Europe PMC Funders Group
Author Manuscript
Stroke. Author manuscript; available in PMC 2015 June 01.

Published in final edited form as:
Stroke. 2014 December ; 45(12): 3589–3596. doi:10.1161/STROKEAHA.114.007362.

Etiologic Ischemic Stroke Phenotypes in the NINDS Stroke Genetics Network

A full list of authors and affiliations appears at the end of the article.

Abstract

Background and Purpose—NINDS Stroke Genetics Network (SiGN) is an international consortium of ischemic stroke studies that aims to generate high quality phenotype data to identify the genetic basis of etiologic stroke subtypes. This analysis characterizes the etiopathogenetic basis of ischemic stroke and reliability of stroke classification in the consortium.

Methods—Fifty-two trained and certified adjudicators determined both phenotypic (abnormal test findings categorized in major etiologic groups without weighting towards the most likely cause) and causative ischemic stroke subtypes in 16,954 subjects with imaging-confirmed ischemic stroke from 12 US studies and 11 studies from 8 European countries using the web-based Causative Classification of Stroke System. Classification reliability was assessed with blinded re-adjudication of 1509 randomly selected cases.

Results—The distribution of etiologic categories varied by study, age, sex, and race ($p<0.001$ for each). Overall, only 40% to 54% of cases with a given major ischemic stroke etiology (phenotypic subtype) were classified into the same final causative category with high confidence. There was good agreement for both causative (kappa 0.72, 95%CI:0.69-0.75) and phenotypic classifications (kappa 0.73, 95%CI:0.70-0.75).

Conclusions—This study demonstrates that etiologic subtypes can be determined with good reliability in studies that include investigators with different expertise and background, institutions with different stroke evaluation protocols and geographic location, and patient populations with different epidemiological characteristics. The discordance between phenotypic and causative

Corresponding Author: Hakan Ay, M.D., A. A. Martinos Center for Biomedical Imaging and Stroke Service, Departments of Neurology and Radiology, Massachusetts General Hospital, Harvard Medical School, 149 13th Street, Room 2301, Charlestown MA 02129, USA, phone: 617-7244507, fax: 617-7267422, hay@mgh.harvard.edu.
*Co-Senior Authors and Co-Chairs of the SiGN Phenotype Committee

Disclosures:
Drs. Ay, Arsava, Andsberg, Benner, Chapman, Cole, Delavaran, Dichgans, Giralt-Steinhauer, Grewal, Gwinn, Jern, Jimenez-Conde, Jood, Katsnelson, Kissela, Kleindorfer, Labovitz, Lanfranconi, Lee, Lehm, Lemmens, Levi, Li, Lindgren, McArdle, Melander, Norrving, Peddareddygari, Pedersen, Pera, Rannikmäe, Rhodes, Rich, Roquer, Rothwell, Rundek, Schmidt, Schürks, Seiler, Sharma, Slowik, Sudlow, Thijs, Woodfield: None
Dr. Brown, Jr received research grant from NIH.
Dr. Engström has an employment position in Astra Zeneca R&D.
Dr. Kittner received research grant from NIH.
Dr. Markus received research grant from NIH.
Dr. Rexrode received research grant from NIH.
Dr. Rosand received research grant from NIH and has a consultant or advisory relationship with Boehringer Ingelheim.
Dr. Sacco received research grant from NIH.
Dr. Worrall received research grant from NIH and has an associate editor affiliation with American Academy of Neurology.
Dr. Meschia received research grant from NIH.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

stroke subtypes highlights the fact that the presence of an abnormality in a stroke patient does not necessarily mean that it is the cause of stroke.

Keywords

etiology; classification; phenotype; stroke subtype

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Introduction

Successful identification of genes that modify ischemic stroke risk relies on accurate delineation of etiologic stroke phenotypes.[1-5] Determination of etiologic stroke subtypes requires integration of several clinical, diagnostic, and imaging features and, therefore, is inherently subject to variability. Reproducible data on frequency of etiologic stroke subtypes based on large multicenter datasets using well-defined and evidence-based criteria do not exist. Published studies on etiologic stroke subtypes are largely constrained by poor to moderate reliability of the classification system,[6-8] suboptimal or uncertain diagnostic work-up,[9,10] small sample size,[7,8,11] single centre design,[8] and use of stringent selection criteria.[10,11]

This analysis sought to better understand the etiological basis of ischemic stroke. We prospectively identified etiologic stroke subtypes using the rule- and evidence-based Causative Classification of Stroke (CCS) system within the context of the NINDS Stroke Genetics Network (SiGN).[12-15] CCS automatically provides both phenotypic and causative stroke subtypes in each case. The former is a summary of positive test findings whereas the latter requires integration of clinical, laboratory and imaging stroke features and diagnostic test results to identify a single most likely causative subtype for each case. Hence, they provide different information. Here, we report distribution characteristics of various CCS-defined ischemic stroke subtypes and interrater reliability of etiologic subtype assignments in the SiGN dataset.

Methods

Contributing studies and patient population

SiGN is a large international consortium of ischemic stroke studies aiming to generate high quality phenotype data to assist in the identification of the genetic basis of ischemic stroke subtypes. This analysis included ischemic stroke cases from the initial 12 US and 11 European ischemic stroke studies in SiGN from 9 countries. Imaging confirmation of the absence of hemorrhagic stroke was required in each subject. Details about the individual contributing studies have previously been described in a separate publication.[15] Seventeen studies recruited consenting cases without using any selection criteria. In contrast, 6 studies were conducted in selected populations based on age, sex, and family history.[15] Recruitment to contributing studies occurred during a 23-year period between 1989 and 2012.

Stroke Subtyping

Etiologic stroke classification in SiGN started in July 2010. The current study included 16,954 cases for whom etiologic subtype information was available in the SiGN database as

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

of March 2014. SiGN used the web-based CCS system for stroke subtyping (available at https://ccs.mgh.harvard.edu).[13] The details of CCS were published elsewhere.[13] For the purpose of SiGN, we customized CCS by generating a confidential, password protected data collection platform. We also made a modification in the online CCS form by separating the single data entry field for small artery occlusion (SAO) in the original CCS into two separate data entry fields: one to indicate whether there is a typical lacunar infarct on neuroimaging and the second one to rule out whether there is an accompanying parent artery disease at the origin of the penetrating artery supplying the site of the lacunar infarct. Thus, it became possible to collect phenotypic data on lacunar infarcts for which vascular imaging for parent artery disease was not available. No modification was made in the decision-making code of the CCS; both customized and original CCS algorithms provided the same subtype for each given test condition.

We determined *phenotypic* subtypes in each subject.[13,14] Phenotypic subtypes referred to abnormal test findings categorized in major etiologic groups without weighting towards the most likely cause in the presence of multiple causes.[14] There were 4 main phenotypic categories: large artery atherosclerosis (LAA), cardiac embolism (CE), lacunar infarction (LI), and other-uncommon causes. There were 4 possible states for LAA and CE (major, minor, absent, incomplete evaluation), 3 for LI (major, absent, incomplete evaluation), and 2 for other-uncommon causes (major and absent), giving rise to a total of 96 phenotypic categories. We collapsed these 96 categories into the following 7 subtypes: LAA-major, CE-major, LI-major, other-major, no major etiology, multiple competing major etiologies, and incomplete investigation. We further collapsed the last 3 categories into “undetermined” category and generated a 5-subtype phenotypic categorization.

We also recorded *causative* subtypes in each case. In contrast to phenotypic subtypes, causative subtyping requires integration of multiple aspects of ischemic stroke evaluation in a probabilistic and objective manner.[12,13] The causative subtype differs from the phenotypic subtype in certain occasions. For instance, in a patient with internal carotid artery stenosis, ipsilateral internal borderzone infarcts, and atrial fibrillation, the causative subtype is “LAA”, whereas the phenotypic subtype is “multiple competing etiologies” due to coexistence of LAA and CE. Major causative categories included LAA, CE, SAO, other-uncommon causes, and undetermined causes. The undetermined group was further divided into 4 subcategories as cryptogenic-embolism, cryptogenic-other, incomplete evaluation, and multiple competing causes (unclassified). We grouped cardiac pathologies with uncertain risk of stroke (minor sources) into the “undetermined cryptogenic-other” category. This allowed us to generate a more refined cardio-embolic category (CE-major). Each causative category in CCS (except for undetermined category) was subdivided based on the weight of available data as “evident”, “probable”, or “possible” in order to identify the level of confidence in assigning an etiology.[12] Overall, CCS generated 17 causative subtypes.

CCS did not require a minimum level of investigations. In cases with missing tests, the system still assigned a subtype based on results of available tests but with a lower level of confidence. For instance, in a patient with typical lacunar infarct in the internal capsule and missing intracranial vascular imaging to rule out a parent artery disease, the level of confidence in attributing lacunar infarct to SAO was reduced from evident to possible. A

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

subtype (both causative and phenotypic) was considered to be "*incomplete evaluation*" only when brain imaging, vascular imaging, or cardiac evaluation were not performed in the absence of an identified etiology.

Data adjudication and quality control

A total of 52 adjudicators (13 stroke neurologists, 17 stroke fellows, 13 neurology residents, 9 non-neurologists) performed stroke subtyping. A centralized Phenotype Committee of 4 expert stroke neurologists met weekly to monitor data quality and site performance. The same committee blindly readjudicated a randomly selected 10% of cases recruited from the US studies for quality control. Similarly, 10% of cases from European studies were readjudicated by blinded European investigators (n=20). Each adjudicator and readjudicator had to complete an interactive online training module. The Phenotype Committee members provided training to adjudicators/readjudicators regarding data entry, data submission, and archiving at scheduled study meetings and via webinars. Every investigator was required to pass an on-line certification examination available at the CCS website.

Data source

Study-specific case report forms and unabstracted medical records served as data source for subtyping. Readjudicators used the same data source available to adjudicators to determine the CCS subtypes. Data sources varied in length and detail among the study sites. Subtype assignments were done based on data available at the time of discharge in the majority. Prolonged ambulatory cardiac monitoring was obtained after discharge in 14% of the subjects. In such cases, we used post-discharge cardiac monitoring findings for stroke subtyping. All data entered into CCS as well as the system output were saved in a confidential SiGN database. In addition to subtype-related data, each study site provided baseline variables such as age, sex, race, and vascular risk factors, using a structured data collection form.

Statistics

Our primary objective was to determine the distribution of CCS subtypes within the SiGN cohort. We also determined etiologic subtype distribution in a subset with complete diagnostic investigation. We defined *complete investigation* as the presence of brain imaging, intracranial and extracranial vascular imaging, and cardiac evaluation with echocardiography if ECG and clinical assessment did not reveal a source. We assessed the heterogeneity among centers in utilization of diagnostic tests using the chi-square test. We used Chi-square test and Student's t-test to evaluate differences between cohorts with and without complete investigation for categorical and continuous variables, respectively. We assessed the correlation between causative and phenotypic subtypes by calculating how often CCS classified a given major abnormal evaluation finding (phenotypic subtype) as the causative stroke mechanism (causative subtype). We performed multinomial logistic regression to evaluate associations between causative CCS subtypes and age, sex, and race. In regression models, undetermined category served as the reference. We assessed the concordance between paired ratings by adjudicators and re-adjudicators by calculating crude agreement rates and unweighted kappa values for both 5-subtype causative and phenotypic

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

classification.[16] We expressed associations as odds ratios (OR) and 95% confidence intervals (CI). We considered p values <0.05 as statically significant.

Results

Study cohort

Table 1 presents characteristics of the study population. Complete diagnostic investigation was available in 46%. Cases with complete investigation were similar to those with incomplete investigation except that they were slightly younger and more likely to be men (p<0.001, Table 1). The proportion of cases with complete investigation varied across the 23 studies (p<0.001, Table 2). US studies had higher complete investigation rate as compared to European centers (53% vs. 40%, p<0.001).

Stroke subtypes

Figure 1 shows the distribution of phenotypic and causative subtypes. Compared to the overall population, subtype distribution differed in the cohorts with complete investigation (Figure 1, <0.001) and after exclusion of the 6 studies that employed selection rules (Supplemental Figure I, <0.001).

Vascular investigations revealed an atherosclerotic lesion causing ≥50% stenosis (LAA-major phenotype) in 3,392 of the 16,954 cases (20%); among these, 2,093 (62%) had extracranial stenosis, 962 (28%) had intracranial stenosis, and 337 (10%) had both extra- and intra-cranial stenoses. LAA-major was an isolated finding in 2,536 (75%); in the remaining 856 (25%), there was another major etiology such as a major cardiac emboli source. Diagnostic tests for other etiologies were missing in 972 (29%). Overall, 1,719 (51%) cases with a major LAA had either a missing test or another competing etiology. The final causative subtype was LAA-evident in only 1,815 (54%) cases (Figure 2a). The remaining individuals were either classified into the category of LAA but with a lower level of confidence (either probable or possible) or into another category.

There was a major cardiac source in 4,496 of the 16,954 (27%) cases. Atrial fibrillation accounted for the largest proportion of the major cardiac source of embolism (3,735; 83%). There was another competing major vascular or systemic abnormality in 816 (18%) cases. Diagnostic investigations were incomplete in 2,229 (50%) cases. The final causative subtype was CE-evident in 2,011 (45%) cases with a major cardiac source of embolism (Figure 2b).

There were 2,458 (15%) cases with a typical lacunar infarct on neuroimaging. Among those with lacunar infarction, intracranial vascular imaging was available in 1,567 (64%). An abnormality in the parent artery at the origin of the penetrating artery supplying the territory of the lacunar infarct was reported in 300 (19%). Cardiac investigations were performed in 1,617 (66%) and revealed a major cardiac source in 208 (13%). Overall, vascular and cardiac evaluations revealed another major etiology in 492 (20%). Investigations were incomplete in 1,344 (55%). The final causative subtype was SAO-evident in 992 (40%) cases with a typical lacunar infarct on neuroimaging (Figure 2c).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Diagnostic investigations revealed a major uncommon etiology in 1,109 (7%) cases. The most frequent uncommon etiology was acute arterial dissection (397 cases, 36%). Overall, incomplete evaluation and completing major etiology rate was 53% in this category. The final causative subtype was evident other-uncommon causes in 526 (47%) patients with a major uncommon etiology (Figure 2d).

The largest etiologic category was “undetermined etiology” (7,272 cases, 43%). This category included 3,947 (55%) with incomplete evaluation, 1,333 (18%) with minor cardiac emboli sources, 655 (9%) with multiple competing etiologies, 1,118 (15%) with cryptogenic-other, and 219 (3%) with cryptogenic embolism. Of note, there were a total of 3,257 (19%) cases in the entire study cohort with multiple competing etiologies (major or minor). Nevertheless, the final causative subtype was “unclassified” in only 655 (4%), suggesting that the CCS algorithm was able to identify a probable etiology in the vast majority of patients with overlapping etiologies.

There was a significant relationship between stroke subtypes and age, sex, and race ($p<0.001$ for each, Supplemental Figure II). The association was most obvious for age. Subjects 50 years or older were approximately 4 times more likely to have cardioembolic stroke and 6 times less likely to have stroke due to other-uncommon causes as compared to those younger than 50 years. In a further analysis where age was classified by decades, we found a continuous increase in LAA and SAO with increasing age with peak values in the age around 50 to age 70 (Figure 3). There was no such age peak in major CE; instead, the probability of major CE continuously increased by increasing age. In contrast, there was a steady decrease in cryptogenic and other-uncommon strokes by increasing age.

Reliability

There were 1509 paired ratings by 52 adjudicators and 24 readjudicators. The crude agreement for 5-subtype causative system was 80% (Supplemental Table I). The corresponding kappa value was 0.72 (95% CI, 0.69-0.75). The crude agreement rate for 5-subtype phenotypic system was 81% with a corresponding kappa value of 0.73 (95% CI, 0.70-0.75, Supplemental Table II). Crude agreements rates for causative system varied between 65% and 99% across the study sites except for one site where the agreement rate was 40% (Supplemental Figure III). After excluding that one outlier, the kappa value increased to 0.75 (95% CI, 0.72-0.77) for causative and 0.75 (95% CI, 0.72-0.78) for phenotypic classifications.

Discussion

This is a very large study of systematic ischemic stroke subtyping using an evidence- and rule-based system. Because of its size, patterns of subtype distribution across age groups are more readily discernible. It is also the largest study of the interrater reliability of ischemic stroke subtyping published thus far, based on 1,509 paired ratings by a total of 76 trained and certified adjudicators and readjudicators. There was simultaneous assessment of phenotypic and causative subtypes allowing examination of subtype distribution and reliability separately for these two types of classification. Our finding of discordance between the causative and phenotypic classifications is expected and reflects the fact that the

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

presence of a phenotypic characteristic in a given patient, such as a vascular or cardiac abnormality, does not necessarily mean that it is the cause of stroke in that patient.

The extent of diagnostic evaluation was heterogeneous for a variety of reasons. Some studies used single site recruitment where strokes were evaluated at tertiary medical centers by vascular neurologists with a highly consistent diagnostic approach, while other studies were regional or national in scope with strokes evaluated primarily at community hospitals by physicians with diverse backgrounds with a less consistent diagnostic approach.[15] This variation in extent of diagnostic evaluation motivated us to provide data separately for the subset with complete vascular and cardiac investigations. The results in this subset are the highest quality data available in the literature on the distribution of stroke subtypes. Of note, the subset with complete investigations resembled the overall study population with respect to the majority of baseline characteristics, suggesting no substantial selection bias.

In the present study, interrater reliability was slightly lower (kappa=0.72) than previously reported for CCS (kappa ≥ 0.80).[12-14] Prior studies had smaller number of raters (n=2 to 20) and smaller number of cases (n=50). As the number of cases and the number of raters increase, the variance introduced to stroke classification also increases and hence the reliability decreases. In contrast to prior studies that used abstracted case summaries, reliability analysis in this study was based on reviews of unabstracted case report forms and patient charts. Differences in raters' ability to pinpoint the medical record data that is critical for subtyping, ambiguities in the source data (for instance, inconsistencies in interpretation of test finding between physician notes), variance in raters' interpretation of the diagnostic data, and lack of data or incomplete data (for instance, unavailability of radiographic images for visual assessment) likely contributed to lower agreement. Despite all these factors, kappa for CCS still exceeds reported kappas for conventional classification systems. The Women's Health Study is the largest reliability study of TOAST including 133 cases and 2 raters and reporting a kappa of 0.49. [9] The Siblings with Ischemic Stroke Study (SWISS) assessed the reliability of TOAST using the largest number of raters (6 raters and 30 cases) and reporting a kappa of 0.54.[11] The kappa for conventional classification would be expected to be much lower when tested in the same test conditions with the present study.

Several limitations merit further discussion. We did not include a specific minimum standard for quality of source data used for phenotyping. The 23 studies included in this analysis represent a broad range of methodologies (hospital-based case-control, pedigree-based, observational cohorts, and population-based studies) and using a broad range of criteria for inclusion ranging from no restriction to targeted recruitment by age, sex, family history, etc. Source documents varied from secondary notes of test results to computerized data repositories where access to source data such as radiographic images was possible. This diversity strengthens the confidence in our findings by capturing the vagaries that may occur in the "real world" as opposed to the rigors of a structured clinical research setting. Moreover, CCS provides refined subtypes by integrating quality and completeness of source data into level of confidence for each subtype, minimizing the impact of diversity in source data on validity of classifications. Insufficient representation of certain racial and ethnic groups (for instance, Asian population) in SiGN may have caused underestimation of certain mechanisms such as intracranial atherosclerosis. Finally, because majority of studies were

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

hospital-based, the study population was vulnerable to survival, severity or consent bias in addition to the impact of specific inclusion and exclusion criteria.

A major strength of this study was systematic adjudication of stroke subtypes using a rule- and evidence-based system. CCS offers several advantages such as good to excellent reliability and web-based interface.[13] Additionally, CCS retains and standardizes individual data points such as atrial fibrillation or arterial dissection that underlie subtype classification. Furthermore, its ability to provide both phenotypic and causative subtypes would allow one to separately explore the genetic basis of the presence of a potential etiology (phenotypic subtype) and the presence of a causative etiology (causative subtype). A gene for large artery atherosclerosis could be different from a gene that makes an atherosclerotic plaque rupture and cause stroke. The ability to study such differential genetic associations would facilitate our understanding of the pathophysiological basis of ischemic stroke.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Hakan Ay, MD[a,b], Ethem Murat Arsava, MD[a], Gunnar Andsberg, MD[c], Thomas Benner, PhD[a], Robert D. Brown Jr., MD[d], Sherita N. Chapman, MD[e], John W. Cole, MD, MS[f], Hossein Delavaran, MD[c], Martin Dichgans, MD[g], Gunnar Engström, MD[h], Eva Giralt-Steinhauer, MD[i], Raji P. Grewal, MD[j], Katrina Gwinn, MD[k], Christina Jern, MD[l], Jordi Jimenez-Conde, MD[i], Katarina Jood, MD[m], Michael Katsnelson, MD[n], Brett Kissela, MD[o], Steven J. Kittner, MD[f], Dawn O. Kleindorfer, MD[o], Daniel L. Labovitz, MD[p], Silvia Lanfranconi, MD[q], Jin-Moo Lee, MD[r], Manuel Lehm, BSc[g], Robin Lemmens, MD[s], Chris Levi, MD[t], Linxin Li, PhD[u], Arne Lindgren, MD[c], Hugh S. Markus, DM[v], Patrick F. McArdle, PhD[w], Olle Melander, MD[c], Bo Norrving, MD[c,h], Leema Reddy Peddareddygari, MD[j], Annie Pedersén, MD[l], Joanna Pera, MD[x], Kristiina Rannikmäe, MD[y], Kathryn M. Rexrode, MD[z], David Rhodes, MPH[aa], Stephen S. Rich, PhD[bb], Jaume Roquer, MD, PhD[i], Jonathan Rosand, MD, MSc[b,cc], Peter M. Rothwell, MD[u], Tatjana Rundek, MD, PhD[n], Ralph L. Sacco, MD, MS[n], Reinhold Schmidt, MD[dd], Markus Schürks, MD[ee], Stephan Seiler, MD[dd], Pankaj Sharma, MD[ff], Agnieszka Slowik, MD[x], Cathie Sudlow, MD[y], Vincent Thijs, MD[s], Rebecca Woodfield, MD[y], Bradford B. Worrall, MD, MSc[e,gg,*], and James F. Meschia, MD[hh,*]

Affiliations

[a]AA Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA [b]Stroke Service, Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA [c]Department of Neurology and Rehabilitation Medicine, Skåne University Hospital, Lund, Sweden [d]Department of Neurology, Mayo Clinic Rochester, Rochester, MN, USA [e]Department of Neurology, University of Virginia, Charlottesville, VA, USA [f]Department of Neurology, University

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

of Maryland School of Medicine, Baltimore, MD, USA [g]Institute for Stroke and Dementia Research (ISD), Klinikum der Universität München Ludwig-Maximilians-University, München, Germany [h]Department of Clinical Science, Lund University, Malmö, Sweden [i]Department of Neurology, Neurovascular Research Group, IMIM-Hospital del Mar, Universitat Autonoma de Barcelona/DCEXS-Universitat Pompeu Fabra, Barcelona, Spain [j]Neuroscience Institute, Saint Francis Medical Center, Trenton, NJ, USA [k]National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA [l]Institute of Biomedicine, the Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden [m]Institute of Neuroscience and Physiology, the Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden [n]Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA [o]Department of Neurology, University of Cincinnati College of Medicine, Cincinnati, OH, USA [p]Stern Stroke Center, Albert Einstein College of Medicine, Bronx, NY, USA [q]Department of Neuroscience and Sensory Organs, Policlinico Hospital Foundation IRCCS Cà Granda, Milan, Italy [r]Department of Neurology, Washington University, St. Louis, MO, USA [s]Department of Neurology, University Hospitals Leuven, Leuven, Belgium [t]Department of Neurology, John Hunter Hospital, The University of Newcastle, Newcastle, New South Wales, Australia [u]Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, Oxford University, Oxford, UK [v]Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK [w]Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD, USA [x]Department of Neurology, Jagiellonian University, Medical College, Krakow, Poland [y]Centre for Clinical Brain Sciences, The University of Edinburgh, Edinburgh, UK [z]Division of Preventive Medicine, Brigham & Women's Hospital, Boston, MA, USA [aa]School of Public Health, University of Alabama, Birmingham, AL, USA [bb]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA [cc]Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA [dd]Department of Neurology, Medical University Graz, Austria [ee]Department of Neurology, University Hospital Essen, Essen, Germany [ff]Cerebrovascular Research Unit, Imperial College London, London, UK [gg]Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA [hh]Department of Neurology, Mayo Clinic Jacksonville, Jacksonville, FL, USA

Acknowledgements

None

Funding:

SiGN: The study was funded by a cooperative agreement grant from the National Institute of Neurological Disorders and Stroke (NINDS) U01 NS069208.

BASICMAR: The BASICMAR Genetic Study was supported by the Ministerio de Sanidad y Consumo de España, Instituto de Salud Carlos III (ISC III) with the grants: "Registro BASICMAR" Funding for Research in Health (PI051737); "GWALA project" from Fondos de Investigación Sanitaria ISC III (PI10/02064) and (PI12/01238);

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

and Fondos FEDER/EDRF Red de Investigación Cardiovascular (RD12/0042/0020). Additional support provided by the Fundació la Marató TV3 with the grant “GOD’s project. Genestroke Consortium” (76/C/2011) Recercaixa’13 (JJ086116). Assistance with data cleaning was provided by the Research in Cardiovascular and Inflammatory Diseases Program of Institute Hospital del Mar of Medical Investigations, Hospital del Mar, and the Barcelona Biomedical Research Park.

BRAINS: BRAINS (Bio-Repository of DNA in Stroke) was supported by the British Council (UKIERI), Henry Smith Charity, and the UK Stroke Research Network. Dr Sharma was supported by a Department of Health (UK) Senior Fellowship.

CIDR: Genotyping services were provided by the Johns Hopkins University Center for Inherited Disease Research (CIDR), which is fully funded through a federal contract from the National Institutes of Health to the Johns Hopkins University (contract number HHSN268200782096C).

EDINBURGH: The Edinburgh Stroke Study was supported by the Wellcome Trust (clinician scientist award to Dr Sudlow) and the Binks Trust. Sample processing occurred in the Genetics Core Laboratory of the Wellcome Trust Clinical Research Facility, Western General Hospital, Edinburgh. Much of the neuroimaging occurred in the Scottish Funding Council Brain Imaging Research Centre (www.sbirc.ed.ac.uk), Division of Clinical Neurosciences, University of Edinburgh, a core area of the Wellcome Trust Clinical Research Facility and part of the Scottish Imaging Network – A Platform for Scientific Excellence (SINAPSE) collaboration (www.sinapse.ac.uk), funded by the Scottish Funding Council and the Chief Scientist Office. Genotyping was performed at the Wellcome Trust Sanger Institute in the UK and funded by the Wellcome Trust as part of the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z and 085475/Z/08/Z and WT084724MA).

GASROS: The Massachusetts General Hospital Stroke Genetics Group was supported by the National Institutes of Health Genes Affecting Stroke Risks and Outcomes Study (GASROS) grant K23 NS042720, the American Heart Association/Bugher Foundation Centers for Stroke Prevention Research 0775010N, and NINDS K23NS042695, R01NS059727, the Deane Institute for Integrative Research in Atrial Fibrillation and Stroke, and by the Keane Stroke Genetics Fund. Genotyping services were provided by the Broad Institute Center for Genotyping and Analysis, supported by grant U54 RR020278 from the National Center for Research Resources.

GCNKSS: The Greater Cincinnati/Northern Kentucky Stroke Study was supported by the National Institutes of Health (NS 030678).

GEOS: The GEOS Study was supported by the National Institutes of Health Genes, Environment and Health Initiative (GEI) grant U01 HG004436, as part of the Gene Environment Association Studies (GENEVA) consortium under GEI, with additional support provided by the Mid-Atlantic Nutrition and Obesity Research Center (P30 DK072488) and the Office of Research and Development, Medical Research Service, and the Baltimore Geriatrics Research, Education, and Clinical Center of the Department of Veterans Affairs. Genotyping services were provided by the Johns Hopkins University CIDR, which is fully funded through a federal contract from the National Institutes of Health to the Johns Hopkins University (contract number HHSN268200782096C). Assistance with data cleaning was provided by the GENEVA Coordinating Center (U01 HG 004446; PI Bruce S Weir). Study recruitment and assembly of data sets were supported by a Cooperative Agreement with the Division of Adult and Community Health, Centers for Disease Control and by grants from the NINDS and the National Institutes of Health (NIH) Office of Research on Women’s Health (R01 NS45012, U01 NS069208-01).

GRAZ: The Austrian Stroke Prevention Study was supported by the Austrian Science Fund (FWF) grant numbers P20545-P05 and P13180 and I904-B13 (Era-Net). The Medical University of Graz supports the databases of the Graz Stroke Study and the Austrian Stroke Prevention Study.

ISGS and SWISS: The Ischemic Stroke Genetics Study (ISGS) was supported by the NINDS (R01 NS42733; PI Dr Meschia). The Sibling with Ischemic Stroke Study (SWISS) was supported by the NINDS (R01 NS39987; PI Dr Meschia). Both SWISS and ISGS received additional support, in part, from the Intramural Research Program of the National Institute on Aging (Z01 AG000954-06; PI Andrew Singleton). SWISS and ISGS used samples and clinical data from the NIH-NINDS Human Genetics Resource Center DNA and Cell Line Repository (http://ccr.coriell.org/ninds), human subjects protocol numbers 2003–081 and 2004–147. SWISS and ISGS used stroke-free participants from the Baltimore Longitudinal Study of Aging (BLSA) as controls with the permission of Dr Luigi Ferrucci. The inclusion of BLSA samples was supported, in part, by the Intramural Research Program of the National Institute on Aging (Z01 AG000015-50), human subjects protocol number 2003–078. This study used the high-performance computational capabilities of the Biowulf Linux cluster at the NIH (http://biowulf.nih.gov).

KRAKOW: Phenotypic data and genetic specimens collection were funded by the grant from the Polish Ministry of Science and Higher Education for Leading National Research Centers (KNOW) and by the grant from the Medical College, Jagiellonian University in Krakow, Poland: K/ZDS/002848.

Europe PMC Funders Author Manuscripts

LEUVEN: The Leuven Stroke genetics study was supported by personal research funds from the Department of Neurology of the University Hospitals Leuven. Vincent Thijs is supported by a Fundamental Clinical Research grant from FWO Flanders (numbers 1.8.009.08.N.00 and 1800913N). Robin Lemmens is a Senior Clinical Investigator of FWO Flanders (FWO 1841913N) and is supported through Fonds Annie Planckaert-Dewaele.

LUND: The Lund Stroke Register was supported by the Swedish Research Council (K2010-61X-20378-04-3), The Swedish Heart-Lung Foundation, Region Skåne, the Freemasons Lodge of Instruction EOS in Lund, King Gustaf V's and Queen Victoria's Foundation, Lund University, and the Swedish Stroke Association. Biobank services were provided by Region Skåne Competence Centre (RSKC Malmö), Skåne University Hospital, Malmö, Sweden, and Biobank, Labmedicin Skåne, University and Regional Laboratories Region Skåne, Sweden.

MCISS: The Middlesex County Ischemic Stroke Study (MCISS) was supported by intramural funding from the New Jersey Neuroscience Institute/JFK Medical Center, Edison, NJ, and The Neurogenetics Foundation, Cranbury, NJ.

MIAMISR and NOMAS(S): The Northern Manhattan Study (NOMAS) was supported by grants from the NINDS (R37 NS029993, R01 NS27517). The Cerebrovascular Biorepository at University of Miami/Jackson Memorial Hospital (The Miami Stroke Registry, Institutional Review Board No. 20070386) was supported by the Department of Neurology at University of Miami Miller School of Medicine and Evelyn McKnight Brain Institute. Biorepository and DNA extraction services were provided by the Hussmann Institute for Human Genomics at the Miller School of Medicine.

MUNICH: The MUNICH study was supported by the Vascular Dementia Research Foundation and the Jackstaedt Stiftung.

NHS: The Nurses' Health Study work on stroke is supported by grants from the National Institutes of Health, including HL088521 and HL34594 from the National Heart Lung and Blood Institute, as well as grants from the National Cancer Institute funding the questionnaire follow-up and blood collection: CA87969 and CA49449.

OXVASC: The Oxford Vascular Study was supported by the Stroke Association, Medical Research Council, Wellcome Trust, Dunhill Medical Trust, National Institutes of Health Research (NIHR), and NIHR Oxford Biomedical Research Centre based at Oxford University Hospitals NHS Trust and University of Oxford. Dr. Rothwell is in receipt of Senior Investigator Awards from the Wellcome Trust and the NIHR.

REGARDS: The Reasons for Geographic and Racial Differences in Stroke Study was supported by a cooperative agreement U01 NS041588 from the NINDS, National Institutes of Health, and Department of Health and Human Service. A full list of participating REGARDS investigators and institutions can be found at http://www.regardsstudy.org.

SAHLSIS: The Sahlgrenska Academy Study of Ischemic Stroke was supported by the Swedish Research Council (K2011-65X-14605-09-6), the Swedish Heart and Lung Foundation (20100256), the Swedish state/Sahlgrenska University Hospital (ALFGBG-148861), the Swedish Stroke Association, the Swedish Society of Medicine, and the Rune and Ulla Amlöv Foundation.

SPS3: The Secondary Prevention of Small Subcortical Strokes trial was funded by the US National Institute of Health and Neurological Disorders and Stroke grant No. U01NS38529-04A1 (principal investigator, O.R.B.; co-principal investigator, R.G.H.). The SPS3 Genetic Substudy (SPS3-GENES) was funded by R01 NS073346 (co-principal investigators, J.A.J, O.R.B, A.R.S.).

ST. GEORGE'S: The principal funding for this study was provided by the Wellcome Trust, as part of the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z and 085475/Z/08/Z and WT084724MA). Collection of some of the St George's stroke cohort was supported by project grant support from the Stroke Association.

WHI: The Women's Health Initiatives (WHI) program was funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, US Department of Health and Human Services through contracts N01WH22110, 24152, 32100–2, 32105–6, 32108–9, 32111–13, 32115, 32118 to 32119, 32122, 42107–26, 42129–32, and 44221. The Hormones and Biomarkers Predicting Stroke (HaBPS) was supported by a grant from the National Institutes of Neurological Disorders and Stroke (R01NS042618).

WUSTL: The collection, extraction of DNA from blood, and storage of specimens were supported by two NINDS NIH grants (P50 NS055977 and R01 NS8541901). Basic demographic and clinical characterization of stroke phenotype was prospectively collected in the Cognitive Rehabilitation and Recovery Group (CRRG) registry. The Recovery Genomics after Ischemic Stroke (ReGenesIS) study was supported by a grant from the Barnes-Jewish Hospital Foundation.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

References

1. Gschwendtner A, Bevan S, Cole JW, Plourde A, Matarin M, Ross-Adams H, et al. Sequence variants on chromosome 9p21.3 confer risk for atherosclerotic stroke. Ann Neurol. 2009; 65:531–9. [PubMed: 19475673]
2. International Stroke Genetics Consortium (ISGC); Wellcome Trust Case Control Consortium 2 (WTCCC2). Bellenguez C, Bevan S, Gschwendtner A, Spencer CC, et al. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. Nat Genet. 2012; 44:328–33. [PubMed: 22306652]
3. Holliday EG, Maguire JM, Evans TJ, Koblar SA, Jannes J, Sturm JW, et al. Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. Nature Genet. 2012; 44:1147–51. [PubMed: 22941190]
4. Lubitz SA, Yi BA, Ellinor PT. Genetics of atrial fibrillation. Heart Fail Clin. 2010; 6:239–47. [PubMed: 20347792]
5. Gudbjartsson DF, Arnar DO, Helgadottir A, Gretarsdottir S, Holm H, Sigurdsson A, et al. Variants conferring risk of atrial fibrillation on chromosome 4q25. Nature. 2007; 448:353–7. [PubMed: 17603472]
6. Goldstein LB, Jones MR, Matchar DB, Edwards LJ, Hoff J, Chilukuri V, et al. Improving the reliability of stroke subgroup classification using the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) criteria. Stroke. 2001; 32:1091–8. [PubMed: 11340215]
7. Gordon DL, Bendixen BH, Adams HP Jr, Clarke W, Kappelle LJ, Woolson RF, The TOAST Investigators. Interphysician agreement in the diagnosis of subtypes of acute ischemic stroke: implications for clinical trials. Neurology. 1993; 43:1021–7. [PubMed: 8492920]
8. Lindley RI, Warlow CP, Wardlaw JM, Dennis MS, Slattery J, Sandercock PA. Interobserver reliability of a clinical classification of acute cerebral infarction. Stroke. 1993; 24:1801–4. [PubMed: 8248958]
9. Atiya M, Kurth T, Berger K, Buring JE, Kase CS, Women's Health Study. Interobserver agreement in the classification of stroke in the Women's Health Study. Stroke. 2003; 34:565–7. [PubMed: 12574576]
10. Selvarajah JR, Glaves M, Wainwright J, Jha A, Vail A, Tyrrell PJ. Classification of minor stroke: intra- and inter-observer reliability. Cerebrovasc Dis. 2009; 27:209–14. [PubMed: 19176952]
11. Meschia JF, Barrett KM, Chukwudelunzu F, Brown WM, Case LD, Kissela BM, et al. Siblings with Ischemic Stroke Study (SWISS) Investigators. Interobserver agreement in the trial of org 10172 in acute stroke treatment classification of stroke based on retrospective medical record review. J Stroke Cerebrovasc Dis. 2006; 15:266–72. [PubMed: 17904086]
12. Ay H, Furie KL, Singhal A, Smith WS, Sorensen AG, Koroshetz WJ. An evidence-based causative classification system for acute ischemic stroke. Ann Neurol. 2005; 58:688–97. [PubMed: 16240340]
13. Ay H, Benner T, Arsava EM, Furie KL, Singhal AB, Jensen MB, et al. A computerized algorithm for etiologic classification of ischemic stroke: the Causative Classification of Stroke System. Stroke. 2007; 38:2979–84. [PubMed: 17901381]
14. Arsava EM, Ballabio E, Benner T, Cole JW, Delgado-Martinez MP, Dichgans M, et al. International Stroke Genetics Consortium. The Causative Classification of Stroke system: an international reliability and optimization study. Neurology. 2010; 75:1277–84. [PubMed: 20921513]
15. Meschia JF, Arnett DK, Ay H, Brown RD Jr, Benavente OR, Cole JW, et al. NINDS SiGN Study. Stroke Genetics Network (SiGN) study: design and rationale for a genome-wide association study of ischemic stroke subtypes. Stroke. 2013; 44:2694–702. [PubMed: 24021684]
16. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960; 20:37–46.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

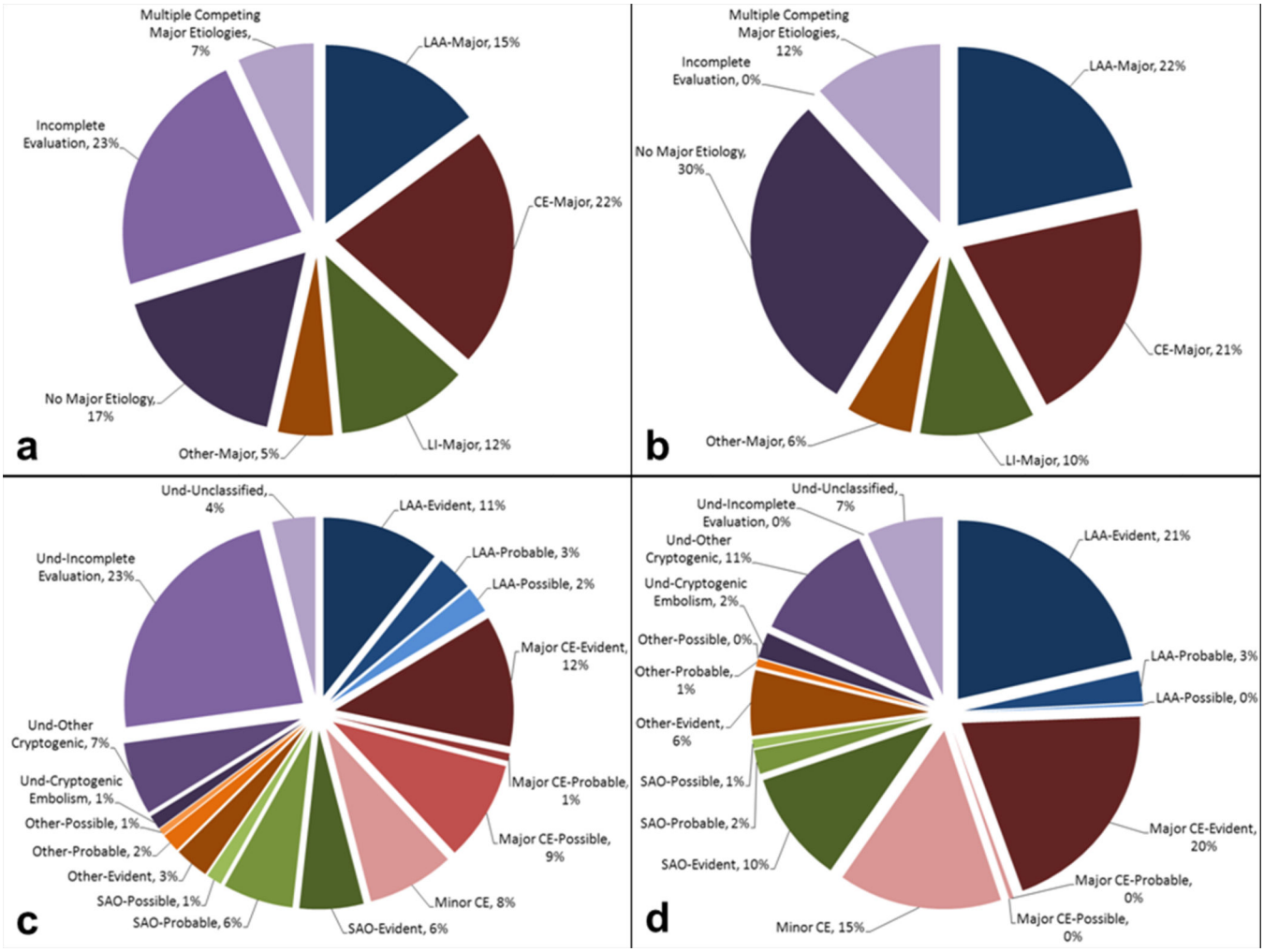

Figure 1.

Distribution of phenotypic and causative stroke subtypes: 1(a), phenotypic subtypes in the entire population; 1(b), phenotypic subtypes in the subset with complete vascular and cardiac investigation; 1(c), causative subtypes in the entire population; 1(d), causative subtypes in the subset with complete vascular and cardiac investigation. Please note that the term “incomplete evaluation” in Figures 1a and 1c designates an etiologic subgroup under “undetermined” category that is considered when diagnostic investigations are not performed in the absence of an identified etiology based on history and physical examination. According to this definition, a case with atrial fibrillation in history is not classified as incomplete evaluation when vascular and cardiac investigations are not done. The term “complete investigation” in Figures 1b and 1d, on the other hand, is solely based on availability of diagnostic tests indicating that brain imaging, vascular imaging, and cardiac evaluation are available. Hence, only 3,947 cases were classified as “incomplete evaluation”, while diagnostic investigations were not complete in 9,206 cases. Und: undetermined

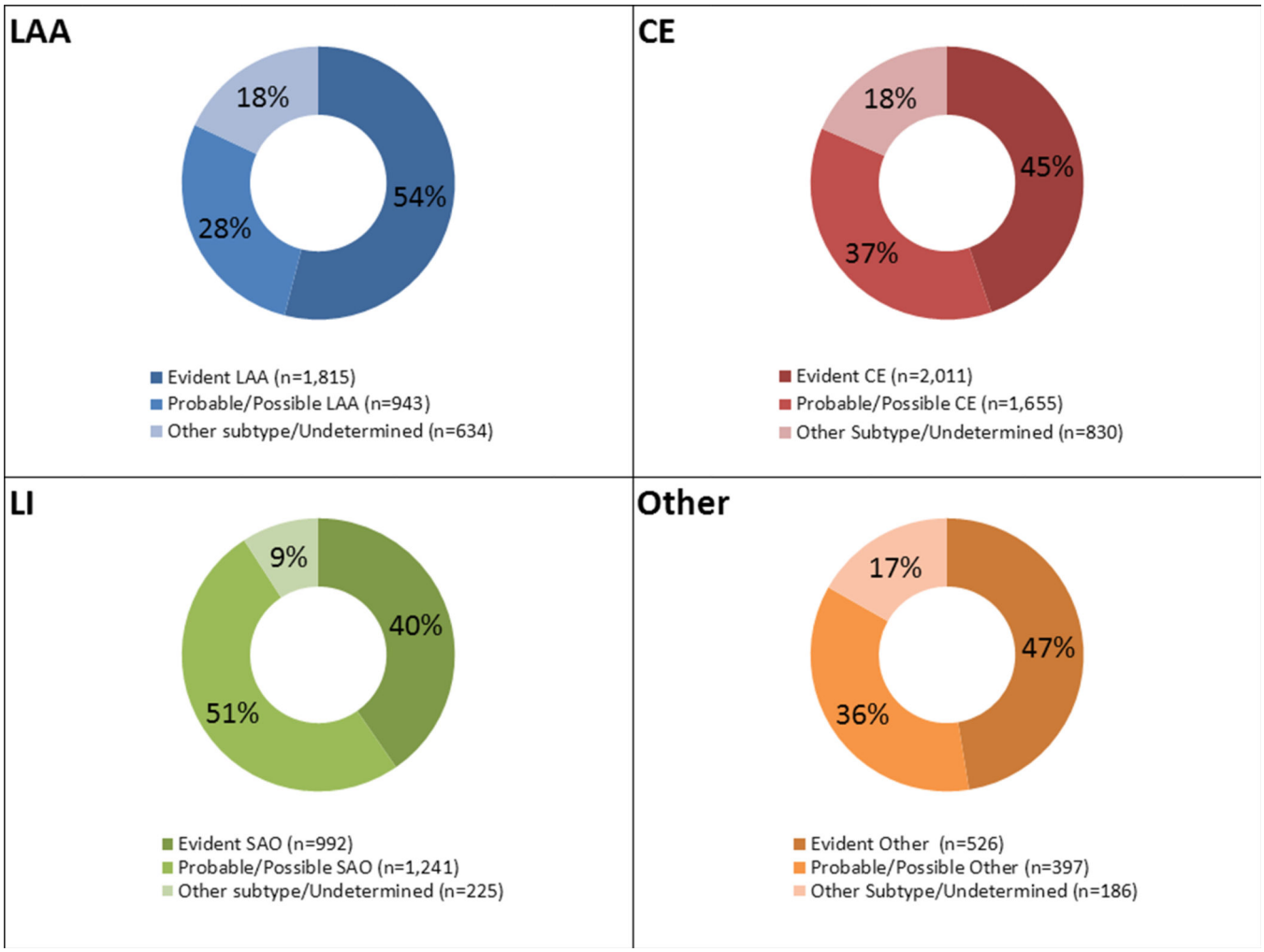

Figure 2.

Correlation between causative and phenotypic subtypes. Segments in each circle indicate proportion of causative subtypes in each major phenotypic category (circles).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

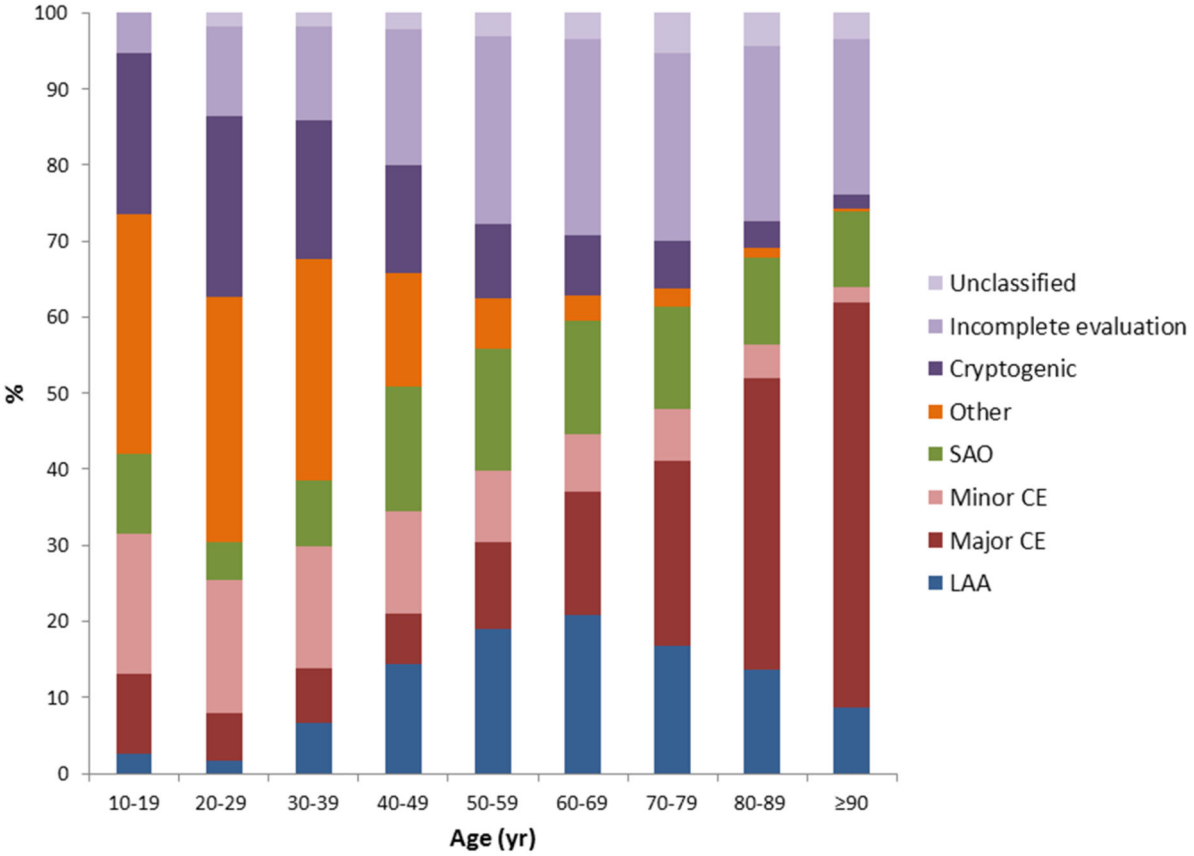

Figure 3.
The relationship between age and causative stroke subtypes.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

Patient characteristics

	Overall Study Population N=16,954	Complete Investigation N=7,748	Incomplete Investigation N=9,206
Age (mean, +/− SD)	67.1 (14.9) years	64.7 (15.7) years	69.1 (13.9) years
Female (%)	48.8	44.5	52.3
Race (%)			
Black	9.7	11.0	8.7
White	79.3	77.5	80.7
Other	11.0	11.5	10.6
Hypertension (%)	67.8	66.0	69.3
Diabetes Mellitus (%)	25.0	25.7	24.4
Atrial Fibrillation (%)	21.4	23.2	19.9
Coronary Artery Disease (%)	22.9	21.3	24.3
Current smoking (%)	24.1	24.4	23.8

Complete investigation is defined as the presence of brain imaging, cardiac evaluation with electrocardiography, echocardiography if other investigations did not reveal a source, and intracranial and extracranial vascular evaluation.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 2

Complete investigation rates across the contributing studies. The heterogeneity among studies was assessed by Chi-square test.

Study	Number of Cases	Complete Cardiac Investigation	Complete Vascular Investigation	Complete Cardiac and Vascular Investigation
1	684	40.9%	20.2%	4.8%
2	578	89.6%	75.3%	70.4%
3	840	71.8%	39.4%	30.1%
4	1,072	79.5%	98.5%	78.7%
5	331	98.8%	97.9%	96.7%
6	876	94.5%	79.3%	75.9%
7	598	58.2%	35.8%	21.9%
8	675	80.0%	79.4%	64.3%
9	1,088	64.0%	97.6%	61.7%
10	626	45.7%	13.9%	5.9%
11	642	81.2%	50.9%	43.3%
12	891	87.8%	68.2%	62.3%
13	470	57.7%	20.9%	13.0%
14	555	73.9%	42.9%	34.8%
15	643	95.2%	32.7%	30.6%
16	407	51.4%	40.8%	29.0%
17	1,085	78.6%	47.3%	37.4%
18	686	83.4%	92.7%	80.2%
19	554	58.3%	31.0%	19.0%
20	685	74.9%	92.3%	69.5%
21	524	93.9%	85.7%	80.5%
22	957	55.3%	31.2%	16.7%
23	1,487	85.3%	33.8%	29.1%
		p<0.001	p<0.001	p<0.001
Europe	9,360	68.8%	54.7%	39.7%
USA	7,594	81.6%	60.6%	53.1%
		p<0.001	p<0.001	p<0.001