Serveur Académique Lausannois SERVAL serval.unil.ch

Author Manuscript
Faculty of Biology and Medicine Publication

This paper has been peer-reviewed but dos not include the final publisher proof-corrections or journal pagination.

Published in final edited form as:

serval
serveur académique lausannois

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Predicting stroke through genetic risk functions: The CHARGE risk score project

Carla A Ibrahim-Verbaas, MD[*], **Myriam Fornage, PhD**[*], **Joshua C Bis, PhD**[*], **Seung Hoan Choi, MA**[*], **Bruce M Psaty, MD, PhD**, **James B Meigs, MD, MPH**, **Madhu Rao, MD, PhD**, **Mike Nalls, PhD**, **Joao D Fontes, MD**, **Christopher J. O'Donnell, MD, PhD**, **Sekar Kathiresan, MD, PhD**, **Georg B. Ehret, MD, PhD**, **Caroline S Fox, MD, MPH, PhD**, **Rainer Malik, PhD**, **Martin Dichgans, MD, PhD**, **Helena Schmidt, MD, PhD**, **Jari Lahti, PhD**, **Susan R Heckbert, MD, PhD**, **Thomas Lumley, PhD**, **Kenneth Rice, PhD**, **Jerome I Rotter, PhD**, **Kent D Taylor, PhD**, **Aaron R Folsom, MD, PhD**, **Eric Boerwinkle, PhD**, **Wayne D Rosamond, MD, PhD**, **Eyal Shahar, MD, MPH**, **Rebecca F. Gottesman, MD, PhD**, **Peter J Koudstaal, MD, PhD**, **Najaf Amin, PhD**, **Renske G. Wieberdink, MD, PhD**, **Abbas Dehghan, MD, PhD**, **Albert Hofman, MD, PhD**, **André G Uitterlinden, MD, PhD**, **Anita L. DeStefano, PhD**, **Stephanie Debette, MD, PhD**, **Luting Xue, PhD**, **Alexa Beiser, PhD**, **Philip A. Wolf, MD**, **Charles DeCarli, MD**, **M. Arfan Ikram, MD, PhD**[*], **Sudha Seshadri, MD**[*], **Thomas H Mosley Jr, PhD**[*], **WT Longstreth Jr, MD**[*], **Cornelia M van Duijn, PhD**[*], and **Lenore J Launer, PhD**[*]

Departments of Epidemiology (CAIV, PJK, NA, RGW, AD, AH, AGU, CvD), Neurology (CAIV, PJK,RGW, MAI), Internal Medicine (AGU) and Radiology (MAI), Erasmus University Medical Center, Rotterdam, The Netherlands; Center for Medical Systems Biology, Leiden, the Netherlands (CAIV, NA, CvD); Institute for Molecular Medicine (MF) and Human Genetics Center (MF, EB), University of Texas Health Science center at Houston, Houston, TX, USA; Cardiovascular Health Research Unit (JCB, BMP) and Departments of Medicine (JCB, BMP), Epidemiology (BMP, SRH, WTL), Health Services (BMP), Biostatistics (KR) and Neurology (WTL), University of Washington, Seattle, WA, USA; Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA (BMP); Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA (SHC, ALdeS, SD,LX,AB,PAW); Department of Neurology(SHC, ALDeS, SD, LX, AB, PAW, SS), and Cardiology section, Whitaker Cardiovascular Institute (JDF) Boston University School of Medicine, Boston, MA, USA; The National Heart, Lung and Blood Institute's Framingham Heart Study, Framingham, MA, USA (SHC, JDF, CO'D, CSF, ALDeS, SD, LX, AB, PAW, SS); Department of Medicine, Harvard Medical School General Medicine Division (JBM), Cardiovascular Research Center and Cardiology Division (SK) and Center for Human Genetic Research (SK), Massachusetts General Hospital, Boston, MA, USA; Division of Nephrology/Tufts Evidence Practice Center, Tufts University School of Medicine, Tufts Medical Center, Boston, MA, USA (MR); Laboratory of Neurogenetics (MN) and Laboratory of Epidemiology and Population Sciences(LJL), National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; Cardiology section, Whitaker Cardiovascular Institute, Boston University School of Medicine, Boston, MA, USA; Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), Cambridge, MA, USA (SK); Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine (GBE) and Department of Neurology (RFG), Johns Hopkins University School of Medicine, Baltimore, MD, USA; Institute of Social and

---

Corresponding author: Lenore J. Launer, Laboratory of Epidemiology and Population Sciences, National Institute on Aging, 7201 Wisconsin Avenue, Gateway Building suite C-309, Bethesda, MD, USA, launerl@nia.nih.gov, fax 301-496-4006, telephone 301-496-1178.
All authors marked with contributed equally

Preventive Medicine (IUMSP), Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, and Cardiology, Department of Specialties of Internal Medicine, Geneva University Hospital, Geneva, Switzerland (GBE); Center for Population Studies of the National Heart, Lung, and Blood Institute, Bethesda, MD, USA (CSF); Endocrinology Division, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA (CSF); Institute for Stroke and Dementia Research (ISD), Klinikum der Universität München, Ludwig-Maximilians-Universität, Munich, Germany (RM, MD); Institute of Molecular Biology and Biochemistry, Centre for Molecular Medicine, Medical University of Graz, Graz, Austria (HS); Institute of Behavioral Sciences, University of Helsinki, and Folkhälsan Research Centre, Helsinki, Finland Helsinki, Finland (JL); Department of Statistics, University of Auckland, Auckland, New Zealand (TL); Medical Genetics Institute Cedars-Sinai Medical Center, Los Angeles, CA, USA (JIR, KDT); Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA (ARF); Department of Epidemiology, School of Medicine, University of North Carolina, Chapel Hill, NC, USA (WDR); Division of Epidemiology and Biostatistics, University of Arizona, Tucson, AZ, USA (ES); Netherlands Consortium for Healthy Ageing, Leiden, The Netherlands (AH, AGU, MAI, CvD); Department of Epidemiology, University of Versailles, St-Quentinen-Yvelines, Department of Neurology, Lariboisière Hospital, and Department of Neuroepidemiology, Institut National de la Santé et de la Recherche Médicale (INSERM), Paris, France (SD); Department of Neurology and Neuroscience, University of California at Davis, Sacramento, CA, USA (CDeC); Department of Medicine and Neurology, University of Mississippi Medical Center, Jackson, MS, USA (THM)

Abstract

Background and Purpose—Beyond the Framingham Stroke Risk Score (FSRS), prediction of future stroke may improve with a genetic risk score (GRS) based on Single nucleotide polymorphisms (SNPs) associated with stroke and its risk factors.

Methods—The study includes four population-based cohorts with 2,047 first incident strokes from 22,720 initially stroke-free European origin participants aged 55 years and older, who were followed for up to 20 years. GRS were constructed with 324 SNPs implicated in stroke and 9 risk factors. The association of the GRS to first incident stroke was tested using Cox regression; the GRS predictive properties were assessed with Area under the curve (AUC) statistics comparing the GRS to age sex, and FSRS models, and with reclassification statistics. These analyses were performed per cohort and in a meta-analysis of pooled data. Replication was sought in a case-control study of ischemic stroke (IS).

Results—In the meta-analysis, adding the GRS to the FSRS, age and sex model resulted in a significant improvement in discrimination (All stroke: Δjoint AUC =0.016, p-value=2.3*10-6; IS: Δ joint AUC =0.021, p-value=$3.7*10^{-7}$), although the overall AUC remained low. In all studies there was a highly significantly improved net reclassification index (p-values $<10^{-4}$).

Conclusions—The SNPs associated with stroke and its risk factors result only in a small improvement in prediction of future stroke compared to the classical epidemiological risk factors for stroke.

Keywords

Introduction

Stroke is a major and debilitating neurological disease that increases in frequency with age; it is estimated that in 2030, 23 million persons will have a first-ever stroke resulting in 7.8 million deaths[1, 2]. Stroke is a complex disease with many modifiable risk factors, and a substantial genetic component, with heritability estimates varying from 17%[3] to 38%[4]. The genetic architecture of stroke has been difficult to unravelAlthough recently some findings have been replicated for specific stroke subtypes,[5, 6] initial discoveries of genetic variants from genome-wide association studies (GWAS) of all stroke (sub-types combined) outcome[7] have failed to replicate. This has led to the concept that different genes may be involved in different subtypes of stroke.

In contrast, many modifiable clinical and epidemiological risk factors consistently have been shown to increase the risk for stroke, and also have well replicated risk-associated genetic variants (single nucleotide polymorphisms – SNPs). Several modifiable risk factors have been combined into validated clinical prediction tools such as the Framingham Stroke Risk Score (FSRS), which incorporates systolic blood pressure, diabetes mellitus, cigarette smoking, prior cardiovascular disease, atrial fibrillation, left ventricular hypertrophy, and the use of antihypertensive medications[8, 9]. The FSRS measures traits that may fluctuate in the short or medium term, thus affecting its predictive properties in any one individual. Insufficiency of the FSRS has been demonstrated in earlier studies[10]. Given the increasing availability of genotyping technology and the promise of using the information in a more personalized medicine approach, it is timely to investigate whether risk scores incorporating genetic information will add to the power to predict an individual's future risk for stroke.

Here we examine the predictive properties of a genetic risk score (GRS) to predict future stroke in community dwelling stroke-free individuals. We hypothesized that the combined effect of individual SNPs with small effects would improve prediction. A previous study by Kathiresan et al. found a GRS based on SNPs from a single class of risk factors, lipids, did have some value in reclassification, but not in improved discrimination of persons at future risk for cardiovascular disease[11]. However, cardiovascular diseases such as stroke have a complex pathophysiology, which can partially be accounted for in GRS, as they are in clinical risk scores. Here, we take the approach of including in a risk score, genetic variants associated with stroke *and* its multiple risk factors, with the goals of: assessing the potential of a score, based on SNPs associated with stroke and its risk factors, to predict stroke in general populations; and investigating whether the score could potentially add to the predictability of a score based on established stroke clinical and epidemiological risk factors. As far as we know, we are the first to try to combine not only a disease specific or risk factor specific set of SNPs into a risk score, but a comprehensive set of risk SNPs from the whole spectrum of non-behavioral risk factors for stroke. We also investigated the performance of the GRS in a higher risk population captured in a clinic-based case-control study of ischemic stroke (IS).

Materials and methods

Our analyses are based on incident cases and stroke-free participants characterized in 4 cohorts participating in the Cohorts for Heart and Ageing Research in Genomic Epidemiology (CHARGE) consortium. CHARGE is a large consortium of major population-based prospective cohort studies of cardiovascular health that aims to identify new genetic variants for multiple quantitative, sub- and clinical factors contributing to health and disease in older persons[12]. The individual cohorts and the combined CHARGE genome wide association study of stroke genes have been previously described[7].

Cohorts and case definition

This analysis is based on the following CHARGE cohorts: the Atherosclerosis Risk in Communities (ARIC) study[13], the Cardiovascular Health Study (CHS)[14], the Framingham Heart Study (FHS)[15, 16] and the first cohort of the Rotterdam Study (RS)[17]. From these cohorts, we included persons who were stroke-free at the age of 55 or older, of European descent, and who had complete outcome and genotype data. (Table 1, Supplemental Table I). For all cohorts, the baseline was established in the late 1980's and early 1990's, and all studies are ongoing. All participants provided informed consent and all studies were approved by their governing institutional review boards.

All cohorts defined stroke as a focal neurological deficit of presumed vascular cause with a sudden onset and lasting for at least 24 hours or until death if the participant died less than 24 hours after the onset of symptoms. All suspected events were adjudicated by stroke experts who reviewed medical records, death certificates, imaging studies, or some combination of these sources. We report on "All" stroke, which includes ischemic, hemorrhagic, and unknown sub-type, and separately on ischemic stroke, which is of presumed cardio-embolic/large vessel/small vessel origin. Subarachnoid hemorrhages were excluded from all analyses.

Genotyping

Each study separately genotyped or imputed SNPs to the same reference panel (see Supplemental Table II for methods) and provided data on imputation quality. Due to imputation, there were no missing genotypes in the datasets. Genotypes for each SNP were coded in terms of the number of risk alleles.

Identifying risk factors, associated SNPs, and selecting SNPs for inclusion in the risk score

SNP selection—Based on a literature review, as well as clinical and neurological expert opinion, we identified 9 domains of established risk factors for stroke that have also been studied in GWAS: high blood pressure, atherosclerosis, arrhythmia, diabetes, inflammation, blood constituents, hematologic changes, obesity, elevated lipids, and impaired kidney function. Within each of these risk factor domains we identified 3–5 traits that contribute to the overall domain (Table 2), resulting in a total of 33 traits. For each of the 33 traits we identified from published, in press, and under review genome-wide association studies, SNP variants that associated with the trait at the standard GWAS significance level (p-value $5*10^{-8}$) and for which there was evidence of independent replication (see Supplemental Table III for a complete list of references). We also included two ischemic stroke-associated SNPs identified and replicated by the International Stroke Genetics Consortium[5]. Although these 4 CHARGE cohorts were often included in the above referenced trait GWAS, in general the meta-analyses were based on many more subjects, with the proportion of subjects from these studies ranging between 0[18] and ~86%[19].

In total we identified 334 autosomal SNPs for 34 traits (including stroke). When several SNPs for a class of risk factors were located in the same loci, we selected the top SNP with the lowest p–value, from the most recent meta-analyses that included the largest number and thus had the largest power and highest precision. In some cases the same SNP was associated with multiple traits. For this, we assigned the SNP to the clinical rather than the sub-clinical trait (for example, diabetes over fasting glucose levels). After these exclusions we included 324 SNPs (see Supplemental Table III), including 28 traits with multiple SNPs and 6 with single SNPs.

Statistics: Construction of the risk score—The construction of the risk score (Figure 1) is described in detail in the Supplemental Methods. In short, weighted risk scores were created for each of the 28 traits with multiple SNPs. Weights were the effects, or log odds ratios, of the risk allele on the outcome it was originally associated to. This effect was multiplied by the number of risk alleles (allele dosage: 0,1 or 2) the individual carried; for imputed SNPs, where the dosage is estimated, this value can obtain any fraction between 0 and 2. The risk score was the sum of the weight*dosage product for all SNPs within the trait. For the traits that had only one SNP associated (singletons), the allele dosage was used. Risk scores and singletons entered as covariates in Cox regressions with age as the time-to-event variable. Each cohort calculated a modified FSRS, which differed from the original FSRS only in the omission of the age variable.

Analysis

The individual risk predictions were calculated from five regression models: model 1 included sex only, model 2 included the GRS only, model 3 included sex and the GRS, model 4 included sex and the mFSRS, and model 5 included sex, the GRS and the mFSRS. CHS models also included a variable for study site and the FHS made adjustments for family relationships[15, 16].

Calculation and comparison of the ROC curves

Receiver-operated characteristic (ROC) curves with corresponding areas under the curve (AUC) [95% confidence interval for each model] were created from the predicted risks derived from the regression models. The model AUCs were tested against the null model (AUC=0.50) and compared to each other. The latter comparisons were tested with the Hanley-McNeil test for comparison of correlated AUCs[20]. The Hanley-McNeil test requires a correlation between the AUCs, which was obtained using the metacor package for R[21]. Finally, we estimated meta-AUCs [95% confidence intervals] based on all cohorts combined using an inverse variance weighted meta-analysis.

Reclassification statistics

In secondary analyses we assessed the added clinical value of the genetic risk score over the FSRS using two statistics: the net reclassification improvement (NRI) without cut-off values (continuous NRI), which gives a summary of the number of subjects with reclassified predicted case status based on the new score; and the integrated discrimination improvement (IDI)[22, 23], which describes the ability of the score to discriminate between cases and non-cases.

External replication

We investigated the predictive power of the same GRS in a German case-control set of IS previously described.[5, 6] Cases (N=985) were the German samples of the Wellcome Trust Case Control Consortium 2 (WTCCC2), recruited in the Klinikum Grosshadern, Department of Neurology, Ludwig-Maximilians-University in Munich, Germany. Controls (N=596) were participants in the KORAgen study, residing in the Augsburg area in Germany (http://www.gsf.de/kora/en/english.html), with no history of stroke or transient ischemic attack. The studies were approved by the respective local Institutional Review Boards and all subjects gave informed consent. Details on genotyping and imputation are given in Supplemental Table II. To account for the case-control design we used a logistic regression model, adjusted for age and sex to estimate the trait scores and their association with stroke. As details on the elements of the FSRS were not generally available for the subjects in this cohort, we were unable to perform replication efforts for the FSRS comparisons in this cohort.

Results

During follow-up, 2047 participants from the 22,720 developed first ever strokes, including 1523 ischemic strokes. The RS (N of cases=783, 13.5%) and CHS (N of cases=560, 17.1%) participants were older and had more incident stroke events (Table 1) than the FHS (N of cases=206, 4.8%) and ARIC (N of cases=498, 5.3%). participants. These cohorts also had a higher prevalence of hypertension than the other two. Most smokers were found in ARIC and RS (Supplemental Table I). The number of risk alleles had similar distribution in all cohorts (Supplemental Figure I) reflecting the population-based study designs of all cohorts.

Predicted risks from the models including the GRS and/or mFSRS were significantly higher in cases than in individuals who remained stroke-free (p-value <0.001, Supplemental Table IV). Across cohorts, the AUC including only the GRS (Table 3, Supplemental Figure II) ranged from 0.563 to 0.617 for all stroke. When combining the findings of the cohorts, the meta-AUC was 0.578 (p-value=$9*10^{-10}$ compared to the model with only sex). For the model with sex and the GRS, the all stroke meta-AUC was 0.572, which was statistically different from the model including only sex (p-value=$9*10^{-18}$). For IS, the AUC of the model including only the GRS, ranged from 0.585 to 0.627 across cohorts; the meta-AUC was 0.592. For the combined model with sex and the GRS, meta-AUC was 0.597 (p-value compared to sex only = $2*10^{-19}$). (Table 4, Supplemental Figure II)

Comparison between the genetic and clinical risk scores

In all cohorts, the AUC for the mFSRS was higher than those seen for the GRS. For all stroke, the mFSRS AUC ranged from 0.587 in the RS to 0.709 in the FHS, with a highly significant meta-AUC of 0.621 (Table 3, Supplemental Figure III). In all cohorts, there was a low correlation between the absolute predicted risks derived from mFSRS and those derived from the genetic risk score (meta-correlation 0.012, meta p-value 0.13). The full model (sex, mFSRS and the GRS) improved prediction significantly in all cohorts except the FHS, the cohort that was used to develop the original FSRS. The meta-AUC for the full model was 0.637. When the full model is compared to the model with sex and mFSRS, the GRS gives an improvement of 0.016 (p-value=$2*10^{-6}$) over classical risk factors. For IS, the meta-AUC for the full model was 0.654, with an improvement of 0.021 (p-value=$4*10^{-7}$) between the full model and the sex and mFSRS model (Table 4, Figure 2). Compared to the model based on sex and the FSRS, the GRS yielded a significant NRI (improvement ranging from 0.18 to 0.32 for all stroke and from 0.24 to 0.28 for IS, *p*-values $1.1*10^{-4}$) and IDI (improvement ranging from 0.005 to 0.02 for all stroke and from 0.008 to 0.021 for ischemic stroke, p-values $5*10^{-5}$) (Supplemental Table V).

Replication

We replicated our findings in the case-control sample of the WTCCC2. The AUC of the sex, age and GRS model was higher than the model with age and sex alone (difference between the two models = 0.014 (p-value 0.04); Supplemental Table VI, Supplemental Figure IV). The reclassification statistics (continuous NRI = 0.309, p-value < $1*10^{-5}$, and IDI = 0.018, p-value < $1*10^{-5}$) showed a small but highly significant improvement in prediction of ischemic stroke when the GRS was added to models based on age and sex.

Discussion

We assessed the predictive properties of a genetic risk score based on stroke risk factors in a population-based sample of 2,047 well-characterized incident stroke cases among 22,720 initially stroke-free individuals. We found that a genetic risk score that included genome-wide associated SNPs for 9 domains of risk factors plus stroke provided a small but very

significant and consistent improvement in the prediction of an individual's risk for future stroke. This small difference was observed when the GRS was compared to sex adjusted models as well as to a widely used clinical-epidemiological risk score. Similar results were found when the score was applied in a large clinic-based case control sample.

Age is an important predictor of many diseases, in particular a late onset outcome as stroke. A large part of the discrimination of a prediction model is therefore determined by age, biasing the interpretation of the value of the other variables, in our case genetic predictors. In this study, we controlled for age as the follow-up time so we could clearly assess whether genetic information could improve prediction conditional on the person's age It is important to note that the AUCs are therefore in general lower than reported in the literature. For instance, in the Rotterdam Study, we find much lower AUCs in the Cox approach with age as the timescale (AUC for FSRS=0,59; p-value=$1*10^{-14}$) than in the logistic model where age is a covariate, essentially examining the GRC with all subjects of the mean age (AUC for FSRS =0.65; p-value=$2*10^{-39}$).

The GRS performs significantly worse than the age+sex model in the replication, but this was not the primary comparison we wished to test. Our goal was to test whether the GRS has added value over and above the clinical risk factors, and we do indeed find modest (Δ AUC (discovery) = 0.021, Δ AUC (replication) = 0.024) but significant (p-value (discovery) = $3.66*10^{-7}$, p-value (replication) = 0.03) (Table 4 & Supplemental Table 6) improvement in the AUCs in both the discovery and replication analyses. The difference in p-values most likely reflects the differences in sample size of the discovery and replication cohorts.

Before interpreting the findings it is important to note the assumptions underlying the construction of the score and consider the limitations: 1) The relationships between variants and risk factors, and between risk factor and outcome are always the same, ie (log)linear ; 2) Each variant associated with the *stroke risk factor*, will also be a risk variant for stroke; 3) The effect is proportional to the number of risk variants (0, 1, or 2 copies) (Whereas in reality, the variant may have dominant or recessive effect on stroke); 4) There is no interaction between genetic variants (we did not find evidence for interaction between the loci in this study). Each of these assumptions is necessary when creating such a score, but they may be simplifications of the true underlying biological model.

Previous studies have examined the association of clinical disease to SNPs that associate with risk factors for the disease. For example studies by Ehret et al and Wain et al[24, 25] showed genetic variants associated with blood pressure traits were also associated with stroke. Paynter et al.[26] assessed the performance of a GRS based on a small number of SNPs associated to cardiovascular disease (CVD) to predict CVD outcomes including stroke. They did not find their genetic risk score improved prediction over the traditional risk factors. A similar conclusion has been reached by other investigators examining risk SNPs for a single class of risk factors[11, 27, 28], although recently some authors have found risk differences between GRS quintiles for stroke[29, 30]. Here, we present a different approach: we test a single outcome (stroke) to a risk score based on SNPs associated with a variety of risk factors. This approach takes into account the complexity of disease and our results suggest incorporating such genetic information into risk scores may be a fruitful even in a very complex disease such as stroke.

Our analyses showed that the genetic risk score improved the discrimination over the mFSRS, although the absolute increase in prediction was small. The correlation between the mFSRS and genetic risk score was low suggesting the genetic variants, which are constant over the lifetime, can add to the information provided by a single assessment of variable risk factors such as blood pressure and glucose levels. Supporting evidence of this improvement

is reflected in the improved reclassification statistics. Although reclassification is seen as clinically relevant, this finding should be interpreted with caution as the value of the IDI has been questioned[31], and both the IDI and NRI may be inflated[32].

One of the most notable findings is that the GRS worked similarly in 3 of the 4 different cohorts, which, although all population-based, had different age distributions. The absence of an improvement in prediction in the FHS beyond the mFSRS may well reflect overfitting of the mFSRS in this cohort, as the original FSRS was developed on the FHS source population. Another important finding is that AUCs and AUC improvements are higher and p-values are lower for ischemic stroke compared to all stroke, despite a smaller number of cases. This difference may reflect the fact IS is an etiologically more homogenous phenotype, and the risk factors we selected are better associated to ischemic stroke than to all stroke. This is consistent with GWAS discoveries which thus far have been limited to defined subtypes[5, 6]. It will be valuable to have both better clinical prediction models and improved genetic prediction models targeting ischemic and hemorrhagic stroke as two distinct clinical entities, resulting in an improvement to the FSRS, which has a low AUC. The combined clinical-genetic-and epidemiological risk model may become a valuable tool for clinicians, and such differential risk models for ischemic versus hemorrhagic stroke may even help guide treatment decisions.

As array-based genotyping becomes more and more affordable and widely available, the possibility of multi-marker genetic risk profiling as part of daily medical practice becomes more realistic. As we show in this article, clinical risk profiling alone is still superior in predictive power to genetic risk profiling alone in the setting of a population-based cohort study of middle-aged and elderly subjects. In current clinical practice, some risk factors like diabetes, hypertension, dyslipidemia and atrial fibrillation often come to light only after a stroke has occurred. Paroxysmal atrial fibrillation may be missed during regular ECG registration or even during a 24-hour Holter monitoring. To act pre-emptively on these risk factors, a subject would have to be screened on a regular basis and at least every few years depending on their age and perhaps other factors, we suggest that their genetic profile might help select persons at higher stroke risk for more frequent or thorough screening. A genetic risk score could be estimated at a single time point early in life, and persons with a higher genetic risk be targeted for more rigorous lifestyle counseling before risk factors emerge, more stringent clinical follow-up for control of risk factors, and even preventive medication such as platelet aggregation inhibitors, statins or antihypertensive medications in persons with borderline levels of these risk factors.. We anticipate that with rapid developments in unraveling the genetic origin of various stroke risk factors, the genetic prediction of stroke risk will improve in the near future improving the efficacy of early genetic risk profiling and targeted preventive interventions. This study in a large prospectively followed population based epidemiological cohort yields a proof of principle that genetic variants associated with risk factors for stroke combined into a risk score improves discrimination of at-risk patients. These results are based on stroke-free individuals living in the community who may be examined in first line health services, and are replicated in individuals who have had a stroke and are identified in the hospital setting. However the small improvement we found is unlikely to be of clinical significance;. and only brings a small improvement over scores based on clinical information. In the future, however, as our understanding of the genetic architecture of stroke -its sub-types and risk factors - improves, GRS could become powerful additions to clinically measured risk factors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Mukherjee D, Patil CG. Epidemiology and the global burden of stroke. World Neurosurg. 2011; 76:S85–S90. [PubMed: 22182277]

2. Strong K, Mathers C, Bonita R. Preventing stroke: Saving lives around the world. Lancet neurology. 2007; 6:182–187. [PubMed: 17239805]

3. Bak S, Gaist D, Sindrup SH, Skytthe A, Christensen K. Genetic liability in stroke: A long-term follow-up study of danish twins. Stroke a journal of cerebral circulation. 2002; 33:769–774. [PubMed: 11872902]

4. Bevan S, Traylor M, Adib-Samii P, Malik R, Paul NL, Jackson C, et al. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. Stroke a journal of cerebral circulation. 2012; 43:3161–3167. [PubMed: 23042660]

5. Bellenguez C, Bevan S, Gschwendtner A, Spencer CC, et al. International Stroke Genetics C, Wellcome Trust Case Control C. Genome-wide association study identifies a variant in hdac9 associated with large vessel ischemic stroke. Nature genetics. 2012; 44:328–333. [PubMed: 22306652]

6. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell JC, Cheng YC, et al. Genetic risk factors for ischaemic stroke and its subtypes (the metastroke collaboration): A meta-analysis of genome-wide association studies. Lancet neurology. 2012; 11:951–962. [PubMed: 23041239]

7. Ikram MA, Seshadri S, Bis JC, Fornage M, DeStefano AL, Aulchenko YS, et al. Genomewide association studies of stroke. The New England journal of medicine. 2009; 360:1718–1728. [PubMed: 19369658]

8. Wolf PA, D'Agostino RB, Belanger AJ, Kannel WB. Probability of stroke: A risk profile from the framingham study. Stroke a journal of cerebral circulation. 1991; 22:312–318. [PubMed: 2003301]

9. D'Agostino RB, Wolf PA, Belanger AJ, Kannel WB. Stroke risk profile: Adjustment for antihypertensive medication. The framingham study. Stroke a journal of cerebral circulation. 1994; 25:40–43. [PubMed: 8266381]

10. Wassertheil-Smoller, S.; McGinn, A.; Allison, M.; Ca, T.; Curb, D.; Eaton, C., et al. [Accessed April 23, 2013] Improvement in stroke risk prediction: Role of c-reactive protein and lipoprotein-associated phospholipase a(2) in the women's health initiative. International journal of stroke : official journal of the International Stroke Society. 2012. [Published online ahead of print October 23 2012] http://www.ncbi.nlm.nih.gov/pubmed/23088183.

11. Kathiresan S, Melander O, Anevski D, Guiducci C, Burtt NP, Roos C, et al. Polymorphisms associated with cholesterol and risk of cardiovascular events. The New England journal of medicine. 2008; 358:1240–1249. [PubMed: 18354102]

12. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, et al. Cohorts for heart and aging research in genomic epidemiology (charge) consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circulation. Cardiovascular genetics. 2009; 2:73–80. [PubMed: 20031568]

13. ARIC-investigators. The atherosclerosis risk in communities (aric) study: Design and objectives. The aric investigators. Am J Epidemiol. 1989; 129:687–702. [PubMed: 2646917]

14. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al. The cardiovascular health study: Design and rationale. Ann Epidemiol. 1991; 1:263–276. [PubMed: 1669507]

15. Dawber TR, Kannel WB. The framingham study. An epidemiological approach to coronary heart disease. Circulation. 1966; 34:553–555. [PubMed: 5921755]

16. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The framingham offspring study. Design and preliminary data. Prev Med. 1975; 4:518–525. [PubMed: 1208363]

17. Hofman A, Breteler MM, van Duijn CM, Janssen HL, Krestin GP, Kuipers EJ, et al. The rotterdam study: 2010 objectives and design update. European journal of epidemiology. 2009; 24:553–572. [PubMed: 19728115]

18. Marroni F, Pfeufer A, Aulchenko YS, Franklin CS, Isaacs A, Pichler I, et al. A genome-wide association scan of rr and qt interval duration in 3 european genetically isolated populations: The eurospan project. Circulation. Cardiovascular genetics. 2009; 2:322–328. [PubMed: 20031603]

19. Dehghan A, Yang Q, Peters A, Basu S, Bis JC, Rudnicka AR, et al. Association of novel genetic loci with circulating fibrinogen levels: A genome-wide association study in 6 population-based cohorts. Circulation. Cardiovascular genetics. 2009; 2:125–133. [PubMed: 20031576]

20. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology. 1983; 148:839–843. [PubMed: 6878708]

21. Laliberté E. Metacor: Meta-analysis of correlation coefficients. R package version 1.0–2.2011. http://CRAN.R-project.org/package=metacor.

22. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med. 2011; 30:11–21. [PubMed: 21204120]

23. Pencina MJ, D'Agostino RB Sr, Demler OV. Novel metrics for evaluating improvement in discrimination: Net reclassification and integrated discrimination improvement for normal variables and nested models. Stat Med. 2012; 31:101–113. [PubMed: 22147389]

24. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature. 2011; 478:103–109. [PubMed: 21909115]

25. Wain LV, Verwoert GC, O'Reilly PF, Shi G, Johnson T, Johnson AD, et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. Nature genetics. 2011; 43:1005–1011. [PubMed: 21909110]

26. Paynter NP, Chasman DI, Pare G, Buring JE, Cook NR, Miletich JP, et al. Association between a literature-based genetic risk score and cardiovascular events in women. JAMA: the journal of the American Medical Association. 2010; 303:631–637. [PubMed: 20159871]

27. Brautbar A, Pompeii LA, Dehghan A, Ngwa JS, Nambi V, Virani SS, et al. A genetic risk score based on direct associations with coronary heart disease improves coronary heart disease risk prediction in the atherosclerosis risk in communities (aric), but not in the rotterdam and framingham offspring, studies. Atherosclerosis. 2012; 223:421–426. [PubMed: 22789513]

28. Hernesniemi JA, Seppala I, Lyytikainen LP, Mononen N, Oksala N, Hutri-Kahonen N, et al. Genetic profiling using genome-wide significant coronary artery disease risk variants does not improve the prediction of subclinical atherosclerosis: The cardiovascular risk in young finns study, the bogalusa heart study and the health 2000 survey--a meta-analysis of three independent studies. PloS one. 2012; 7:e28931. [PubMed: 22295058]

29. Yiannakouris N, Katsoulis M, Dilis V, Parnell LD, Trichopoulos D, Ordovas JM, et al. Genetic predisposition to coronary heart disease and stroke using an additive genetic risk score: A population-based study in greece. Atherosclerosis. 2012; 222:175–179. [PubMed: 22429504]

30. Havulinna AS, Kettunen J, Ukkola O, Osmond C, Eriksson JG, Kesaniemi YA, et al. A blood pressure genetic risk score is a significant predictor of incident cardiovascular events in 32 669 individuals. Hypertension. 2013; 61:987–994. [PubMed: 23509078]

31. Kerr KF, McClelland RL, Brown ER, Lumley T. Evaluating the incremental value of new biomarkers with integrated discrimination improvement. American journal of epidemiology. 2011; 174:364–374. [PubMed: 21673124]

32. Hilden, J.; Gerds, TA. [Accessed April 10, 2013] A note on the evaluation of novel biomarkers: Do not rely on integrated discrimination improvement and net reclassification index. Stat Med. 2013. [Published online ahead of print April 2 2013] http://www.ncbi.nlm.nih.gov/pubmed/23553436.
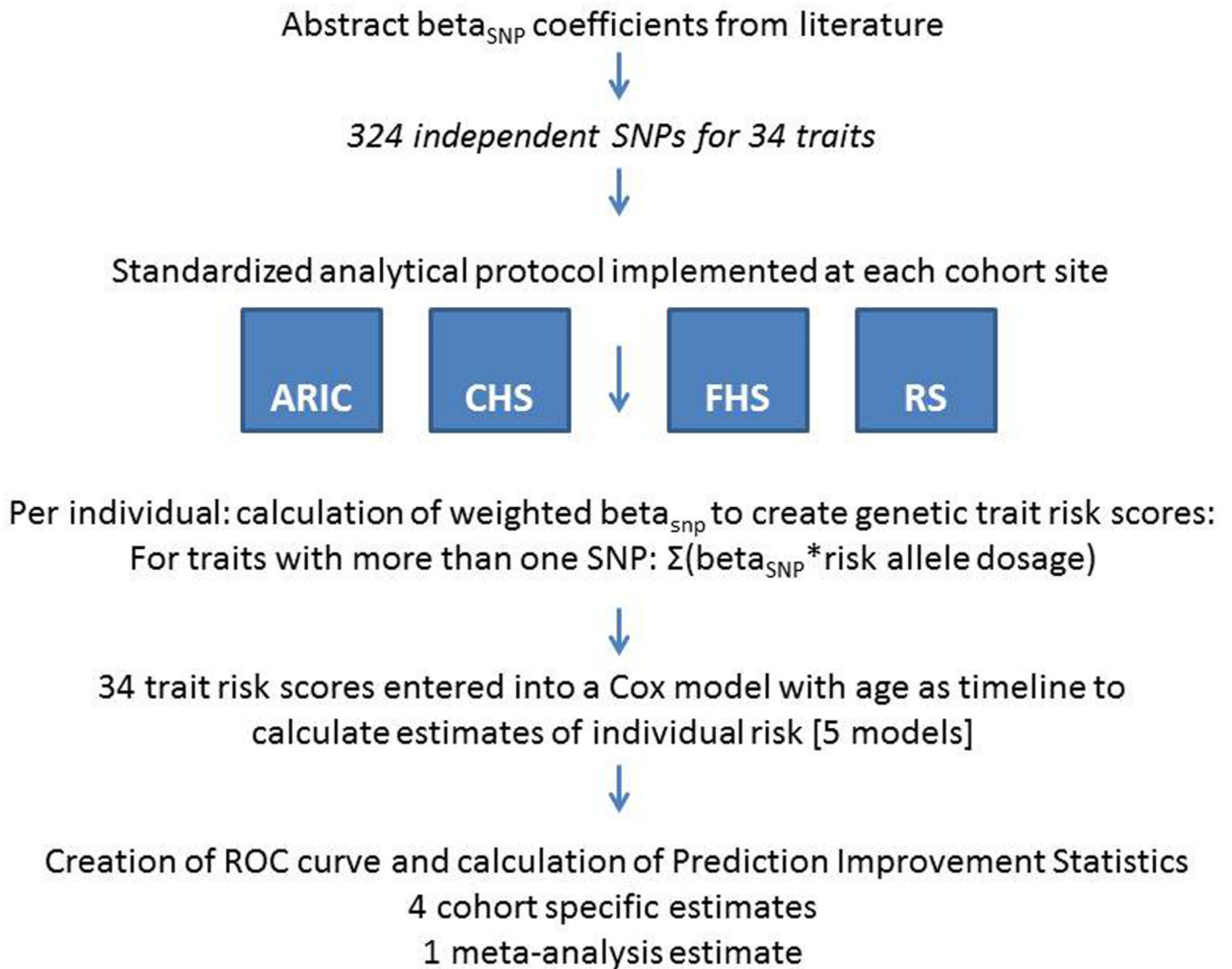
Figure 1. Flow diagram: Construction of the Genetic Risk Score for Stroke

Panel A: All stroke
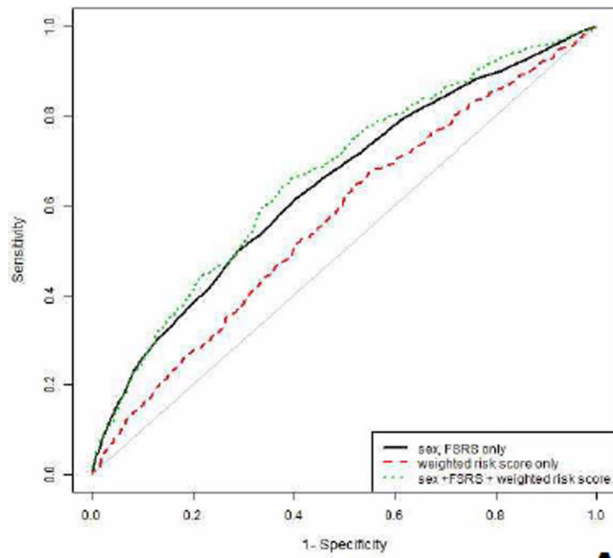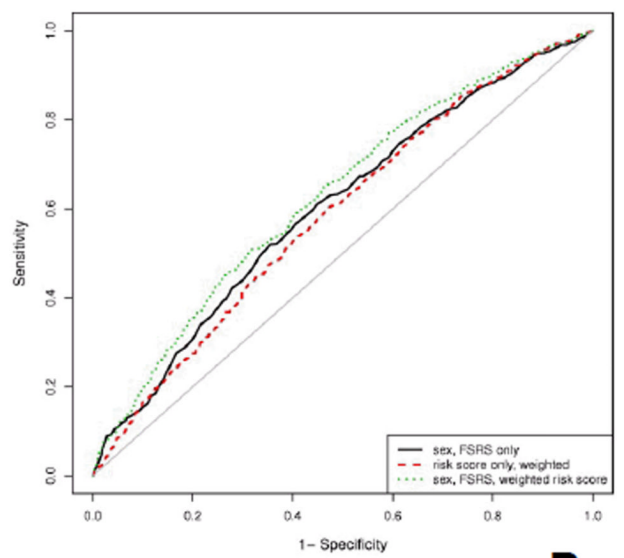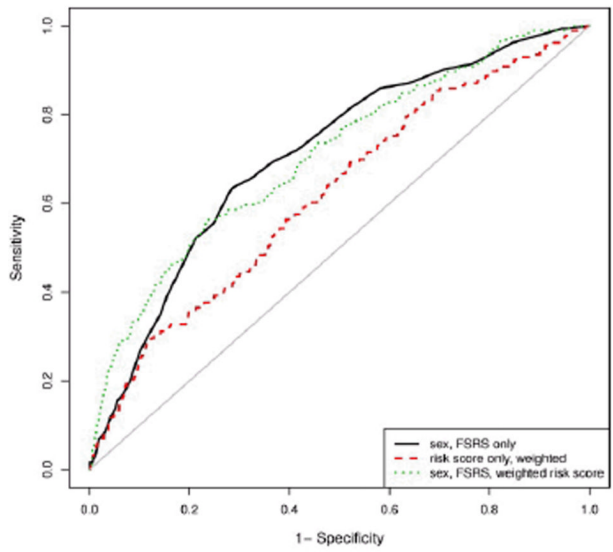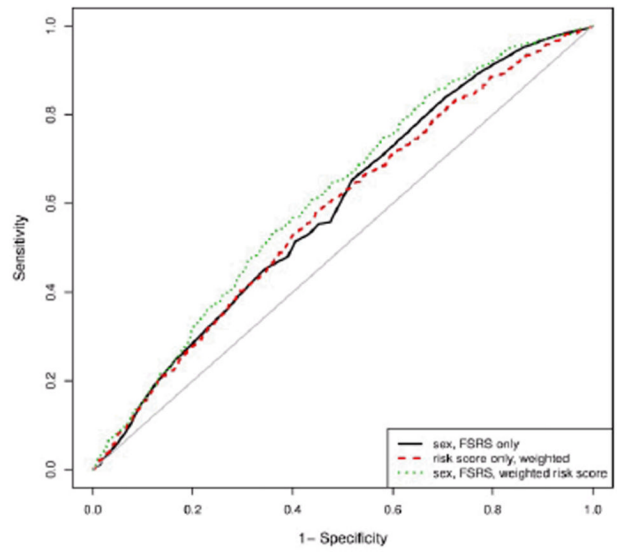
Panel   B: Ischemic stroke
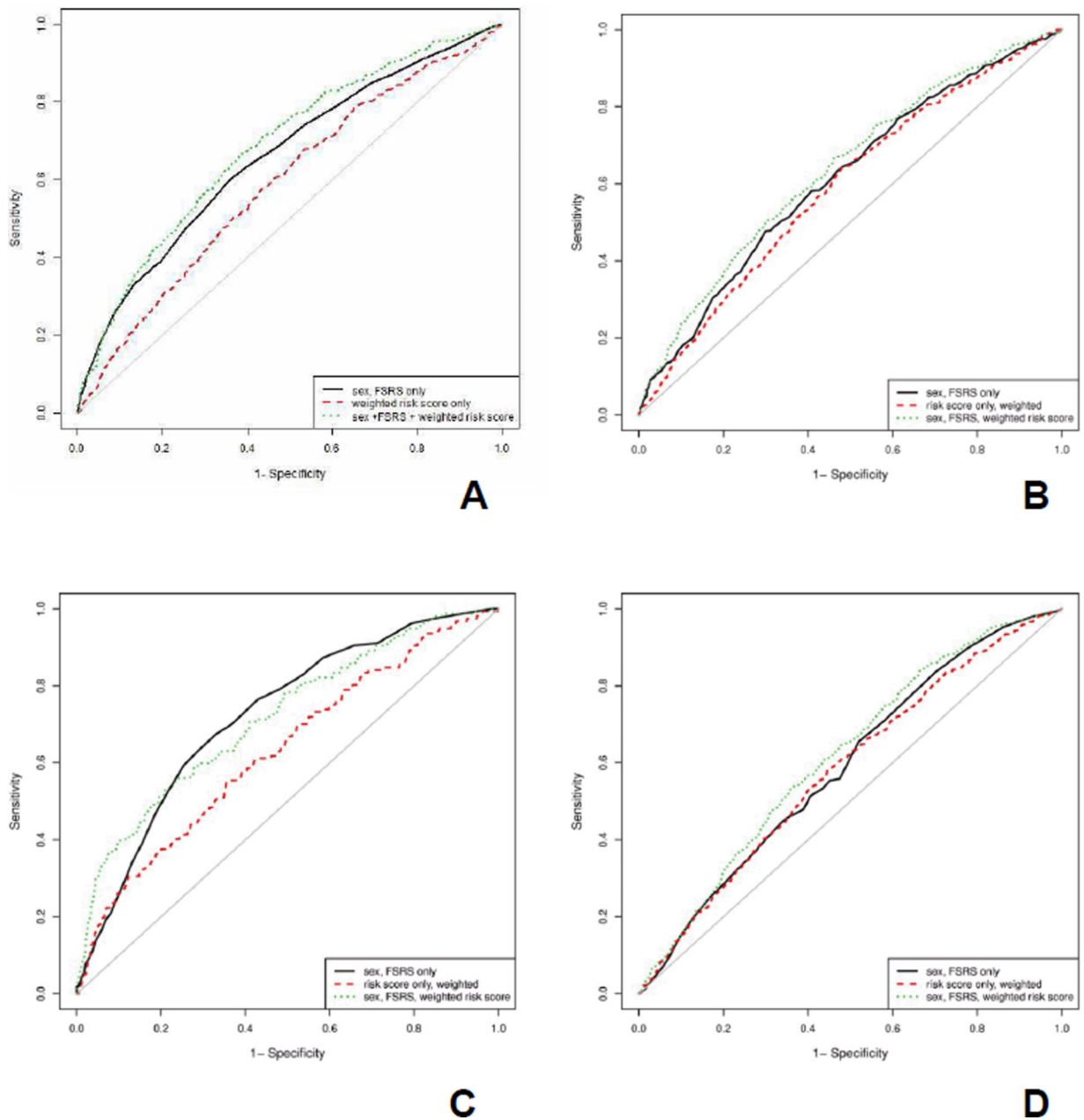
Figure 2. ROC curves for the discovery cohorts, ischemic stroke: CHARGE Genetic Risk Score for Stroke

Each lettered panel gives the sensitivity*(1-specificity) curves for the clinical prediction (sex +FSRS), prediction based on the GRS, and on the two combined. Panels: A=ARIC, B=CHS, C=FHS, D=RS.

Table 1

Participants included in the sample to develop the CHARGE Genetic Risk Score for Stroke and the Replication set

Descriptive		ARIC (n=9349)	CHS (n=3268)	FHS (n=4340)	RS (n=5763)	WTCCC (n=1581)
		Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)
Cases (N (%))	All stroke	498 (5.3%)	560 (17.1%)	206(4.8%)	783 (13.5%)	-
	Ischemic stroke	437 (4.7%)	453 (13.8%)	166 (4.1%)	467 (8.2%)	985 (62%)
Age (baseline)	Cases (All stroke)	57.17 (5.27)	73.44 (5.45)	75.14(9.93)	72.12 (8.97)	71.2 (8.7)
	Non-cases	54.13 (5.67)	72.1 (5.34)	66.16(11.69)	68.65 (8.96)	66.8 (7.9)
Age (end)	Cases (All stroke)	68.55 (7.19)	81.87 (6.19)	80.59(9.64)	83.95 (7.21)	-
	Non-cases	72.88 (6.25)	85.69 (4.97)	73.27(11.01)	81.44 (7.46)	-
Sex (females, N (%))	Cases (All stroke)	218 (43.8%)	351 (62.7%)	109(55.6%)	469 (58.6%)	383 (38.9%)
	Non-cases	4721 (53.3%)	1639 (60.5%)	2143(55.3%)	2965 (59.5%)	291 (48.7%)

Table 2

Traits included in the CHARGE Genetic Risk Score for Stroke model

Risk score group	SNPs implicated in:
Arrythmia	Atrial fibrillation, PR-interval, RR-interval
Atherosclerosis	Intima-media thickness, subclinical coronary artery disease, clinical coronary artery disease
Blood pressure	Hypertension, pulse pressure, mean arterial pressure
Diabetes	Type 2 diabetes, fasting glucose levels, insulin levels
Hematology	White blood cell count, hemoglobin, hematocrit, platelet count, mean platelet volume
Inflammation	Fibrinogen levels, C-reactive protein
Lipids	Total cholesterol, high-density lipoprotein, low-density lipoprotein, triglycerides
Nephrology	Estimated glomerular filtration rate, albumin/creatinine ratio, creatinine, end-stage renal disease
Obesity	Body mass index, waist-to-hip ratio, waist circumference
Stroke	Stroke genome-wide association studies

Table 3

Area under the Curve for all stroke: CHARGE Genetic Risk Score for Stroke

Model		AUC	95%CI lower	95%CI upper	p-value compared to null	p-value (compared to sex only model)
Sex only	ARIC	0,548	0,525	0,570	$2,908*10^{-5}$	
	CHS	0,535	0,509	0,561	$8,328*10^{-3}$	
	FHS	0,499	0,463	0,534	$9,387*10^{-1}$	
	RS	0,505	0,486	0,523	$6,060*10^{-1}$	
	Meta	**0,523**	**0,511**	**0,535**	**$1,722*10^{-4}$**	
Riskscore only	ARIC	0,566	0,540	0,592	$9,181*10^{-7}$	$3,049*10^{-1}$
	CHS	0,595	0,570	0,620	$9,477*10^{-14}$	$6,892*10^{-4}$
	FHS	0,617	0,577	0,658	$1,301*10^{-8}$	$1,963*10^{-5}$
	RS	0,563	0,542	0,584	$4,103*10^{-9}$	$6,503*10^{-5}$
	Meta	**0,578**	**0,565**	**0,591**	**$6,273*10^{-32}$**	**$8,993*10^{-10}$**
Sex + Riskscore	ARIC	0,584	0,558	0,609	$1,102*10^{-10}$	$3,852*10^{-2}$
	CHS	0,587	0,562	0,613	$9,053*10^{-12}$	$8,019*10^{-5}$
	FHS	0,612	0,571	0,653	$8,596*10^{-8}$	$2,747*10^{-6}$
	RS	0,545	0,525	0,566	$1,302*10^{-5}$	$7,736*10^{-9}$
	Meta	**0,572**	**0,560**	**0,585**	**$6,273*10^{-32}$**	**$9,133*10^{-18}$**
Sex + FSRS	ARIC	0,645	0,618	0,671	$2,26*10^{-26}$	
	CHS	0,602	0,576	0,628	$1,33*10^{-14}$	
	FHS	0,709	0,672	0,746	$2,66*10^{-22}$	
	RS	0,587	0,565	0,608	$1,37*10^{-14}$	
	Meta	**0,621**	**0,608**	**0,634**	**$6,62*10^{-73}$**	
Sex + FSRS + Riskscore	ARIC	0,664	0,639	0,690	$2,63*10^{-33}$	$1,56*10^{-3}$
	CHS	0,628	0,603	0,653	$6,70*10^{-22}$	$5,46*10^{-4}$
	FHS	0,707	0,667	0,747	$6,97*10^{-22}$	0.91
	RS	0,601	0,579	0,622	$5,57*10^{-19}$	$7,70*10^{-3}$
	Meta	**0,637**	**0,624**	**0,650**	**$6,29*10^{-93}$**	**$2,31*10^{-6}$**

Table 4

Area under the Curve for ischemic stroke: CHARGE Genetic Risk Score for Stroke

Model		AUC	95%CI lower	95%CI upper	p-value (compared to null)	p-value (compared to sex only model)
Sex only	ARIC	0,560	0,536	0,584	$4,950*10^{-7}$	
	CHS	0,527	0,499	0,555	$3,018*10^{-2}$	
	FHS	0,497	0,458	0,535	$5,713*10^{-1}$	
	RS	0,521	0,498	0,545	$3,561*10^{-2}$	
	Meta	**0,532**	**0,518**	**0,545**	$\mathbf{7,464*10^{-6}}$	
Riskscore only	ARIC	0,585	0,558	0,612	$8,561*10^{-10}$	0.17
	CHS	0,592	0,565	0,620	$2,414*10^{-11}$	$6,511*10^{-4}$
	FHS	0,627	0,583	0,671	$1,135*10^{-8}$	$1,302*10^{-5}$
	RS	0,586	0,560	0,613	$8,987*10^{-11}$	$2,681*10^{-4}$
	Meta	**0,592**	**0,578**	**0,607**	$\mathbf{5,834*10^{-38}}$	$\mathbf{2,493*10^{-9}}$
Sex + Riskscore	ARIC	0,607	0,581	0,634	$2,291*10^{-15}$	$1,596*10^{-5}$
	CHS	0,597	0,570	0,624	$1,902*10^{-12}$	$8,510*10^{-7}$
	FHS	0,622	0,579	0,666	$4,430*10^{-8}$	$1,969*10^{-6}$
	RS	0,578	0,552	0,604	$4,103*10^{-9}$	$9,108*10^{-9}$
	Meta	**0,597**	**0,582**	**0,611**	$\mathbf{8,171*10^{-37}}$	$\mathbf{1,551*10^{-19}}$
Sex + FSRS	ARIC	0,658	0,629	0,686	$2,25*10^{-27}$	
	CHS	0,613	0,585	0,640	$6,56*10^{-15}$	
	FHS	0,721	0,682	0,759	$3,12*10^{-21}$	
	RS	0,590	0,564	0,616	$1,90*10^{-10}$	
	Meta	**0,633**	**0,618**	**0,648**	$\mathbf{2,32*10^{-70}}$	
Sex + FSRS + Riskscore	ARIC	0,684	0,657	0,711	$2,00*10^{-36}$	$2,35*10^{-4}$
	CHS	0,637	0,610	0,665	$2,87*10^{-21}$	0.002
	FHS	0,716	0,673	0,759	$1,90*10^{-20}$	0.804
	RS	0,617	0,591	0,642	$2,06*10^{-16}$	$7,76*10^{-4}$
	Meta	**0,654**	**0,639**	**0,669**	$\mathbf{3,45*10^{-96}}$	$\mathbf{3,66*10^{-7}}$

All predictions result from Cox regressions with age as the time scale