥ 1	<10 (<3.9)	743 (0.6)	21 (2.9)	2574 (1.9)	10 (0.6)	1214 (0.9)
	Abortion ≥ 1	<10 (<3.9)	2356 (1.8)	32 (4.5)	5617 (4.1)	34 (1.9)	2518 (1.8)
	Menstrual irregularity	12 (4.7)	3182 (2.5)	28 (3.9)	4159 (3.1)	85 (4.7)	5859 (4.1)
	Infertility	<10 (<3.9)	996 (0.8)	11 (1.6)	3459 (2.5)	16 (0.9)	1376 (1.0)
	Hysterectomy	<10 (<3.9)	215 (0.2)	<10 (1.4)	221 (0.2)	<10 (0.6)	239 (0.2)
	Poor fetal growth	<10 (<3.9)	434 (0.3)	<10 (1.4)	1221 (0.9)	10 (0.6)	525 (0.4)
	Complications during birth	22 (8.7)	4858 (3.8)	93 (13.2)	14878 (10.9)	56 (3.1)	5349 (3.8)
	Hormonal replacement therapy	<10 (<3.9)	283 (0.2)	<10 (1.4)	670 (0.5)	33 (1.8)	1658 (1.2)
	Combined hormonal contraceptive use	120 (47.2)	59471 (46.2)	297 (42.1)	46287 (34.0)	606 (33.8)	35950 (25.4)
Psychosocial risk factors	Socioeconomic status score, *mean (± SD)***	0.18 (0.77)	0.19 (0.76)	0.21 (0.77)	0.29 (0.74)	0.24 (0.72)	0.32 (0.68)
	Depression, *n (%)*	19 (7.5)	6056 (4.7)	83 (11.8)	10534 (7.7)	327 (16.3)	15844 (10.2)
	Psychotic disorder, *n (%)*	<10 (<3.9)	1145 (0.9)	<10 (1.4)	1795 (1.3)	57 (3.2)	2189 (1.5)

**Age at baseline*

***Mean socioeconomic status score based on principal component analysis, with higher scores indicating higher socioeconomic status*

Table 3. Univariate association between women-specific risk factors and stroke per age group, and association independent of traditional cardiovascular risk factors

Risk factors		20–29 years		30–39 years		40–49 years	
		HR (95% CI)	aHR* (95% CI)	HR (95% CI)	aHR* (95% CI)	HR (95% CI)	aHR* (95% CI)
Traditional cardiovascular (reference model)	Age	1.13 (1.08–1.18)	NA	1.12 (1.09–1.15)	NA	1.06 (1.04–1.07)	NA
	Hyperlipidemia	6.81 (2.81–16.5)	NA	4.04 (2.64–6.18)	NA	2.40 (1.97–2.91)	NA
	Diabetes mellitus	4.80 (1.98–11.6)	NA	2.54 (1.52–4.23)	NA	2.23 (1.74–2.84)	NA
	Hypertension	1.48 (0.84–2.58)	NA	2.34 (1.83–2.98)	NA	1.99 (1.77–2.24)	NA
	Smoking	2.90 (1.37–6.14)	NA	2.23 (1.46–3.40)	NA	2.12 (1.71–2.64)	NA
Women-specific	Migraine	0.90 (0.42–1.91)	0.79 (0.37–1.69)	1.69 (1.26–2.28)	1.48 (1.10–2.01)	1.11 (0.91–1.35)	0.98 (0.81–1.20)
	Preeclampsia	NA	NA	1.97 (1.26–3.08)	1.83 (1.17–2.85)	1.09 (0.55–2.19)	1.15 (0.57–2.30)
	Preterm birth ≥ 1	0.74 (0.10–5.29)	0.58 (0.08–4.14)	1.54 (0.99–2.40)	1.50 (0.96–2.34)	0.72 (0.39–1.34)	0.79 (0.43–1.48)
	Abortion ≥ 1	1.28 (0.57–2.88)	1.11 (0.49–2.50)	1.06 (0.74–1.51)	1.05 (0.74–1.50)	1.14 (0.81–1.60)	1.26 (0.90–1.78)
	Irregular menstruation	2.14 (1.17–3.91)	1.98 (1.08–3.63)	1.23 (0.83–1.84)	1.14 (0.77–1.71)	1.15 (0.93–1.43)	1.06 (0.85–1.32)
	Infertility	1.58 (0.50–4.92)	1.21 (0.39–3.79)	0.62 (0.34–1.12)	0.61 (0.34–1.10)	1.01 (0.62–1.66)	1.11 (0.68–1.82)
	Complication during birth	2.69 (1.74–4.16)	2.14 (1.37–3.34)	1.29 (1.04–1.61)	1.27 (1.02–1.58)	0.91 (0.70–1.19)	1.00 (0.76–1.30)
	Poor fetal growth	1.33 (0.19–9.48)	1.10 (0.15–7.82)	1.34 (0.67–2.69)	1.30 (0.65–2.60)	1.61 (0.87–3.00)	1.82 (0.98–3.39)
	Hormonal replacement therapy	5.22 (1.67–16.3)	4.48 (1.43–14.1)	0.80 (0.26–2.50)	0.72 (0.23–2.25)	1.49 (1.06–2.10)	1.30 (0.92–1.84)
	Combined hormonal contraceptive use	0.90 (0.70–1.15)	0.92 (0.72–1.18)	1.20 (1.03–1.39)	1.20 (1.03–1.40)	1.37 (1.24–1.51)	1.35 (1.22–1.49)
Psychosocial	Socioeconomic status score	0.93 (0.79–1.09)	0.92 (0.79–1.08)	0.80 (0.73–0.88)	0.80 (0.73–0.88)	0.82 (0.77–0.87)	0.83 (0.78–0.89)
	Depression	1.58 (0.99–2.52)	1.38 (0.86–2.21)	1.51 (1.20–1.90)	1.32 (1.04–1.66)	1.69 (1.49–1.91)	1.52 (1.34–1.72)
	Psychotic disorder	1.40 (0.45–4.38)	1.20 (0.38–3.77)	1.05 (0.54 - 2.03)	0.89 (0.46–1.73)	2.21 (1.70–2.88)	1.96 (1.50–2.55)

**HR = Hazard Ratios; aHR = Hazard Ratios adjusted for age, hyperlipidemia, diabetes mellitus, hypertension and smoking*

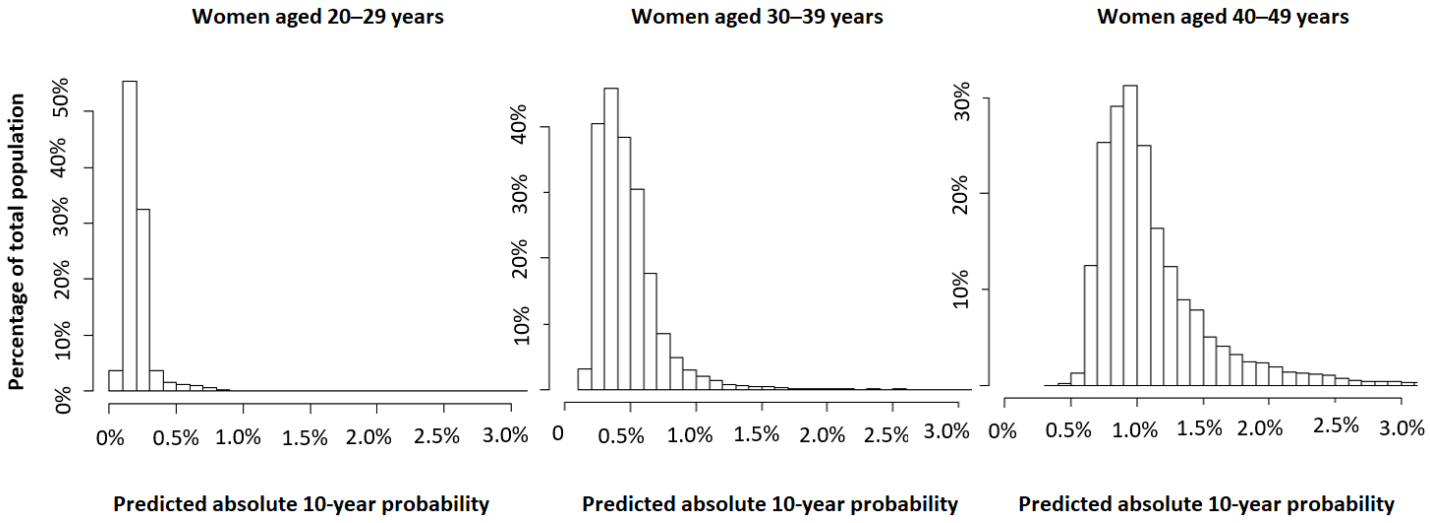
Table 4. Performance of female-specific Cox PH models with different risk factor selections across the three age groups

Age range	Risk factor selections	C-statistic (95% CI)	Δ C-statistic*	Calibration curve slope (95% CI)
20–29 years	Traditional cardiovascular	0.616 (0.592–0.639)	ref.	0.949 (0.894–0.978)
	Traditional + female-specific	0.625 (0.590–0.652)	0.009	0.871 (0.801–0.939)
	Traditional + female-specific + psychosocial	0.632 (0.606–0.660)	0.016	0.868 (0.808–0.920)
30–39 years	Traditional cardiovascular	0.613 (0.595–0.630)	ref.	0.977 (0.952–1.000)
	Traditional + female-specific	0.626 (0.604–0.649)	0.013	0.930 (0.905–0.955)
	Traditional + female-specific + psychosocial	0.636 (0.619–0.663)	0.023	0.937 (0.894–0.960)
40–49 years	Traditional cardiovascular	0.584 (0.573–0.597)	ref.	0.984 (0.962–1.000)
	Traditional + female-specific	0.602 (0.592–0.610)	0.018	0.957 (0.941–0.975)
	Traditional + female-specific + psychosocial	0.613 (0.602–0.625)	0.029	0.959 (0.943–0.970)

**Difference between c-statistices of reference models (traditional cardiovascular risk factors) and models including female-specific and psychosocial risk factors*

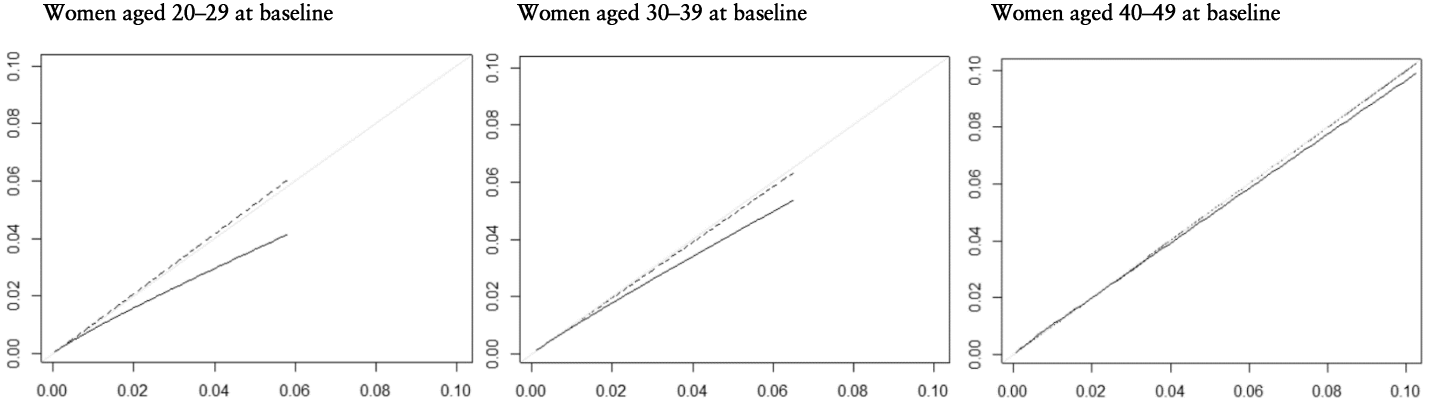
***Difference between C-statistices expressed as difference with the reference model relative to full scale (C-statistic range of 0.5–1.0)*
Model performance metrics were optimism-corrected using 100 boorstraps and empirical confidence intervals were derived by repeating the bootstrapping procedure 50 times

Figure 1. Absolute ten-year risk predictions of female-specific prediction models across the three age ranges



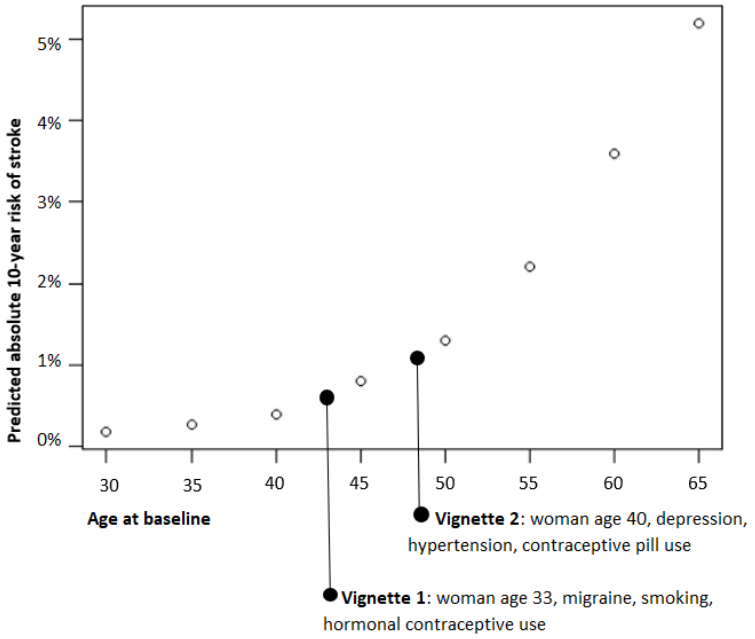
On the X-axis predicted probabilities from optimism-corrected prediction models including traditional cardiovascular, female-specific, and psychosocial risk factors. Predicted probabilities are divided into bins based on increments 0.1%, and on the Y-axis the fraction of the population within each bin is plotted.

Figure 2. Calibration plots of female-specific prediction models across three age ranges



Calibration plots with on the Y-axis the observed, and on the X-axis the predicted probabilities at 10 years of follow-up, for the three models containing traditional cardiovascular, female-specific, and psychosocial risk factors. For each model, two calibration curves are constructed using restricted cubic splines with a Cox proportional hazards model (10 knots): the original model (continuous line) and the optimism-corrected model (dashed line).

Figure 3. Visualization of the stroke risk age tool



This figure shows the graph of predicted absolute ten-year risk of stroke for women without any traditional cardiovascular, female-specific, and psychosocial risk factor levels and age increasing from 30 to 65, and the absolute ten-year risks of women from vignettes 1 and 2 plotted in the same graph.

Discussion

In this study we show (i) that female-specific factors such as migraine, irregular menstruation, complications during childbirth, preeclampsia, hormonal replacement therapy, and combined hormonal contraceptive use, as well as psychosocial risk factors such as social status score and a history of depression or psychotic disorders, are associated with an increased risk of stroke in women aged 20–49 years, (ii) that this association is independent of that caused by traditional cardiovascular risk factors, and (iii) that associations change across the three ten-year age strata. Moreover, addition of these risk factors to prediction models that include only traditional cardiovascular risk factors, substantially increase the predictive performance of models for the prediction of stroke in women aged 20–49 years.

Three studies previously investigated the added value of female-specific risk factors in cardiovascular risk models.[5,6,29] In the Women's Health Initiative Observational Study, pregnancy loss, absence of breastfeeding for ≥ 1 month, and irregular menstruation were independently associated with an increased future risk of cardiovascular events in post-menopausal women.[6] However, adding these factors to the model only modestly improved the c-statistic from 0.726 to 0.730. In a Norwegian study, only preeclampsia remained associated with the risk of cardiovascular events after adjustment for established risk factors (HR: 1.60; 95% CI: 1.16–2.17).[5] The addition of pregnancy complication history to the established prediction model led to small improvements in discrimination (c-statistic difference 0.004, 95% CI 0.002–0.006) and correct reclassification of events (net reclassification improvement 0.02, 95% CI 0.002–0.05). A Swedish study found that low birth weight of a woman's child was associated with cardiovascular events (aHR: 1.68; 95% CI: 1.19–2.37).[29] The addition of a history of hypertensive disorders during pregnancy or low birth weight of the offspring to the traditional cardiovascular risk factors did not meaningfully improve the ten-year prediction of cardiovascular risk in women aged 50 years or older.

Importantly, all of these studies were conducted mainly or exclusively in peri- or postmenopausal women, whereas the stroke risk increasing effect of female-specific risk factors appears to be mainly or only present in young woman. In contrast, our study was conducted in pre-menopausal women and aimed to determine whether female-specific factors had a potential added predictive value for stroke and whether this differed in different age groups. For example, in the study of Kurth et al, [24] migraine only increased the risk of stroke in women aged 45–49 years, but not in older age. In our study, migraine was an independent risk factor in women aged 30–39 years at baseline. However, as the median follow-up time was 11 years, this probably corresponds to a relative risk increase for stroke in the mid-40s age range.

Also for preeclampsia, there is mainly evidence for an increased risk of stroke during the reproductive age (RR: 1.81; 95% CI: 1.45–2.27), which is consistent with our findings in the age group between 30–39 years. However, another study found an increased risk of stroke in women with a history of preeclampsia up to the sixth life decade.[25] This finding contrasts with our study which found no increased risk in women aged 40–49 years. Interestingly, we found a strongly increased risk of stroke in women aged 20–29 who used hormonal replacement therapy. This finding may be confounded by premature ovarian insufficiency, which itself may be underreported in the EHR.[30]

In contrast to these earlier studies, we found a substantial improvement in the discrimination of the stroke prediction models in women aged 30–39 years after adding female-specific and psychosocial factors. This may be explained by a differential effect of female-specific factors on stroke specifically versus general cardiovascular outcomes, the selection of other female-specific factors, the addition of psychosocial factors to our prediction models, or the stratification into three ten-year age groups.

Limitations and strengths

Our study also has limitations. First, there are a number of quality problems with EHR-derived data, in particular the underreporting of clinical conditions.[31,32] For example, the lifetime prevalence of migraine in women is estimated to be around 33%.[33] However, in the Dutch primary care EHR data, on average migraine is only recorded in 2.5% of the general population.[34] In our study, we found an EHR registration for migraine in 4.0% of women younger than 50 years. There are several reasons for the underreporting of migraine. Many patients with migraine do not visit the general practitioner for their migraine[35], and if they do, migraine is probably not always accurately reported in the EHR by the general practitioner.[32] It is probable that patients who do visit the general practitioner have a more severe migraine phenotype, which is more likely to be recorded in the EHR. Because migraine with a high attack frequency has a relatively stronger relation with the risk of ischemic stroke, in our study the association between migraine and stroke could be overestimated.[34] Not only migraine, but also other factors such as smoking (only 3% of women) were underreported in our data. Moreover, although primary care EHR systems have already been widely used since 1990, the quality of the records has increased in recent decades due to improvements in quality control.[35] Therefore, the reporting of female-specific factors related to pregnancy and childbirth may be less accurate in the 40–49 age group than in younger age groups. For the derivation of our prediction models, however, the underreporting of traditional cardiovascular, female-specific and psychosocial factors do not necessarily pose a problem. After all, measurement error (including underreporting) in predictors is unlikely to affect

the generalizability and transportability of our prediction models, if the measurement error is similar in the deployment setting of the models.[36] Second, our reference models included predictors based on the ICPC, ICD-9, ICD-10 or ATC codes for hypertension, hyperlipidemia and diabetes, instead of continuous measurements of blood pressure, cholesterol or serum glucose which are used in most cardiovascular risk prediction models.[36] Not including continuous measurement data in our reference models may have reduced the predictive performance. However, more than 90% of our population lacked measurement data, and the values were probably not missing at random. Therefore, imputation would likely have resulted in biased imputed values.[18] Third, the discriminatory performance of the prediction models in this study is moderate (C-statistics of around 0.61–0.63), but is comparable to prediction models that have been implemented in clinical practice such as CHA_2DS_2-VASc.[37] In addition, because age is by far the most important predictor for the risk of stroke, the restriction of age at baseline to a ten year range also reduces the C-statistic. Further, good model calibration around the absolute risks that are relevant for clinical decisions may be a better indicator for the clinical use of a model compared with discrimination.[38] Fourth, the clinical outcome in our study included both ischemic and hemorrhagic stroke subtypes. Female-specific risk factors may have a differential effect on these two subtypes. In a meta-analysis, migraine had a larger effect on hemorrhagic (aHR: 1.43; 95% CI: 1.03–1.99) than on ischemic stroke (aHR: 1.29; 95% CI: 1.08–1.54.[39] From a clinical utility perspective however, the overall stroke outcome of prediction models may be more practical because in the context of primary prevention no distinction is made between ischemic and hemorrhagic stroke.[2] Fifth, the registration of non-cardiovascular death outside of the hospital in the primary care EHR is known to be relatively incomplete. However, this problem is likely to be limited due to the relatively small fraction of non-cardiovascular deaths in our young population. Sixth, the EHR data on which our study is based do not contain specific information regarding gender. Therefore, we could not discern between cis- and transgender, and gender expansive individuals, and it is unclear whether results can be generalized to transgender and gender expansive individuals

Strengths of our study include the use of the largest dataset to date to study female-specific risk factors in women under 50 and to develop female-specific prediction models. Moreover, in our cohort study, primary care and hospital diagnosis codes were linked. This allowed for a more valid determination of the clinical outcome compared with the use of primary care data alone. Furthermore, by stratifying our population into ten-year age groups, we were able to account for variation in the associations between female-specific risk factors and risk of stroke across the lifespan.

Clinical implications

Although many different prediction models for the risk of cardiovascular events have been developed, female-specific factors or women under 40 years are rarely included.[40] Our study is the first to develop prediction models for stroke risk, including female-specific risk factors specifically in a young population. A challenge in using prediction models for risk of cardiovascular events and stroke in individuals under 50 years is that the predicted absolute ten-year risks are generally very low. In our population these risks were generally lower than 2.5%. The European Society of Cardiology guideline for prevention of cardiovascular disease recommends preventive medication from an absolute ten-year SCORE2 risk of 2.5% and onwards in individuals under 50 years.[2] This, however, does not mean that predicted ten-year risks under 2.5% are irrelevant. The stroke risk age tool developed in this study can help select young women with an absolute risk of stroke that is relatively high due to combinations of female-specific, psychosocial and modifiable cardiovascular risk factors, compared with women without these factors. Currently, a lack of risk awareness is a major factor contributing to the lack of preventive measures and healthy lifestyle choices among women.[41] These women could be proactively advised to eliminate modifiable risk factors early in life to prevent cardiovascular events and other diseases such as dementia.[28] [42,43] Moreover, in younger women, female-specific risk factors often precede the occurrence of traditional cardiovascular risk factors – for example, preeclampsia preceding the occurrence of hypertension.[44] After successful external validation, the stroke risk age tool can be used to counsel women in clinical practice. Moreover, all risk factors used in our models are based solely on medical history present in the primary care EHR. Therefore simple, non-invasive, relatively inexpensive, and even fully automated population stratification procedures can be performed to proactively identify and screen at-risk women. Based on our results, we advise healthcare professionals – and especially general practitioners – to take female-specific and psychosocial factors into account for the estimation of the risk of stroke, and to invest in the quality of registrations of these factors in the EHR. Importantly, it is likely that psychosocial factors 'depression' and 'psychotic disorders' are indicators for social determinants of health, which could practically not have been retrieved from the EHR. Therefore, in the implementation phase of prediction models that utilize these factors, we should invest in education of all end users to prevent any form of stigmatization.

Conclusion

The addition of female-specific and psychosocial risk factors to traditional cardiovascular predictors improves the discriminatory performance of prediction models for women under age 50. Our newly developed stroke risk age tool can help discuss stroke risk in clinical practice.

References

1. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Cheng S, Delling FN, et al. Heart Disease and Stroke Statistics-2021 Update: A Report From the American Heart Association. *Circulation*. 2021;143:e254-e743.
2. Visseren FLJ, Mach F, Smulders YM, Carballo D, Koskinas KC, Back M, Benetos A, Biffi A, Boavida JM, Capodanno D, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice. *Eur Heart J*. 2021;42:3227-3337.
3. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, Himmelfarb CD, Khera A, Lloyd-Jones D, McEvoy JW, et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation*. 2019;140:e563-e595.
4. Baart SJ, Dam V, Scheres LJJ, Damen J, Spijker R, Schuit E, Debray TPA, Fauser B, Boersma E, Moons KGM, et al. Cardiovascular risk prediction models for women in the general population: A systematic review. *PLoS One*. 2019;14:e0210329. doi: 10.1371/journal.pone.0210329
5. Markovitz AR, Stuart JJ, Horn J, Williams PL, Rimm EB, Missmer SA, Tanz LJ, Haug EB, Fraser A, Timpka S, et al. Does pregnancy complication history improve cardiovascular disease risk prediction? Findings from the HUNT study in Norway. *Eur Heart J*. 2019;40:1113-1120.
6. Parikh NI, Jeppson RP, Berger JS, Eaton CB, Kroenke CH, LeBlanc ES, Lewis CE, Loucks EB, Parker DR, Rillamas-Sun E, et al. Reproductive Risk Factors and Coronary Heart Disease in the Women's Health Initiative Observational Study. *Circulation*. 2016;133:2149-2158.
7. Adelborg K, Szepligeti SK, Holland-Bill L, Ehrenstein V, Horvath-Puho E, Henderson VW, Sorensen HT. Migraine and risk of cardiovascular diseases: Danish population based matched cohort study. *BMJ*. 2018;360:k96.
8. Kurth T, Winter AC, Eliassen AH, Dushkes R, Mukamal KJ, Rimm EB, Willett WC, Manson JE, Rexrode KM. Migraine and risk of cardiovascular disease in women: prospective cohort study. *BMJ*. 2016;353:i2610.

9. Lidegaard O, Lokkegaard E, Jensen A, Skovlund CW, Keiding N. Thrombotic stroke and myocardial infarction with hormonal contraception. *N Engl J Med*. 2012;366:2257-2266.
10. Schurks M, Rist PM, Bigal ME, Buring JE, Lipton RB, Kurth T. Migraine and cardiovascular disease: systematic review and meta-analysis. *BMJ*. 2009;339:b3914.
11. Li M, Fan YL, Tang ZY, Cheng XS. Schizophrenia and risk of stroke: a meta-analysis of cohort studies. *Int J Cardiol*. 2014;173:588-590.
12. Peters S, Carcel, C., Millett, E., & Woodward, M. (2020). Sex differences in the association between major risk factors and the risk of stroke in the UK Biobank cohort study. Neurology, 95(20), e2715–e2726.
13. Salaycik KJ, Kelly-Hayes M, Beiser A, Nguyen AH, Brady SM, Kase CS, Wolf PA. Depressive symptoms and risk of stroke: the Framingham Study. *Stroke*. 2007;38:16-21.
14. Seifert CL, Poppert H, Sander D, Feurer R, Etgen T, Ander KH, Purner K, Bronner M, Sepp D, Kehl V, et al. Depressive symptoms and the risk of ischemic stroke in the elderly--influence of age and sex. *PLoS One*. 2012;7:e50803.
15. Kuiper JG, Bakker M, Penning-van Beest FJA, Herings RMC. Existing Data Sources for Clinical Epidemiology: The PHARMO Database Network. *Clin Epidemiol*. 2020;12:415-422.
16. Lamberts H. WM. Oxford University Press; USA: 1987. ICPC, international classification of primary care.
17. WHO. Collaborating Centre for Drug Statistics Methodology. ATC index with DDDs. Oslo; Norway. 2002
18. Beaulieu-Jones BK, Lavage DR, Snyder JW, Moore JH, Pendergrass SA, Bauer CR. Characterizing and Managing Missing Structured Data in Electronic Health Records: Data Analysis. *JMIR Med Inform*. 2018;6:e11.
19. Appelman Y, van Rijn BB, Ten Haaf ME, Boersma E, Peters SA. Sex differences in cardiovascular risk factors and disease prevention. *Atherosclerosis*. 2015;241:211-218.
20. Poorthuis MH, Algra AM, Algra A, Kappelle LJ, Klijn CJ. Female- and Male-Specific Risk Factors for Stroke: A Systematic Review and Meta-analysis. *JAMA Neurol*. 2017;74:75-81.
21. Zhou Y, Wang X, Jiang Y, Ma H, Chen L, Lai C, Peng C, He C, Sun C. Association between polycystic ovary syndrome and the risk of stroke and all-cause mortality: insights from a meta-analysis. *Gynecol Endocrinol*. 2017;33:904-910.
22. Wiksten-Almströmer M HA, Hagenfeldt K. . Menstrual disorders and associated factors among adolescent girls visiting a youth clinic. Acta Obstet Gynecol Scand. 2007;86(1):65-72.

23. Sociaal Cultureel Planbureau, www.scp.nl/Onderzoek/Lopend_onderzoek/A_Z_alle_lopende_onderzoeken/Statusscores, (Updated).
24. Kurth T, Schurks M, Logroscino G, Gaziano JM, Buring JE. Migraine, vascular risk, and cardiovascular events in women: prospective cohort study. *BMJ.* 2008;337:a636.
25. Bellamy L, Casas JP, Hingorani AD, Williams DJ. Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis. *BMJ.* 2007;335:974.
26. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* 1996;15:361-387.
27. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, Cooney MT, Corra U, Cosyns B, Deaton C, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice. Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). *Eur Heart J.* 2016;37:2315-2381.
28. Cooney MT, Vartiainen E, Laatikainen T, De Bacquer D, McGorrian C, Dudina A, Graham I, Score, investigators F. Cardiovascular risk age: concepts and practicalities. *Heart.* 2012;98:941-946.
29. Timpka S, Fraser A, Schyman T, Stuart JJ, Asvold BO, Mogren I, Franks PW, Rich-Edwards JW. The value of pregnancy complication history for 10-year cardiovascular disease risk prediction in middle-aged women. *Eur J Epidemiol.* 2018;33:1003-1010.
30. Roeters van Lennep JE, Heida KY, Bots ML, Hoek A, collaborators of the Dutch Multidisciplinary Guideline Development Group on Cardiovascular Risk Management after Reproductive D. Cardiovascular disease risk in women with premature ovarian insufficiency: A systematic review and meta-analysis. *Eur J Prev Cardiol.* 2016;23:178-186.
31. Spasoff RA. Epidemiologic Methods for Health Policy. New York: Oxford University Press I.
32. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JP. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc.* 2017;24:198-208.
33. Launer LJ, Terwindt GM, Ferrari MD. The prevalence and characteristics of migraine in a population-based cohort: the GEM study. *Neurology.* 1999;53:537-542.

34. Nielen MMJ VDMV, Schellevis FG: Evaluatie pilot PreventieConsult cardiometabool risico, Report in Dutch. Utrecht: NIVEL; 2010. http://www.nivel.nl/pdf/Rapport-Evaluatie-pilot-PreventieConsult., pdf.
35. Lipton RB, Scher AI, Kolodner K, Liberman J, Steiner TJ, Stewart WF. Migraine in the United States: epidemiology and patterns of health care use. *Neurology*. 2002;58:885-894.
36. Score Working Group. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur Heart J*. 2021;42:2439-2454.
37. Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest*. 2010;137:263-272.
38. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21:128-138.
39. Mahmoud AN, Mentias A, Elgendy AY, Qazi A, Barakat AF, Saad M, Mohsen A, Abuzaid A, Mansoor H, Mojadidi MK, et al. Migraine and the risk of cardiovascular and cerebrovascular events: a meta-analysis of 16 cohort studies including 1 152 407 subjects. *BMJ Open*. 2018;8:e020498.
40. Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, Tzoulaki I, Lassale CM, Siontis GC, Chiocchia V, Roberts C, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ*. 2016;353:i2416.
41. Oertelt-Prigione S, Seeland U, Kendel F, Rucke M, Floel A, Gaissmaier W, Heim C, Schnabel R, Stangl V, Regitz-Zagrosek V. Cardiovascular risk factor distribution and subjective risk estimation in urban women--the BEFRI study: a randomized cross-sectional study. *BMC Med*. 2015;13:52.
42. Graham IM, Di Angelantonio E, Visseren F, De Bacquer D, Ference BA, Timmis A, Halle M, Vardas P, Huculeci R, Cooney MT, et al. Systematic Coronary Risk Evaluation (SCORE): JACC Focus Seminar 4/8. *J Am Coll Cardiol*. 2021;77:3046-3057.
43. Seshadri S, Wolf PA. Lifetime risk of stroke and dementia: current concepts, and estimates from the Framingham Study. *Lancet Neurol*. 2007;6:1106-1114.
44. Ghossein-Doha C, Spaanderman M, van Kuijk SM, Kroon AA, Delhaas T, Peeters L. Long-Term Risk to Develop Hypertension in Women With Former Preeclampsia: A Longitudinal Pilot Study. *Reprod Sci*. 2014;21:846-853.