Ischemic Stroke Outcomes

Analyses of Protein and Genetic Biomarkers

Annie Pedersen

Department of Laboratory Medicine
Institute of Biomedicine
Sahlgrenska Academy, University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2019

Cover photo/illustration by Tommy Pedersen/Maria Lagging

Illustrations by Maria Lagging

Ischemic Stroke Outcomes - Analyses of Protein and Genetic Biomarkers.

annie.pedersen@gu.se

ISBN 978-91-7833-344-8 (PRINT)
ISBN 978-91-7833-345-5 (PDF)
http://hdl.handle.net/2077/59055

Printed in Gothenburg, Sweden 2019
Printed by BrandFactory

ABSTRACT

The overall aim of this thesis was to identify novel biomarkers for ischemic stroke outcomes. The specific aims were to test the hypotheses that circulating concentrations of hemostatic biomarkers predict the long-term post-stroke risk of recurrent vascular events/death (paper I) and/or cognitive impairment (paper II) and that circulating concentrations of a marker of neuronal damage (neurofilament light chain, NfL) predict post-stroke functional and neurological outcomes (paper III) as well as to identify genetic variants associated with post-stroke functional outcome through a genome wide association study (GWAS) approach (paper IV).

Papers I-III are based on the first 600 cases and 600 controls recruited to the Sahlgrenska Academy Study on Ischemic Stroke, which includes consecutive ischemic stroke cases aged 18-69 years and sex- and age-matched population-based controls. In cases, blood sampling was performed in the acute phase, after three months, and in a subset also 7 years post-stroke. Controls were sampled once. These blood samples were used to analyze the protein and genetic biomarkers investigated in this thesis. Vascular events and death up to 14 years after inclusion were identified. In cases, functional and neurological outcomes were assessed at 3 months by the modified Rankin scale (mRS) and the NIH Stroke Scale (NIHSS), respectively. At 2 years, the mRS was assessed again, and in a subsample, long-term (7-year) outcomes were assessed by mRS and NIHSS. The 7-year follow-up also included cognitive testing with the Barrow Neurological Institute Screen for Higher Cerebral Functions (BNIS) and Trailmaking Test. Paper IV was based on a GWAS approach, i.e. genetic variations spread throughout the entire genome were analyzed with respect to their association to 3-month post-stroke functional outcome in a hypothesis free manner. This study was performed within the Genetics of Ischemic Stroke Functional Outcome (GISCOME) network, and included 6,165 ischemic stroke cases from 12 studies in Europe, USA and Australia.

In paper I, we found that plasma levels of hemostatic protein biomarkers were associated with vascular death and coronary events, but not with recurrent stroke. In paper II, we found that, in cases <50 years at index stroke, higher concentrations of the hemostatic protein fibrinogen were independently associated with worse cognitive outcome. In paper III, we found that acute phase and 3-month serum levels of NfL were independently associated to NIHSS and mRS both in short- and long-term follow-up. In paper IV, we

identified one genetic variant associated with functional outcome (mRS score 0-2 vs 3-6) at genome-wide significance. In addition, several genetic variants demonstrated suggestive associations, and some of these are located within or near genes with experimental evidence of influence on ischemic stroke volume and/or brain recovery.

In conclusion, the results from this thesis demonstrate associations between circulating protein levels as well as genetic markers and ischemic stroke outcomes. These results add knowledge on potential mechanisms influencing outcomes after ischemic stroke and may in the long run contribute to a more personalized post-stroke management.

Keywords: Stroke, Prognosis, Biomarkers, Genetics, Genome-Wide Association Study

POPULÄRVETENSKAPLIG SAMMANFATTNING (Swedish summary)

Stroke är den näst vanligaste dödsorsaken globalt och den vanligaste orsaken till förvärvad funktionsnedsättning hos vuxna. På senare år har insjuknande i stroke generellt minskat, men hos unga har tyvärr strokeinsjuknanden ökat. Unga strokepatienter överlever i större utsträckning jämfört med äldre, men kommer att leva med konsekvenserna av sin stroke under en lång tid. Trots detta vet man lite om faktorer som påverkar långtidsprognosen. Därför var det övergripande målet med den här avhandlingen att identifiera så kallade biomarkörer, mätbara ämnen, i det här fallet i form av äggviteämnen i blodet, vars nivåer kan predicera långtidsutfall, och variationer i arvsmassan som är associerad med utfall efter stroke. Biomarkörer kan bidra till en bättre information om prognos till den enskilda individen, men kan också ge fördjupad kunskap om mekanismerna bakom den stora variation som ses individer emellan avseende utfall efter stroke. På sikt kan sådan kunskap bidra till utvecklingen av en mer individanpassad och effektiv behandling och rehabilitering vid stroke.

Stroke uppstår då ett område i hjärnan drabbas av nedsatt blodtillförsel, som leder till syrebrist och vävnadsskada. Orsaken kan vara antingen en blodpropp, som hindrar blodtillförseln till en del av hjärnan, eller ett brustet kärl, det vill säga en blödning. Utifrån dessa två huvudorsaker brukar stroke delas in i ischemisk (hjärninfarkt) och hemorrhagisk (hjärnblödning). I den här avhandlingen studeras enbart ischemisk stroke, vilket är den vanligast stroketypen. I Sverige utgör den drygt 85% av alla strokefall.

I de ingående delarbetena i den här avhandlingen undersökte vi om olika äggviteämnens nivåer i blodet är kopplade till utfall efter stroke. De äggviteämnen vi studerat är dels ämnen som är involverade i blodets levringsförmåga, eftersom blodproppsbildning är en nyckelhändelse i utvecklingen av ischemisk stroke, och dels en nervskademarkör; ”neurofilament light chain” (NfL). Vi har också genom en så kallad ”genome wide association study” (GWAS) sökt efter varianter i arvsmassan som är kopplade till utfall efter stroke. Det finns många möjliga utfall att studera efter stroke. Vi har fokuserat på långtidsrisk för nya insjuknanden i kärlsjukdomar, såsom stroke och hjärtinfarkt, och på olika typer av funktionsnedsättningar genom att studera allmän funktionsnivå, kvarstående neurologiska bortfallssymtom och kognition.

Delarbete 1-3 baserades på de första 600 fallen med ischemisk stroke och 600 kontrollpersoner inkluderade i “the Sahlgrenska Academy Study on Ischemic Stroke” (SAHLSIS). Sammanfattningsvis inkluderar SAHLSIS individer som drabbats av ischemisk stroke mellan 18 och 69 års ålder och friska kontrollpersoner matchade för ålder och kön. Hos fallen togs blodprov vid insjuknandet, vid en 3-månadersuppföljning, samt hos en subgrupp också vid ett ytterligare uppföljningstillfälle efter 7 år. För kontrollpersonerna togs blodprov i samband med inklusion i studien. Blodproverna har använts för att analysera de biomarkörer som studerats i denna avhandling, både nivåer av äggviteämnen och genetiska varianter. Studiedeltagarna har följts upp med registrering av nya insjuknanden och död under upp till 14 år. Fallen med ischemisk stroke har också följts upp med tester för att skatta olika typer av utfall. Vid 3 månader skattades funktionellt utfall med ”the modified Rankin Scale” (mRS) och neurologiskt utfall med ”the NIH Stroke Scale” (NIHSS). Vid 2 år skattades mRS igen och vid 7 år både mRS och NIHSS. Vid 7 år skattades även kognitiv funktion. Delarbete 4 var en GWAS, vilket innebär att vanliga genvarianter utspridda över hela arvsmassan undersöks. Vi letade efter kopplingar mellan dessa varianter och funktionsnivå vid 3 månader, mätt med mRS. GWAS kräver ett stort antal deltagare och därför utfördes projektet som ett internationellt samarbete inom ramen för nätverket ”the Genetics of Ischemic Stroke Functional Outcome” (GISCOME). Genom detta nätverk samlades data från 6,165 fall med ischemisk stroke in från 12 studier i Europa, USA och Australien. SAHLSIS bidrog med data från 1,158 fall.

Resultaten av studierna visade att äggviteämnen involverade i blodets levringsförmåga var associerade med ökad långtidsrisk för hjärtinfarkt/koronarsjukdom och död, men inte med risk för återinsjuknande i stroke. För personer <50 år fanns också en association till sämre kognitiv funktion vid långtidsuppföljningen. För nervskademarkören NfL fann vi associationer till funktionellt och neurologiskt utfall både i ett kortare (3 månader till 2 år) och längre (7 år) perspektiv. Vi hittade också en genetisk variant som med stor statistisk säkerhet var associerad, och ett flertal varianter med möjlig koppling, till funktionellt utfall vid 3 månader.

Sammanfattningsvis visar resultaten från den här avhandlingen på kopplingar mellan i blodet cirkulerande äggviteämnen samt varianter i arvsmassan och olika utfall efter stroke. Resultaten bidrar med kunskap om potentiella mekanismer bakom utfall efter stroke och kan i ett långtidsperspektiv gynna utvecklingen av ett mer individanpassat omhändertagande av personer som drabbas av stroke.

PAPERS INCLUDED IN THE THESIS

This thesis is based on the following papers, referred to in the text by their Roman numerals:

I. Pedersen A, Redfors P, Lundberg L, Gils A, Declerck PJ, Nilsson S, Jood K, Jern C. Haemostatic biomarkers are associated with long-term recurrent vascular events after ischaemic stroke. *Thromb Haemost. 2016;116: 537-543.*

II. Pedersen A, Stanne TM, Redfors P, Viken J, Samuelsson H, Nilsson S, Jood K, Jern C. Fibrinogen concentrations predict long-term cognitive outcome in young ischemic stroke patients. *Res Pract Thromb Haemost. 2018;2:339-346.*

III. Pedersen A, Stanne TM, Nilsson S, Klasson S, Rosengren L, Holmegaard L, Jood K, Blennow K, Zetterberg H, Jern C. Circulating neurofilament light in ischemic stroke: Temporal profile and outcome prediction. *Submitted manuscript.*

IV. Söderholm M*, Pedersen A*, Lorentzen E, Stanne TM, Bevan S, Olsson M, Cole JW, Fernandez-Cadenas I, Hankey GJ, Jimenez-Conde J, Jood K, Lee J-M, Lemmens R, Levi C, Mitchell BD, Norrving B, Rannikmäe K, Rost NS, Rosand J, Rothwell PM, Scott R, Strbian D, Sturm JW, Sudlow C, Traylor M, Thijs V, Tatlisumak T, Woo D, Worrall BB, Maguire JM**, Lindgren A**, Jern C**, on behalf of the International Stroke Genetics Consortium, the NINDS-SiGN Consortium, and the Genetics of Ischaemic Stroke Functional Outcome (GISCOME) network. Genome-wide association meta-analysis of functional outcome after ischemic stroke. *Neurology, 2019;92:e1271-e1283.* *These authors contributed equally to this work. **These authors jointly supervised this work.

The papers are appended at the end of the thesis. Reprints were made with permission from the publishers.

Paper IV was the subject of an editorial, A SNP-it of stroke outcome. *Neurology, 2019;92:549-550*.

CONTENTS

ABBREVIATIONS

A	adenine
AUC	area under the ROC curve
BBB	blood brain barrier
BNIS	Barrow Neurological Institute Screen for Higher Cerebral Functions
C	cytosine
CCS	Causative Classification of Stroke
CE	cardioembolic
CNS	central nervous system
CNV	copy number variation
CRP	C-reactive protein
CT	computer tomography
DALYs	disability-adjusted life years
DNA	deoxyribonucleic acid
ECG	electrocardiogram
ECL	electrochemiluminescence immunoassay
ELISA	enzyme-linked immunosorbent assay
eQTL	expression quantitative trait locus
ESUS	embolic stroke of undetermined source
G	guanine
GISCOME	Genetics of Ischemic Stroke Functional Outcome
GWAS	genome wide association study
HIV	human immunodeficiency virus
HR	hazard ratios
ICD10	International Classification of Diseases 10th Revision
ISGC	International Stroke Genetics Consortium
LAA	large-artery atherosclerosis
LACI	lacunar infarct
LD	linkage disequilibrium
LDL	low-density lipoprotein
lncRNA	long non-coding RNA
LVD	large-vessel disease
MAF	minor allele frequency
MMSE	Mini-Mental State Examination
MRI	magnetic resonance imaging
mRNA	messenger RNA
mRS	modified Rankin Scale
NCBI	National Center for Biotechnology Information
NfL	neurofilament light chain
NIHSS	National Institutes of Health Stroke Scale
OCSP	Oxfordshire Community Stroke Project
OR	odds ratio
PACI	partial anterior circulation infarct

PAI-1	plasminogen activator inhibitor type 1
PCA	principal component analysis
PFO	patent foramen ovale
POCI	posterior circulation infarct
PP1	protein phosphatase 1
PROM	patient reported outcome measure
QC	quality control
RCT	randomized controlled trial
RNA	ribonucleic acid
ROC	receiver operating characteristic curve
SAHLSIS	the Sahlgrenska Academy Study on Ischemic Stroke
SAO	small-artery occlusion
SiMoA	single-molecule array
SNP	single nucleotide polymorphism
SSS	Scandinavian Stroke Scale
SVD	small vessel disease
T	thymine
TACI	total anterior circulation infarct
TAFI	thrombin activatable fibrinolysis inhibitor
TAFI-AP	thrombin activatable fibrinolysis inhibitor activation peptide
TIA	transient ischemic attack
TOAST	Trial of Org. 10172 in Acute Stroke Treatment
t-PA	tissue-type plasminogen activator
UTR	untranslated region
VWF	von Willebrand factor
WHO	World Health Organization

INTRODUCTION

Stroke – past and present

Acute cerebrovascular disease, or what we refer to as a "stroke" was recognized by Hippocrates and the ancient Greeks as "apoplexy", meaning "struck with violence as if by a thunderbolt" [1]. Several aspects of the condition were noted at the time; an association of right-sided paralysis with loss of speech, that attacks of numbness or anaesthesia could be signs of impending apoplexy, and that people between the ages of 40 and 60 were most susceptible [1]. Ever since then, advancements have been made towards an understanding of the pathophysiology underlying the condition. In a book on apoplexy published in 1619 by Gregor Nymman of Wittenberg it was recognized that an apoplectic attack could result from closure of the vessels that bore the vital spirits to the brain by emboli from concretions within the heart [1]. In the 17th century, Johann Jacob Wepfer first suggested that apoplexy could be due to hemorrhage, but also to blockage of one of the main arteries supplying blood to the brain [2, 3]. Thus, stroke was recognized as a cerebrovascular disease. In the 20th century, with the development of cerebral angiography, the role of extracranial vessels in the etiology of stroke was more deeply understood [1]. Today an etiologic mechanism can be established in a majority of cases. However, as will be described, there are still a considerable proportion of cases with ischemic stroke, the type of stroke studied in this thesis, for which the underlying etiology cannot be established despite an extensive work-up.

Throughout history, people suffering and surviving a stroke have often been left with severe disabilities. Available care has been limited and focused on strategies to manage with sequelae such as persisting weakness, speech impairments or eating difficulties. Not until the 1990s, an effective pharmacological acute treatment of stroke was implemented with the approval of the use of tissue-type plasminogen activator (t-PA) for ischemic stroke. During the same period, organizational changes were adopted to promote increased teamwork and knowledge, including care and early mobilization at specific stroke units. In the beginning of the 21st century, additional advancement in acute ischemic stroke treatment was made by the introduction of mechanical thrombectomy, i.e. physical removal of the blood clot causing the stroke. These acute treatments represent recent and important achievements in acute stroke care [4-6]. Also rehabilitation strategies as well as primary and secondary prevention have developed during this period. Despite this, stroke is today the second most common cause of death world-wide [7] and the third

most common cause of acquired disability in adults, measured in disability-adjusted life years (DALYs) [8].

Over the last decades, the age-standardized incidence of stroke in high-income regions as well as the global mortality caused by stroke has decreased. However, due to ageing and population growth, the overall stroke burden in terms of absolute number of people affected by, or who remain disabled from stroke has still increased [9]. Moreover, the incidence of stroke is increasing in younger ages [10, 11]. This means there will be an increasing number of stroke survivors with life-long impairments. Today, 60% of the people currently living who have experienced a stroke are under the age of 70 [12]. The type and degree of persisting impairments can be designated the post-stroke outcome, and is a result of the brain injury as well as recovery. Outcome can also refer to mortality or new cardiovascular events. There is a great inter-individual variability in post-stroke outcomes and the mechanisms behind this variability are not fully understood. These unexplored mechanisms represent the starting point of this thesis. They are potential keys in the search for a personalized post stroke management approach that include accurate prognostic prediction as well as patient-tailored treatment, secondary prevention and rehabilitation. Such an approach has the potential to reduce morbidity and mortality and improve quality of life for stroke sufferers.

Stroke – definition, pathophysiology and etiologic subtypes

Today, the World Health Organization (WHO) definition of stroke is "rapidly developing clinical signs of focal, and at times global, loss of cerebral function, with symptoms lasting more than 24 hours or leading to death, with no apparent cause other than that of vascular origin" [13]. These clinical manifestations are the result of reduced blood flow to an area of the brain leading to local injury. The individual consequences are dependent on the size and location of the injury, the general brain condition, and also the underlying pathophysiological mechanism.

There are two main causes of stroke, a blocked vessel (ischemic stroke) or a ruptured vessel (hemorrhagic stroke). The focus of this thesis is ischemic stroke, which is most common world-wide and in high-income countries, such as Sweden, accounts for over 85% of all stroke cases [14]. Most commonly, ischemic stroke is caused by a blood clot that partially or totally obstructs a vessel. The result is a sudden decrease or stop of the blood flow leading to

focal ischemia and, subsequently cellular death. There are different pathophysiological mechanisms underlying ischemic stroke, and based on these mechanisms, it can be further divided into etiological subtypes, Figure 1. The most frequently used classification system for etiologic subtypes of ischemic stroke is the Trial of Org. 10172 in Acute Stroke Treatment (TOAST) [15], which includes the subtypes: large-artery atherosclerosis (LAA), cardioembolism, small-artery occlusion (SAO), other determined etiology (a mix of specified unusual causes), and undetermined etiology. In this thesis, the subtypes LAA and SAO are referred to as large vessel disease (LVD) and small vessel disease (SVD), respectively. The TOAST classification focuses on pathophysiological mechanisms and is based on clinical symptoms as well as results from investigations in the diagnostic work-up. There are also other classification systems for stroke, such as the Oxfordshire Community Stroke Project (OCSP) [16]. The OCSP classification is based on clinical presentation and separates strokes of different sizes and locations into: total anterior circulation infarct (TACI), partial anterior circulation infarct (PACI), posterior circulation infarct (POCI), and lacunar infarct (LACI).

In ischemic stroke, prognosis as well as optimal treatment and prevention strategies are dependent on the underlying mechanism. Therefore, etiologic subtypes are of importance in ischemic stroke research. In the following sections, the main etiologic subtypes are further described.

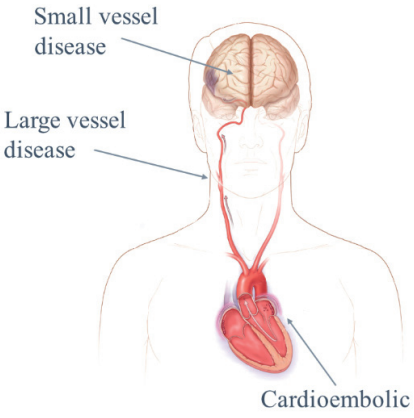


Figure 1. Three of the four major etiologic subtypes of ischemic stroke. The fourth, cryptogenic stroke, is characterized by the lack of an identifiable cause, despite an extensive work-up.

Large vessel disease

LVD refers to stroke caused by atherosclerosis in large or medium sized cerebral or precerebral arteries. The cerebral ischemia may be the result of artery to artery embolization from an atherosclerotic lesion, or of hemodynamic mechanisms. LVD is recognized by significant stenosis or occlusion of a major brain artery or branch cortical artery, and typically leads to cortical lesions, but also cerebellar, brain stem or subcortical infarcts. This subtype also has a less favourable prognosis regarding survival and short-term recurrence risk, as well as long-term risk of cardiovascular events [17-19]. Subtype-specific secondary prevention includes carotid endarterectomy to reduce the risk of recurrent stroke. LVD accounts for around 20% of all ischemic strokes [20], but is more frequent in Asian populations [20, 21].

Small vessel disease

SVD leads to brain ischemia and stroke by occlusion of a single deep perforating end-artery arising from the circle of Willis or from the basilar artery [22]. The mechanism causing the vessel occlusion in SVD is not fully understood, but microatheroma, lipohyalinosis and intimal thickening, and wall fibrosis has been suggested [22, 23]. SVD infarcts are small, usually with a diameter less than 15 mm. Typical locations are in the deep white matter of the brain, the thalamus, striatum and the paramedian and lateral regions of the brainstem. Clinical manifestations are characterized by a so called lacunar syndrome, including “pure motor stroke”, “pure sensory stroke”, “sensorimotor stroke”, “ataxic hemiparesis” and “dysarthria-clumsy hand syndrome” [24]. Common to these syndromes is the absence of cortical symptoms. The main risk factor for SVD is hypertension, thus blood pressure control is of great importance also in secondary prevention. This subtype typically leads to milder strokes and has a comparatively favourable post-stroke prognosis [17-19]. SVD causes approximately 20% of all ischemic strokes[20].

Cardioembolic stroke

Cardioembolism is the cause of cardioembolic (CE) stroke. In other words, an embolus originating from the heart follows the blood stream until it obstructs a cerebral artery leading to focal brain ischemia. Several conditions are associated with increased risk of thrombus formation in the heart. The most common cause of CE stroke is atrial fibrillation. Other examples of high-risk sources of cardioembolism include mechanical prosthetic valves, recent myocardial infarction, dilated myocardiopathy and endocarditis. CE strokes tend to be large and severe and as a majority is caused by atrial fibrillation this subtype is more common in older individuals [25, 26]. The prognosis is comparatively poor with a high risk for recurrent cardiovascular events and death [17-19]. Anticoagulant drugs are effective for secondary prevention in patients with atrial fibrillation, but also in other types of CE stroke. About 22% of ischemic strokes are caused by cardioembolism, but apart from being more common in elderly, this subtype is more frequent in white, and less frequent in Asian populations [20].

Cryptogenic stroke

The group classified as undetermined etiology according to the TOAST criteria is either due to an incomplete evaluation, more than one identified etiology, or no identified cause despite an extensive work-up. In this thesis, the latter group is referred to as cryptogenic stroke. Of note, this definition of cryptogenic stroke resembles the concept of embolic stroke of undetermined source (ESUS) [27]. The mechanisms leading to cryptogenic stroke is by definition unknown, but most probably, this group is etiologically heterogeneous. The role of patent foramen ovale (PFO) and paradoxical embolism has been debated for a long time. Around one-half of cases with cryptogenic stroke <60 years of age have a PFO, which is almost double the prevalence in the general population [28]. Since, in patients with PFO and cryptogenic stroke, PFO closure is associated with a small reduction of recurrent stroke compared with medical therapy, this seems to be a contributing factor in some cases [28], and especially in specific subgroups, e.g. in individuals with activated protein C resistance. Generally, patients suffering from cryptogenic stroke are relatively young and have a more favourable prognosis compared to for instance LVD and CE stroke [17]. In young and middle-aged stroke sufferers cryptogenic stroke accounts for almost 30% of all ischemic stroke cases [29].

Other determined etiology

The subtype of other determined etiology includes a variety of more unusual causes of ischemic stroke that collectively account for around 3% of all cases [20]. In younger age groups this proportion is higher, 18-26% in cases up to 55 years [30, 31]. The most common other determined cause is arterial dissection. Other examples are vasculitis, monogenic diseases and hypercoagulable states or hematologic disorders.

Prognosis and outcomes

Ischemic stroke is a heterogeneous disease, not only regarding pathophysiological mechanisms, but also prognosis and persisting disabilities post-stroke. Many patients with ischemic stroke recover from their initial neurological impairments during the first 3 months after the event, but also beyond 3 months [32]. However, there is a great inter-individual variability, with some individuals reaching full recovery while others are left with persistent

severe disability [33]. Long-term outcomes, i.e. several years, after stroke are less studied. This is of specific significance for younger stroke patients who have a lower case-fatality and may live with the consequences of their stroke for decades. Long-term consequences include persistent neurological and functional impairments, but also the risk of recurrent vascular events [34] and excess mortality [35]. In addition, impairments such as cognitive impairment, depressive symptoms and fatigue are common [36]. These often have a great impact on quality of life as they determine the ability to resume daily activities such as work, general mobility and participation in social life.

To conclude, post-stroke consequences are diverse. Therefore, studies of post-stroke outcomes should preferably include measures that take this diversity into account. In addition, outcomes in different time intervals post-stroke are of relevance, from the first months to several years after the event. In this thesis, the main focus is on different outcomes assessed several years after ischemic stroke in relatively young stroke sufferers. In the following sections the outcome measures that we used are further described.

Functional outcome

Functional outcome after stroke refers to the ability to manage activities of daily life or the degree of persisting disability and dependence. This ability depends on motor and perceptual functions as well as cognitive functions, including language [37]. For overall stroke, there is data showing that up to 28% of stroke survivors are dependent on others to manage self-care and personal activities of everyday living one year after their stroke [38]. In ischemic stroke specifically, recent data shows that 35% of survivors that were previously independent are classified as dependent at five years post-stroke [39]. To measure functional outcome different scales can be used. The Barthel Index [40] is commonly applied as a measure of basic activity of daily living, but the most widely used scale for measuring post-stroke functional outcome in clinical trials is the modified Rankin Scale (mRS) [41], Figure 2. The mRS measures the degree of disability or dependence in daily activities. It is a 7-grade scale that is easy to apply and not very time-consuming. However, the mRS is a relatively crude and non-specific outcome measure that can be influenced also by comorbidities and socioeconomic factors [42]. Another potential problem is the inter-rater variability, which can be improved by structured assessments [43, 44].

Figure 2. Description of the scores in the modified Rankin Scale (mRS).

0	No symptoms.
1	No significant disability. Able to carry out all usual activities, despite some symptoms.
2	Slight disability. Able to look after own affairs without assistance, but unable to carry out all previous activities.
3	Moderate disability. Requires some help, but able to walk unassisted.
4	Moderately severe disability. Unable to attend to own bodily needs without assistance, and unable to walk unassisted.
5	Severe disability. Requires constant nursing care and attention, bedridden, incontinent.
6	Dead.

Neurological outcome

The neurological impairments caused by a stroke can be assessed by stroke severity scales. The most widely used stroke severity scale is the National Institutes of Health Stroke Scale (NIHSS), Figure 3. The NIHSS includes assessments of level of consciousness, eye movements, visual fields, facial palsy, motor and sensory function, ataxia, language and speech, and neglect. It is used during the acute phase of stroke both in clinical practice and in observational and interventional studies. Repeated measurements can be used to quantify a patient's improvement or decline or to monitor the effectiveness of a treatment. The NIHSS score correlates to infarct volume, but the score is dependent on the infarct location [45, 46]. The NIHSS score is also a predictor of functional outcome and mortality at 3 months post-stroke [47]. Another example of a stroke severity scale is the Scandinavian Stroke Scale (SSS) that is less widely used. The SSS does not include as many items as the NIHSS and lacks evaluation of visual fields, ataxia and neglect, but it predicts 3-month outcome with a similar accuracy as the NIHSS [48].

Figure 3. Description of the National Institutes of Health Stroke Scale (NIHSS).

1a. Level of Consciousness	0 = alert; keenly responsive 1 = not alert, but arousable by minor stimulation 2 = not alert; requires repeated stimulation 3 = unresponsive or responds only with reflex
1b. Level of Consciousness questions: What is the month? What is your age?	0 = answers two questions correctly 1 = answers one question correctly 2 = answers neither question correctly
1 c. Level of Consciousness commands: Open and close your eyes. Grip and release your hand.	0 = performs both tasks correctly 1 = performs one task correctly 2 = performs neither task correctly
2. Best gaze	0 = normal 1 = partial gaze palsy 2 = forced deviation
3. Visual fields	0 = no visual loss 1 = partial hemianopia 2 = complete hemianopia 3 = bilateral hemianopia
4. Facial palsy	0 = normal symmetric movements 1 = minor paralysis 2 = partial paralysis 3 = complete paralysis of one or both sides
5. Motor arm 5a. Left arm 5b. Right arm	0 = no drift 1 = drift 2 = some effort against gravity 3 = no effort against gravity; limb falls 4 = no movement
6. Motor leg 6a. Left leg 6b. Right leg	0 = no drift 1 = drift 2 = some effort against gravity 3 = no effort against gravity 4 = no movement
7. Limb ataxia	0 = absent 1 = present in one limb 2 = present in two limbs
8. Sensory	0 = normal; no sensory loss 1 = mild-to-moderate sensory loss 2 = severe to total sensory loss
9. Best language	0 = no aphasia; normal 1 = mild to moderate aphasia 2 = severe aphasia 3 = mute, gobal aphasia
10. Dysarthria	0 = normal 1 = mild to moderate dysarthria 2 = severe dysarthria
11. Extinction and inattention	0 = no abnormality 1 = one of the sensory modalities 2 = profound hemi-inattention or extinction

Cognitive outcome

Cognitive impairment post-stroke is common and impacts both social functioning and the ability to work [49-51]. Batteries of cognitive tests measuring different cognitive domains can be used to detect post-stroke cognitive impairment, but with the disadvantage of being time-consuming and tiring for the patients. Screening tests can be useful in order to identify individuals that would benefit from more comprehensive neuropsychological assessments. A commonly used scale to screen for cognitive dysfunction is the Mini-Mental State Examination (MMSE), but due to the fact that it has a ceiling effect and that it cannot provide a full cognitive profile, it seems suboptimal for the screening of cognitive dysfunction post-stroke [52]. As an alternative, the Barrow Neurological Institute Screen for Higher Cerebral Functions (BNIS) that includes assessments of a broader range of cognitive functions has proven useful for screening cognitive function in the long-term post-stroke [52], Figure 4. A good validity has been described for BNIS for the discrimination of patients with brain lesions [53] and for the screening of cognitive impairments in stroke patients [54]. While the BNIS represents a screening test for global cognitive function, there are numerous tests assessing specific cognitive domains. One example that we used in this thesis is the Trailmaking Test [55] that includes two parts, A and B. Part A primarily assesses processing speed, and part B further requires executive functions such as working memory and attentional shifting.

Recurrent vascular events and mortality

In addition to persistent impairments, stroke sufferers also face an increased risk of recurrent vascular events, such as stroke and coronary events, and death. The reported risk of recurrent stroke varies between studies with five-year recurrence rates between 10 and 30% [19, 34, 56]. In studies with cases of all ages, corresponding figures for coronary events are 5-14% [18, 57, 58]. For mortality, five-year rates between 40 and 60% have been reported [18, 59, 60]. However, risks are generally lower in younger cases [34, 56, 61]. In Sweden, several national registers enable investigations of recurrence rates and mortality with high coverage.

Figure 4. Description of the Barrow Neurological Institute Screen for Higher Cerebral Functions (BNIS) items.

BNIS items	Score	Subscale score
Pre-screening		
Level of consciousness	3	
Basal communication	3	
Cooperation	3	9
Speech and language		
Fluency	1	
Paraphasia	1	
Dysarthria	1	
Comprehension	2	
Naming	1	
Repetition	2	
Reading	1	
Writing – sentence copying	1	
Writing – dictamen	1	
Spelling – irregular	1	
Spelling – phonetic	1	
Arithmetic – number/symbol alexia	1	
Arithmetic – dyscalculia	1	15
Orientation		
Left-right orientation	1	
Place orientation	1	
Time orientation	1	3
Attention/concentration		
Arithmetic memory/concentration	1	
Digits – forward	1	
Digits - backward	1	3
Visuospatial and visual problem-solving		
Visual object recognition	1	
Constructional praxis dominant hand	1	
Constructional praxis non-dominant hand	1	
Visual scanning	2	
Visual sequencing	1	
Pattern copying	1	
Pattern recognition	1	8
Memory		
Number/symbol test	4	
Delayed recall	3	7
Affect		
Affect expression	1	
Affect perception	1	
Affect control	1	
Spontaneous affect	1	4
Awareness		
Awareness vs performance	1	1
Total score	50	

Protein and genetic biomarkers of ischemic stroke outcomes

As highlighted in previous sections, there is a wide range of inter-individual variability in ischemic stroke outcomes. There are some well-known clinical factors associated with outcomes such as age, sex, stroke severity and etiologic stroke subtype [17, 62-65]. Other prognostic factors include prior functional status, social factors, post-stroke depression and comorbidities, and the treatment and rehabilitation obtained [62, 66-71]. However, prediction models based on clinical variables do not fully explain the variability in stroke outcomes. Thus, there is reason to suspect a role of additional, not yet identified determinants, for instance factors that modulate the response to ischemic brain injury, brain plasticity and recovery or risk of recurrent vascular events and death. The search for such novel biological pathways can be conducted by evaluating "biomarkers" in relation to different traits such as outcomes or severity. A biomarker is defined as a characteristic that is measured objectively and evaluated as an indicator of normal biological or pathogenic processes, or of pharmacological responses to a therapeutic intervention [72]. This definition includes both circulating proteins and genetic variants that are the focuses of this thesis, but also other measures such as radiological characteristics or proteins in for instance the cerebrospinal fluid.

The study of biomarkers of stroke outcomes can be considered to have two conceptually different purposes: 1) To improve our understanding of the biological pathways underlying disease mechanisms, which can inform future studies and possibly guide in the search for new molecular targets for treatments and secondary prevention. 2) To identify biomarkers that are directly useful in a clinical setting, for instance biomarkers that contribute to better individual prognostic prediction or patient-tailored management by identifying subgroups that benefit from distinct treatments, secondary prevention and/or rehabilitation strategies.

Examples of circulating biomarkers with associations to stroke outcomes are copeptin [73], brain natriuretic peptide [74], C-reactive protein (CRP) [75], and lipoprotein-associated phospholipase A2 [76]. However, the requirements for a biomarker to be clinically useful are several. It should provide information at the individual level by being specific, but also sensitive to avoid false negatives. There should be normative values that are easily interpretable and ultimately a specific cut off for an intervention or clinical decision. The provided information should be additive to already established parameters such as clinical factors and imaging. In addition, measurements using standardized,

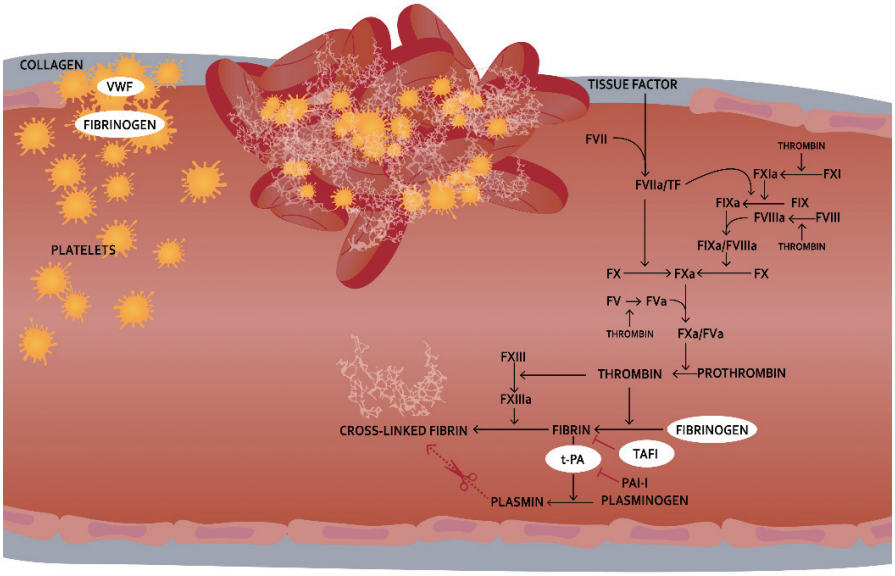
precise, reliable and reproducible assays must be available at a reasonable cost. Currently no biomarker of stroke outcomes has fully met these requirements and qualified for use in clinical routine [77]. Nevertheless, given the heterogeneous nature of ischemic stroke a more individualized management is necessary in order to optimize outcomes for ischemic stroke patients. Novel biomarkers have the potential to provide some of the lacking information for this to be achieved. Therefore, the aim of this thesis was to identify biomarkers for ischemic stroke outcomes. In the following sections, the specific biomarkers studied in the included papers are described.

Hemostasis

Hemostasis is the physiological process that serves to limit bleeding from injured vessels through the formation of blood clots. At the same time, the maintenance of blood fluidity within the vascular system needs to be ensured. Consequently, hemostasis is a strictly balanced interplay between a large number of proteins including both positive feedback mechanisms and inhibitors at multiple steps of the pathway. Abnormalities in proteins involved in these pathways can result in either excessive bleeding or pathologic blood clot formation, thrombosis. Since blood clot formation is a key mechanistic event in both ischemic stroke and myocardial infarction, hemostatic proteins are of interest when searching for biomarkers of significance for these conditions. For the same reason, hemostatic proteins are candidate predictors also of recurrent vascular events after ischemic stroke. Regarding other post-stroke outcomes, there are experimental data showing that some hemostatic proteins influence processes in the brain related to stroke recovery making them of interest also for functional and cognitive outcomes. As an example, the tissue-type plasminogen activator (t-PA) is directly involved in neuronal plasticity [78]. Moreover, results from both prospective and case-control studies have suggested a relationship between hemostatic biomarkers and cognitive impairment [79, 80], as well as vascular dementia [81, 82].

The hemostatic response to a vessel wall injury includes the following key steps: 1) vasoconstriction, i.e. instant narrowing of the vessel wall in order to limit bleeding, 2) platelet activation leading to the formation of a first platelet plug, 3) formation of a blood clot through stabilization by a fibrin network, and 4) eventually dissolving the clot (fibrinolysis). The main features of the 3 latter steps are briefly described below and illustrated in Figure 5.

Figure 5. Schematic overview of selected key components in the formation of a platelet plug, the coagulation cascade (formation of a blood clot), and fibrinolysis. The proteins studied in the papers of this thesis are highlighted in the figure.



F designates factor, and a the active form of the factors.

Platelet activation and the formation of a platelet plug

Upon vessel wall injury the subendothelial matrix (collagen) is exposed to the bloodstream. Exposed collagen interacts with receptors on the surface of circulating platelets leading to platelet adhesion and platelet activation. Platelet activation results in a change in shape of the platelets that facilitates aggregation, and secretion of the content of the platelets' granules, including the proteins von Willebrand factor (VWF) and fibrinogen. This in turn attracts more platelets and allows platelet aggregation mediated by VWF and fibrinogen, which act as bridges between the platelets. The result is a short-lived primary platelet plug.

Coagulation cascade and the formation of a blood clot

The coagulation cascade is triggered by vessel wall injury leading to exposure of the glycoprotein tissue factor. This constitutes the initiation of the cascade that includes a series of enzymatic conversions of proenzymes to activated enzymes. A key event is prothrombin being activated to thrombin, which subsequently converts fibrinogen to fibrin. Fibrin monomers spontaneously polymerize and become cross-linked, thus forming a three-dimensional mesh in which red blood cells and platelets are trapped. This constitutes the blood clot. In addition to the prothrombotic steps of the cascade, including positive feedback mechanisms, there are several antithrombotic control mechanisms and inhibitors essential to maintain a well-controlled system.

Fibrinolysis

Blood clots are removed by the body in the process of fibrinolysis. In the presence of factors involved in blood clot formation, i.e. thrombin, t-PA is released from the endothelium. T-PA can activate a circulating proenzyme, plasminogen, to plasmin. Plasmin then cleaves fibrin at multiple sites, which breaks down the clot and produces circulating fragments that are cleared by other proteases or by the kidney and liver. The presence of fibrin leads to a dramatic increase in t-PA activity, and in this way fibrinolysis is triggered where and when it is needed. Again, to maintain a balanced and controlled system there are also inhibitors of the fibrinolysis. For instance, plasminogen activator inhibitor type 1 (PAI-1) inhibits t-PA, and thrombin activatable fibrinolysis inhibitor (TAFI) acts as a functional inhibitor of fibrinolysis by making fibrin more resistant to degradation.

One of the aims of this thesis was to investigate associations between hemostatic proteins and post-stroke outcomes, specifically recurrent vascular events and cognitive function in the long-term. The hemostatic proteins that we selected are described in the following sections and highlighted in Figure 5.

Fibrinogen

Fibrinogen is a circulating glycoprotein primarily synthesized and secreted into the blood by hepatocytes. As described above, it is involved in platelet aggregation where it binds to platelet receptors linking the platelets together to form the platelet plug. Fibrinogen also plays a key role in the coagulation

cascade by being converted by thrombin to fibrin monomers, subsequently constituting the blood clot. In addition, fibrinogen is involved in several biological pathways relevant to atherothrombotic diseases such as inflammation and atherogenesis. It also contributes to plasma viscosity. Along with a number of other inflammatory proteins, e.g. IL-6 and CRP, fibrinogen is an acute phase reactant, which means that it increases as part of the body's acute response to systemic inflammation or tissue injury such as cerebral ischemia.

Large prospective studies have established a relationship between increased circulating fibrinogen concentrations and the risk of future cardiovascular events including stroke [83, 84]. Fibrinogen is also the hemostatic biomarker that has been most studied in relation to post-stroke outcomes. Independent associations between high acute phase fibrinogen concentrations and functional outcome or death during the first months after overall stroke have been demonstrated [85], but conflicting results exist in several studies of ischemic stroke [86-89]. Acute phase fibrinogen concentrations have also been found to predict 1-year mortality after ischemic stroke as well as recurrent vascular events and death during the first two years post-stroke [90-92]. Still, results from different studies are contradicting and a study with a follow-up of 7.4 years found no independent association [93]. Fewer studies have investigated fibrinogen concentrations measured after passing the acute phase of stroke, but associations to recurrent vascular events and cardiovascular death have been demonstrated during follow-up times ranging between around 2 and 10 years [94, 95].

After blood brain barrier (BBB) disruption fibrinogen is deposited in the central nervous system (CNS) where it has direct effects on microglia, astrocytes, and neurons. In the acute phase of CNS injuries, these depositions of fibrinogen may be beneficial. However, at later stages after injury, excessive fibrinogen deposition can be deleterious by increasing inflammation and preventing repair processes such as neurite outgrowth [96, 97] and re-myelination of axons [98]. Experimental data also indicate that fibrinogen has a role in the pathophysiology of Alzheimer's disease [99]. In light of this, fibrinogen is the most studied circulating hemostatic biomarker for cognitive impairment and dementia in man. Associations to incident vascular dementia have been demonstrated in prospective studies with mean follow-up times of up to seventeen years [81, 100]. In addition, in a population with atherosclerosis, but no cardiovascular disease at baseline, an association to cognitive decline during a five-year follow-up was demonstrated [101].

von Willebrand factor (VWF)

VWF is a large multimeric protein synthesized and stored both in megakaryocytes and endothelial cells, but a majority of circulating VWF is derived from the endothelium [102]. As described above, VWF is involved both in platelet adhesion and aggregation, and in the coagulation cascade. Upon vascular wall injury, VWF is released and attaches to endothelial cells and to collagen in the subendothelial matrix. In addition, it binds to platelet receptors linking the platelets together to form the platelet plug. In the coagulation cascade VWF serves as a carrier for coagulant factor VIII. In addition to its role in hemostasis, VWF is suggested to be involved in the development of atherosclerosis [103].

Increased circulating VWF concentrations have been identified as a predictor of incident coronary heart disease and stroke [84, 104, 105]. In addition, VWF concentrations have been reported to associate with the risk of recurrent events in patients with coronary disease [104]. VWF has also been studied in relation to post-stroke outcomes. VWF concentrations in the acute phase of ischemic stroke have been investigated in relation to 3 months functional outcome or death, but no associations were found after adjustment for confounding factors, including age and stroke severity [87, 89]. However, for all-cause mortality in the long-term, i.e. several years after ischemic stroke or transient ischemic attack (TIA), independent associations with acute phase VWF concentrations have been reported [93, 106]. In contrast, a large prospective study on cases with non-disabling ischemic stroke found no independent association between VWF activity in the acute phase and mortality. The same study, however, did find an independent association with recurrent stroke risk [107].

With regards to the CNS, VWF is less studied than fibrinogen. However, experimental studies have shown that deficiency of VWF reduces ischemic cerebral injury [108]. Moreover, there is experimental data showing that VWF increases cerebral inflammation and BBB damage after intracerebral hemorrhage, thereby contributing to poor outcome [109]

Tissue-type plasminogen activator (t-PA)

As described above, t-PA is a key enzyme in the endogenous breakdown of blood clots, i.e. fibrinolysis. Circulating t-PA is derived from the vascular endothelial cells and acts through the conversion of plasminogen to plasmin, thus enabling the cleavage of fibrin and clot breakdown. The main inhibitor of t-PA is PAI-1. Only a few percentage of circulating t-PA constitutes active t-

PA. Instead, most t-PA is bound to PAI-1 or other inhibitors in a complex and is enzymatically inactive. This complex can be measured as t-PA antigen, and thus, although it at first glance might seem paradoxical, high concentrations of t-PA antigen indicate a procoagulant tendency.

Several studies have demonstrated associations between increased plasma levels of t-PA antigen and incident stroke and coronary heart disease [110 84, 111]. Concentrations of t-PA antigen in the acute phase of ischemic stroke have been evaluated in relation to functional outcome within the first three months post-stroke, but do not seem to have a major impact on this outcome [87, 89, 112]. With regards to long-term outcomes, one study found that acute phase concentrations of t-PA antigen were associated with all-cause mortality during a mean follow-up of 7.4 years after ischemic stroke, but this association did not remain after adjustment for confounding factors [93]. Another study evaluated convalescent plasma concentrations of t-PA antigen in relation to recurrent stroke, but found no association during a mean follow-up of 3.9 years [113].

Apart from its role as a key player in the intravascular fibrinolytic system, t-PA has been shown to have several additional functions, also outside the bloodstream and particularly in the CNS [114]. Just like fibrinogen, within the CNS t-PA has been shown to act on several cell types and to have both beneficial and deleterious effects. In the acute phase of CNS injury, such as ischemia, t-PA can promote neurotoxicity and increase BBB permeability [114]. However, t-PA also promotes neuronal plasticity and has a functional role in processes related to learning and memory, and t-PA may thus promote recovery [78, 114].

Thrombin activatable fibrinolysis inhibitor (TAFI)

As described above, TAFI is involved in the regulation of fibrinolysis. Activated TAFI removes lysine residues from partly degraded fibrin, which restricts t-PA binding and thus further activation of plasminogen to plasmin. In this way, the rate of fibrinolysis is attenuated. TAFI is activated by trypsin-like enzymes such as thrombin, plasmin or the thrombin/thrombomodulin complex. When TAFI is activated the activation peptide (AP) is released from the catalytic domain. Available measures of TAFI includes both intact TAFI and activated TAFI and TAFI-AP. High TAFI concentrations are associated with an increased risk of ischemic stroke [115]. Moreover, increased activation during the acute phase of ischemic stroke has been associated with more severe strokes and worse functional outcomes [116]. In line with this, results from ischemic stroke animal models have shown that inhibition of TAFI leads to

decreased infarct size as well as improved functional outcome, suggesting it as a novel therapeutic target [117]. In addition, in man, there is evidence of associations between TAFI-AP concentrations and the risk of recurrent vascular events and death during the first two years after ischemic stroke [118], implying that TAFI could influence the risk of recurrent vascular events also in the more long-term setting after stroke.

Neurofilament light chain

Neurofilament light chain (NfL) is one of five subunits forming protein polymers denoted neurofilaments. Neurofilaments are intermediate filaments, with a diameter of approximately 10 nm, that are found in the cytoplasm of neurons where they are part of the neuronal cytoskeleton. Neurofilaments are thought to be of importance for radial growth and stability of axons [119]. Under normal conditions neurofilaments are stable. However, as a result of neuroaxonal damage, the neurofilament proteins are released into the extracellular space and subsequently into the cerebrospinal fluid and, at lower concentrations, into the blood. There is a known association between age and increased neurofilament concentrations, both in cerebrospinal fluid and serum [120, 121]. The mechanisms behind this association are not known in detail, but speculatively it reflects neurodegenerative processes of normal ageing. The neurofilament proteins are exclusively expressed in neurons and thereby represent markers specific for neuroaxonal injury, independent of underlying causal pathways. A biomarker with these properties that accurately reflects the extent of neuronal injury has great potential and several possible applications in various traits involving neuronal damage, for instance neurodegenerative, inflammatory and cerebrovascular diseases, such as stroke. Among the possible applications are longitudinal assessment of disease activity, monitoring of treatment responses and prediction of prognosis, both for clinical and research purposes.

The neurofilament protein with the most promise as a biomarker is NfL. For decades, studies of NfL were limited by the fact that measurements required samples of cerebrospinal fluid and thus, the invasive procedure of lumbar puncture. The reason was that available assays were not sensitive enough to detect the low NfL concentrations in blood samples. With the development of electrochemiluminescence (ECL) immunoassays, measurements of NfL in serum became possible. However, first with the most recent single-molecule array (SiMoA) technology, which is 25-fold more sensitive for quantification of NfL compared to the ECL assays, reliable measurement of the full range of

NfL concentrations present in blood samples has become possible [122, 123]. Thereby, also small variations in NfL can be detected, which broadens the applicability and facilitates longitudinal studies of NfL particularly in diseases where lumbar puncture is not part of the clinical routine, such as stroke. Importantly, several studies have shown that the cerebrospinal and serum concentrations of NfL are highly correlated which allows conclusions about ongoing neuroaxonal injury to be drawn from serum measurements [122, 124]. This recent development of highly sensitive determination of serum NfL (sNfL) has been a prerequisite for the implementation of paper III of this thesis and has likewise resulted in several recent reports on sNfL in neurodegenerative disorders and traumatic brain injury [121, 125-127].

With regards to stroke specifically, there are multiple potential applications for sNfL including outcome prediction and longitudinal monitoring of injury processes during the first months as well as in the long-term. The latter is of specific interest in small vessel disease, a subtype that can have a more progressive disease course. However, so far data on sNfL in stroke are limited to a few reports. Results from these studies show increased sNfL in acute ischemic stroke compared to controls [128, 129] and TIA [130, 131], and as expected correlations between sNfL and infarct volume [128, 129]. There is also evidence that sNfL increases with the time from symptom onset to blood draw [128-130], and that sNfL remains elevated at 3 months post-stroke [128, 129]. Furthermore, recent data indicate that sNfL measured at day 7 after symptom onset, but not as early as within 24 hours, is independently associated with 3-month functional outcome [128, 131]. This illustrates that the time-point of measurement is of great importance when evaluating sNfL post-stroke. Taken together, these results are promising with respect to the clinical utility of sNfL in stroke. However, several aspects need to be further investigated, including the temporal profile post-stroke and the prognostic value, both in short- and long-term, as well as to further explore sNfL in different subgroups such as the etiologic subtypes of ischemic stroke.

Genetics

The human genome – basic concepts

The human genome consists of deoxyribonucleic acid (DNA), the structure of which was described in 1953 by James Watson and Francis Crick [132]. The genome is located in the cell nucleus where it is organized into 23 pairs of chromosomes, consisting of tightly coiled DNA. In addition, there is a much

smaller DNA molecule located in the mitochondria of human cells, however, the genome studied and referred to in this thesis is solely nuclear DNA. The DNA is formed as a double helix built up by two intertwining nucleotide chains. There are four different nucleotides or bases: adenine (A), cytosine (C), guanine (G) and thymine (T). The two nucleotide chains are held together by complementary hydrogen bonds; A to T and C to G, Figure 6. In the human genome there are approximately 3 billion such basepairs and their sequence make up the genetic code. Only around 1% of the genome constitutes the protein-coding part, i.e. genes. The total number of protein-coding genes is estimated around 20,000 [133, 134]. The traditional view of a gene is a DNA sequence consisting of exons, introns and a promotor region. The exons are the actual protein-coding parts, separated from one another by non-coding introns. According to the central dogma of molecular biology, exons and introns are transcribed into ribonucleic acid (RNA) that subsequently, after splicing off the introns, is translated into proteins. The promoter is the region that initiates transcription of the gene and has binding sites for RNA polymerase and so called transcription factors, and is thus also involved in regulating the transcription of the gene. This model serves well as a basic thinking model of the relation between DNA and proteins although it has proven simplistic. For instance, RNAs that are transcribed from DNA, but not translated into proteins (also called noncoding genes) have gained more attention in recent years. These non-coding RNAs have multiple functions often involving regulation of gene expression at different levels [135].

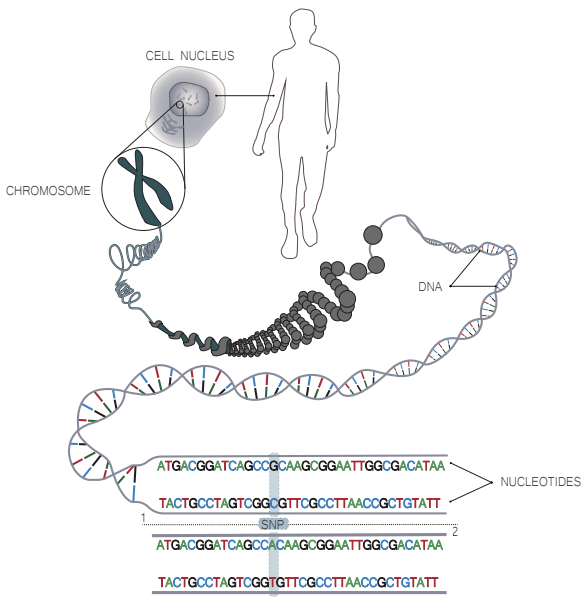
Genetic variation - Single Nucleotide Polymorphisms (SNPs) and haplotypes

All human beings are over 99% identical in their genetic makeup. The remaining part differs between individuals due to genetic variation. This variation is caused by insertions and deletions of various sizes, but the most common variation is due to so called single nucleotide polymorphisms (SNPs), meaning that one nucleotide at a specific position differs between individuals, Figure 6. On average one SNP occurs once in every 300 nucleotides. The resulting sequence variants are called alleles and a pair of alleles at a locus, i.e. a specific position in the genome, is referred to as genotype. The frequency in a population of the rarer variant, also called the minor allele frequency (MAF), can be up to nearly 50%. A commonly used definition when referring to a polymorphism is that it has a MAF of at least 1% of the population. These common variants have survived in populations for many generations, which implies they are unlikely to be highly pathogenic. However, they can be

susceptibility factors contributing to the risk of developing common diseases. The idea that susceptibility factors for common diseases are ancient common genetic variants is a theory referred to as the common disease – common variant hypothesis.

There are many millions of SNPs in the genome, but these variants are not inherited independently of each other. Loci in close proximity on a chromosome are less likely to be separated by recombination, i.e. exchange of genetic material between homologous chromosomes during meiosis, and therefore tend to be inherited together. Genetic variants within such regions are said to be linked, they show high linkage disequilibrium (LD), which implies that an allele at one position on one chromosome can predict an allele at an adjacent position on the same chromosome with high probability. Such regions are called haplotypes and are the result of recombination events occurring throughout the history of a population. Thus, they represent ancestral chromosome segments that are shared within a population and that may differ between individuals of different ethnic origins. As will be described in the section on genome-wide association studies (GWAS), genetic association studies take advantage of this structure of haplotypes.

Figure 6. Levels of organization of the human genome. At specific positions in the genome, the sequence differs between individuals by one base pair exchange, so called single nucleotide polymorphisms (SNPs).



Monogenic versus complex inheritance

Monogenic inheritance is caused by variation at a single locus in the genome and follows the Mendelian laws for autosomal dominant, autosomal recessive or X-linked inheritance. In contrast, complex inheritance refers to variation at many loci that each, and often together with environmental factors, make a small contribution to a trait or disease. Simply put, the first are variants that directly cause disease and the latter are variants that modulate the susceptibility for a disease. However, these are conceptual models and most traits are actually somewhere on a scale between the two. With regards to stroke, there are rare monogenic forms, but the majority is multifactorial, i.e. the result of complex inheritance as well as environmental factors.

Genome-Wide Association Studies (GWAS)

For at least a decade, GWAS has been the dominant tool for studying the genetics of complex diseases. In a GWAS, SNPs spread throughout the whole genome are analyzed. Thanks to the structure of haplotype blocks described above, it is not necessary to genotype every SNP. Selected SNPs serve as markers for different regions and the remaining SNPs can subsequently be determined or imputed, i.e inferred from the closest surrounding SNPs. Typically, the frequencies of variants are then compared between subjects with and without a disease. If a variant is statistically more frequent in the disease group it is said to be associated with the disease. The frequency of variants can also be statistically analyzed for association to a certain outcome, as was done in this thesis, Figure 7. If an association is found this does not mean that the detected variant is causal itself. Instead, the variant should be considered a marker for a region harbouring the actual causal gene or variant. The rationale for this is again the existence of haplotype blocks as explained above.

In contrast to the previously more common, hypothesis-driven candidate gene studies, the GWAS approach is hypothesis-free. One of the advantages with this approach is that there is a possibility to identify variants unravelling novel pathways for pathophysiological mechanisms. Thereby, this approach also has the potential to provide completely new targets for interventions and drug development. The down-side is that it results in a very large number of statistical tests and subsequently risk of falsely positive results. To avoid this, a very stringent significance level must be used, typically $P < 5\times10^{-8}$, which in turn implies a need for very large sample sizes. For this to be achieved international multi-center collaborations are often required.

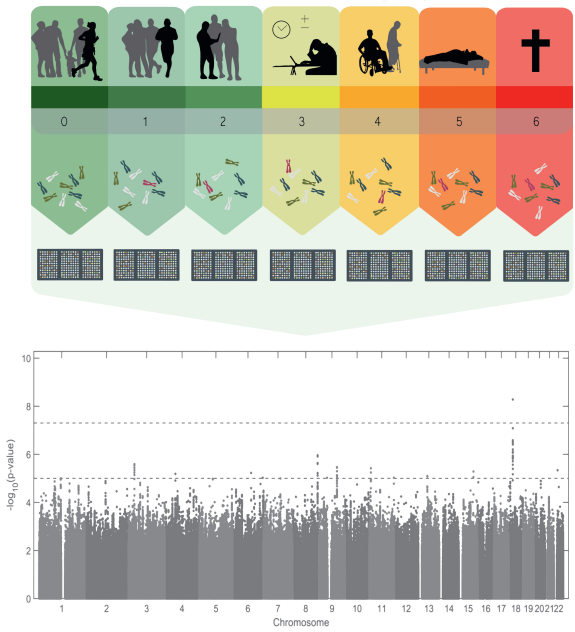


Figure 7. In a genome wide association study (GWAS) genetic variants (single nucleotide polymorphisms, SNPs) spread throughout the genomes of the participants are analyzed for associations to a disease or an outcome. In this example, the outcome measure is the modified Rankin Scale, a 7-grade scale of functional outcome after stroke. The results are presented in a Manhattan plot, in which each dot represents an SNP. The x-axis shows the genomic location and the y-axis shows the association level.

As explained above, in a GWAS the actual causal variants can be the detected ones, or some other variants within the same region. Either way, if the causal variant is in the coding part of a gene it can, for instance, introduce an amino acid change leading to alterations in structure, folding or stability of the protein. However, many causal variants are not within coding regions of genes. Instead, they affect gene expression in different ways [136]. Genetic variants that explain variation in gene expression levels (i.e. messenger RNA, mRNA, abundance) are called expression quantitative trait loci (eQTLs). Such variants may be located in regulatory regions, such as promoters and enhancers, or affect mRNA stability. eQTLs can be identified by simultaneous assessments of genetic variation and gene expression. Reported eQTLs are available in public databases that can be used when interpreting results from GWAS.

Genetics of stroke outcomes

As for many complex diseases, there is evidence for a genetic contribution to stroke susceptibility. This has been demonstrated in twin and family history studies as well as in calculation of heritability from so called genome-wide complex trait analysis [137-140]. In line with this, several common variants contributing to stroke risk have been reported [141]. The genetic contribution to stroke outcomes has been much less studied. As mentioned in previous sections, clinical variables do not fully explain the inter-individual variability in stroke outcomes and cannot provide detailed prognostic information in the individual case [142]. Most likely, environmental as well as genetic factors play a role. In support of this, there are studies in both humans and animals that suggest a genetic influence on brain function and recovery after brain injury [143-145]. In addition, genetic variants identified through individual candidate gene studies have been reported to associate with functional outcome after stroke, e.g. within the brain-derived neurotrophic factor and cyclooxygenase-2 genes [146-148], although results from these candidate gene studies have been inconsistent and need replication [149]. To date, there is only one published GWAS investigating SNPs in relation to stroke outcome. This recent study included a selected ischemic stroke population, i.e. only moderate and severe cortical infarcts of the anterior circulation, and demonstrated a genome-wide significant association for intronic variants in the *PATJ* gene and functional outcome at 3 months post-stroke [150]. Another recent study used a different approach to study common genetic variation in relation to ischemic stroke outcome. By analyzing insertions and/or deletions, i.e. copy number variations (CNVs), in relation to 3-month functional outcome they found that the risk of unfavourable outcome increased with the number of protein-coding genes affected by CNVs [151].

AIM OF THE THESIS

The overall aim of this thesis was to identify novel biomarkers for ischemic stroke outcomes with primary focus on the long-term.

Specific aims

Paper I

To test the hypothesis that circulating concentrations of the hemostatic biomarkers fibrinogen, von Willebrand factor (VWF), tissue-type plasminogen activator (t-PA), and/or thrombin activatable fibrinolysis inhibitor (TAFI) predict long-term risk of recurrent vascular events after ischemic stroke.

Paper II

To test the hypothesis that circulating concentrations of the hemostatic biomarkers fibrinogen, VWF, and/or t-PA predict long-term cognitive outcome after ischemic stroke.

Paper III

To to explore serum Neurofilament light chain (sNfL) in ischemic stroke in a longitudinal manner and test the hypotheses that circulating concentrations of NfL measured in the acute phase and/or three months after ischemic stroke predict post-stroke functional and neurological outcomes, and that the concentrations and/or the dynamics of sNfL differ between etiologic stroke subtypes.

Paper IV

To, through a genome-wide association study (GWAS) approach, identify genetic variants associated with post-stroke functional outcome, measured by the modified Rankin Scale (mRS).

SUBJECTS AND METHODS

The Sahlgrenska Academy Study on Ischemic Stroke (SAHLSIS)

Papers I-III of this thesis are based on, and paper IV includes, participants from the Sahlgrenska Academy Study on Ischemic Stroke (SAHLSIS). This study was initiated in 1998 under the leadership of professors Christina Jern and Christian Blomstrand [29]. The main purpose of SAHLSIS is to investigate genetic factors involved in ischemic stroke, with primary focus on hemostasis, in patients aged 18 to 69 years. The reason for choosing an upper age limit is several-fold. Firstly, atherosclerosis and arterial fibrillation are common causes of ischemic stroke at higher ages, whereas the etiology of ischemic stroke is more heterogeneous, and a much larger proportion remains unexplained, at younger ages. Secondly, there are studies showing that the influence of genetic factors is more pronounced in younger stroke sufferers [138]. It has also been speculated that prothrombotic mechanisms have a greater impact at younger ages. Thirdly, despite the fact that the total number of years that young stroke patients will live with the consequences of stroke far exceeds that of older stroke survivors, there are few studies on long-term outcomes for young stroke sufferers.

Cases

Between 1998 and 2003, 600 consecutive cases presenting with first-ever or recurrent acute ischemic stroke were recruited at four stroke units in Western Sweden ("Sahlgrenska Universitetsjukhuset (SU)/Sahlgrenska", "SU/Östra", "Skaraborgs Sjukhus", and "Södra Älvsborgs sjukhus"). This first phase of SAHLSIS constitutes the study sample of papers I-III. During this period 45 eligible cases were not included, 29 due to unwillingness to participate and 16 because they died before detailed informed consent had been obtained. After the first phase of SAHLSIS, cases have been continuously recruited at the stroke unit of the main center at SU/Sahlgrenska in Gothenburg. By December 2017, 1590 cases had been included in total.

All cases were included according to the criteria presented below.

Inclusion criteria:

- Acute onset of symptoms suggestive of stroke
- No hemorrhage on computer tomography (CT) or magnetic resonance imaging (MRI) of the brain
- Age 18 – 69 years

Exclusion criteria:

- Evaluation showed another etiology of the presenting symptoms than stroke
- A diagnosis of cancer at advanced stage, infectious hepatitis or human immunodeficiency virus (HIV)
- Non-Caucasian ethnicity

Controls

For each of the first 600 included cases, one control was randomly selected from participants in a population-based health survey [152] or the Swedish Population Register to match the cases with regards to age, gender and geographical residence area. They were invited to a study visit at the hospital, and individuals with a history or signs of stroke, coronary heart disease or peripheral arterial disease were excluded. In total 1,107 controls were contacted. Of those, 208 did not respond, 191 were unwilling to participate and 108 fulfilled the exclusion criteria.

Baseline data – collection and definitions

For the first 600 cases, baseline examinations were performed both in the acute phase (1-10 days after index stroke) and at a 3-month follow-up visit. Controls were examined once at inclusion. For all participants the protocol included both questionnaires and examination providing extensive data on cardiovascular risk factors, including personal and family history of vascular disease, socioeconomic factors, perceived psychological stress, physical activity at both work and leisure, smoking habits, and alcohol consumption as well as measurements of blood pressure, serum cholesterol, low-density lipoprotein (LDL), plasma glucose, anthropometric variables, and electrocardiogram (ECG) recording. Detailed description of the collection of baseline data has been previously published [29].

Hypertension was defined as pharmacological treatment for hypertension and/or systolic blood pressure ≥160 mm Hg, and/or diastolic blood pressure ≥90 mm Hg. Diabetes mellitus was defined as dietary or pharmacological treatment and/or fasting plasma glucose ≥7.0 mmol/L. Hyperlipidemia was defined as pharmacological treatment, total fasting serum cholesterol >5.0 mmol/L, and/or LDL > 3.0 mmol/L. Among cases, measurements performed at 3-month follow-up were used to define hypertension, diabetes, and hyperlipidemia. Smoking was coded as current versus never or former (smoking cessation within the last year from inclusion). The educational level was categorized as low, medium or high based on answers obtained through questionnaires. Low education corresponds to compulsory school (9 years or less), medium corresponds to secondary school, and high to post-secondary education. History of coronary artery disease was defined as having suffered myocardial infarction or having ECG changes indicating previous myocardial infarction.

Stroke severity and subtypes

Index stroke severity was scored as maximum severity within the first 7 days using the SSS for the first 600 cases [153]. Thereafter, stroke severity was assessed using the NIHSS [154]. For papers III and IV we converted the SSS to the NIHSS using an algorithm [155].

The cases in SAHLSIS have been classified into stroke subtypes using different classification systems; TOAST, Causative Classification of Stroke (CCS) and OCSP. In this thesis, TOAST and OCSP subtypes have been used. TOAST and OCSP subtyping of the first 600 cases in SAHLSIS were performed by two stroke neurologists (Katarina Jood and Christian Blomstrand). This author performed the majority of the CCS subtyping of all cases as well as TOAST and OCSP subtyping of part of the cases.

As described in the Background section, the TOAST classification [15] categorizes cases based on the underlying etiology of the ischemic stroke. In SAHLSIS, the original TOAST criteria were slightly modified, as described in detail elsewhere [156]. In brief, LVD was defined as occlusive or significant stenosis (equivalent to >50% according to the North American Symptomatic Carotid Endarterectomy Trial, NASCET, criteria) of a clinically relevant cerebral or precerebral artery, presumably due to atherosclerosis. LVD also included large plaques (>4 mm) in the aortic arch. Potential causes of cardiac embolism should be excluded. CE stroke was defined as the presence of atrial fibrillation, sick sinus syndrome, myocardial infarction in the past four weeks,

cardiac thrombus, infective endocarditis, atrial myxoma, prosthetic mitral or aortic valve, valvular vegetations, left ventricular akinetic segment, dilated cardiomyopathy, or patent foramen ovale in combination with either atrial septal aneurysm or deep venous thrombosis. CE presumes the absence of an LVD source. SVD was defined as a clinical lacunar syndrome in combination with a relevant infarct of <15 mm or normal CT/ MRI and the absence of a potential CE or LVD source. Cryptogenic stroke was defined as no identified cause despite an extensive evaluation. Undetermined stroke included cases with more than one possible etiology or when the evaluation was cursory.

The CCS is a computerized algorithm for etiologic stroke subtyping and includes the same major etiologic subtypes as TOAST, but the criteria differ slightly [157]. Some of the major differences include the requirement of intracranial vascular imaging to classify a case as SVD, the size of a lacunar infarction (<20 mm) and that patent foramen ovale can be classified as possible CE stroke also in the absence of atrial septal aneurysm or deep venous thrombosis.

The OCSP classification [16] separates strokes of different sizes and locations based on clinical presentation into TACI, PACI, POCI and LACI, as described in the Background section.

Follow-up

A large effort has been put on performing a prospective cohort study of participants in SAHLSIS. Figure 8 shows selected parts of the longitudinal design, with focus on the parts used for analyses in the papers of this thesis.

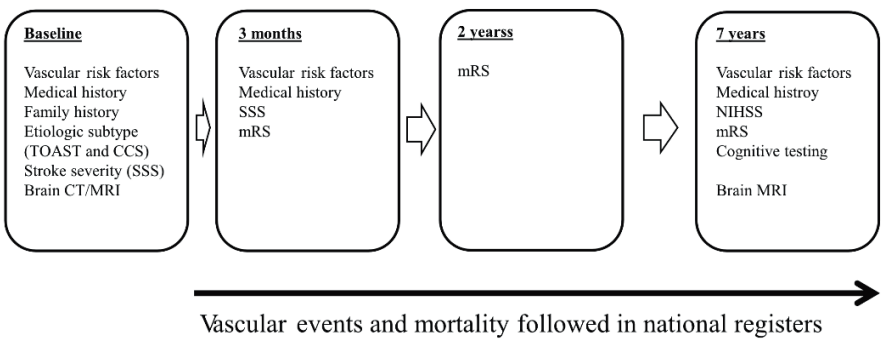
Vascular events and death during follow-up were identified using multiple overlapping methods; different national registers, interviews, and review of medical records as described [17]. In paper I, we used these data from the first 600 cases in order to investigate associations of concentrations of hemostatic proteins with risk of recurrent vascular events and vascular death. By using the unique 10-digit Swedish person identity number, we assessed participants regarding survival rates, recurrent strokes, and coronary events until the 31 of December 2012.

Information on the cause of death was obtained from the Swedish Cause of Death Register, which is based on the International Classification of Diseases 10th Revision (ICD10). We also reviewed existing medical records within 6 months prior to death, both for participants who died in hospital and for participants who died at home. Causes of death were categorized as vascular

and non-vascular. The following ICD10 codes were classified as vascular deaths: I10-I28, I42-I50, I60-I89 and F01. Death was also considered to be vascular if occurring within 30 days from a major stroke or a myocardial infarction.

To obtain data on non-fatal recurrent strokes and coronary events, we used information on hospital discharge diagnoses from the National Patient Register. Sweden has a publicly financed health care system that offers health care to all citizens at comparatively low cost. Hospitals are required to report discharge diagnoses of all patients to the registry, which contains almost complete data (99%) on dates and codes for hospital discharge diagnosis and surgical procedures. We additionally confirmed events and procedures by reviewing the corresponding medical record. When we were unable to find the medical record (15%), we registered the event if it was the main diagnosis or the main surgical procedure. The definition of stroke was the same as for index stroke, but also included primary intracerebral hemorrhage and subarachnoid hemorrhage. For stroke, we screened the ICD10 codes I60.0-I68.8. A coronary event was defined as any of the following: a myocardial infarction, an acute or elective percutaneous intervention, or a coronary by-pass grafting. For coronary event, we screened the ICD10 codes I21.0-I22.9 for myocardial infarction, FNG00-FNG96 for percutaneous intervention, and FNA00-FNF96 and FNH00-FNW98 for coronary by-pass grafting. In addition, all medical records from hospitals and primary care units collated from presentation to 3 months after index stroke, were reviewed for early stroke recurrence or coronary events. Classification of events was performed blind to biomarker results.

Figure 8. Illustration of selected parts of the longitudinal follow-up in SAHLSIS including data collected at different time-points.



SSS, Scandinavian Stroke Scale; NIHSS, National Institutes of Health Stroke Scale; TOAST, Trial of Org 10172 in Acute Treatment; CCS, the Causative Classification of Stroke system; mRS, modified Rankin Scale; CT, computer tomography; MRI, magnetic resonance imaging

In all cases, functional and neurological outcomes were assessed at three months post-stroke by a stroke neurologist. Functional outcome was assessed by the mRS and neurological outcome by the SSS. After two years mRS was scored again through a telephone interview by a study nurse trained in stroke medicine and specifically trained to score the mRS by the same stroke neurologists who performed the 3-month follow-up.

Surviving cases (n=358 out of 411) who were included at the main center, SU/Sahlgrenska were also invited to participate in a more comprehensive follow-up at 7 years after index stroke with focus on cognition. The protocol included several visits; to a research nurse, to a physician and to a visit for MRI scan of the brain. Participants who were <50 years of age at index stroke were also invited to a visit to a neuropsychologist. Home visits were offered to patients unable to visit the clinic. The research nurse took blood samples, assessed mRS and performed cognitive testing by the BNIS. The nurse had been trained in administering the BNIS by a neuropsychologist, was continuously supervised by this neuropsychologist, and was blinded to biomarker results. As described in the Background section and figure 4, the BNIS is a test in which the total score (maximum 50p) reflects the overall cognitive function and consists of a pre-screen (level of arousal 3p, basic communication 3p, and co-operation 3p) to evaluate whether the patient is capable to take part in further testing, and 7 subscales; speech and language 15p, orientation 3p, attention/concentration 3p, visual and visuospatial problem solving 8p, memory 7p, affect 4p, and awareness 1p [158].

The physician performed a physical and neurological examination including NIHSS. The neuropsychologist investigated the young stroke patients using a set of tests that assess different cognitive domains with focus on psychomotor speed, attention, executive functioning, visuospatial memory, verbal memory, and working memory. These tests included Trailmaking Test A and B [55]. For Part A, individuals rapidly connect numbered circles sequentially. Part B is a more complex task that requires examinees to shift between number sequencing and alphabetic sequencing. Part A requires rapid visual scanning and processing speed, Part B further requires executive functions such as working memory and attentional shifting.

In paper II, we used data from the 7-year follow-up visits to investigate associations between circulating hemostatic proteins and cognitive function (BNIS and Trailmaking test A and B) after stroke. In paper III, we used the 3-month mRS/SSS data, 2-year mRS data and 7-year mRS and NIHSS data to assess associations between circulating NfL concentrations and functional and

neurological outcomes after stroke. In paper IV, we used the 3-month mRS data in order to search for associations to genetic variants.

Blood sampling and protein measurements

For the first 600 patients and 600 controls in SAHLSIS, standardized blood sampling was performed at 8:30 AM - 10:30 AM after overnight fasting. Plasma and serum was isolated within 2 h by centrifugation 2000 x *g* at 4 °C for 20 min. All samples were aliquoted and stored at -80 °C pending analysis. In patients, plasma and serum were biobanked both in the acute and in the convalescent phase (3 months after index stroke). Plasma and serum have also been collected at the 7-year follow-up in SAHLSIS described above.

The plasma concentrations of t-PA antigen and VWF antigen were measured by enzyme-linked immunosorbent assays (ELISAs) in our lab as described [159, 160]. The plasma concentrations of the TAFI-AP, were measured in the Declerck lab using an ELISA that has been described in detail [161]. Plasma fibrinogen concentrations were measured with an automated clot rate assay, as described [162].

Serum concentrations of NfL were measured using a homebrew assay on the SiMoA platform (Quanterix, Lexington, MA) at the Department of Neurochemistry at the Sahlgrenska University hospital. This NfL assay is a magnetic bead-based digital ELISA that allows detection of proteins at subfemtomolar concentrations [163], and it has previously been described in detail[125]. The intra-assay coefficient of variation was 6.4% for quality control (QC) samples with concentrations of 12.4 pg/mL and 110 pg/mL (no difference between the low and high QC sample in this study).

All analyses were performed by laboratory technicians who were blind to clinical information, and all samples were analyzed with the same batch of reagents.

The Genetics of Ischemic Stroke Functional Outcome (GISCOME) network

The Genetics of Ischaemic Stroke Functional Outcome (GISCOME) network was initiated by my main supervisor Christina Jern and the co-PIs professors Arne Lindgren and Jane Maguire within the framework of the International Stroke Genetics Consortium (ISGC, www.strokegenetics.org) with the aim to identify genetic variants with influence on outcome after ischemic stroke [164].

Study population

6,165 ischemic stroke patients from 12 studies in Europe, USA and Australia were collected to be included in paper IV, which is the first GWAS from GISCOME. SAHLSIS contributed with 1,158 of these cases. The GISCOME study protocol has been published and includes detailed description of the study population and analysis plan [164]. After the publication of the GISCOME study protocol, two additional sets of data from SAHLSIS and the Malmö Diet and Cancer study have been added. In brief, the study population consists of ischemic stroke cases of mainly European ancestry ≥18 years of age. The criteria for inclusion in the GISCOME study were available mRS data 60-190 days post-stroke and GWAS genotype data. Phenotypic and genotypic data for the other participating sites in GISCOME were transferred to our site.

Outcome

The mRS as close as possible to 90 days was selected to assess functional outcome. The majority of the included studies (≈80%) assessed mRS at 3 months ± 2 weeks and in most studies this was done by face-to-face interviews. In three studies (Lund Stroke Register, Malmö Diet and Cancer study, and parts of later phases of SAHLSIS), data from the Swedish quality register for stroke (Riksstroke) was used to assess mRS by a validated translation algorithm [164, 165]. This approach prevented a differentiation between the mRS grades 0, 1 and 2. We analyzed mRS as two dichotomous outcomes (mRS 0-2 versus 3-6 and mRS 0-1 versus 2-6) and also as the full ordinal scale variable.

Genotyping, imputation and quality control

For SAHLSIS, around 20% of the participants were genotyped at the Broad Institute at Harvard Medical School with the Illumina Omni Express 750K array plus exome content; 60% at the Center for Inherited Disease Research, John Hopkins University with the Illumina 5M array with exome content; and 20% at deCoDE genetics, Iceland, in collaboration with Sólveig Grétarsdóttir with the Illumina HumanOminExpress-24v1-1_A array. These arrays include SNPs spread over the genome as well as additional exome variants.

For the other sites in GISCOME, genotyping was performed as described [164]. Genotyping at different occasions or with different genotyping arrays within the same studies resulted in a total of 17 genotyping data sets for the analyses in paper IV. The three SAHLSIS data sets were imputed separately and then merged into one data set. This set was subsequently split into one part with mRS assessed face-to-face and another part with mRS assessed by data from the Swedish quality register for stroke (Riksstroke). This division allowed the two parts with different mRS data to be analyzed separately in the ordinal analysis.

Several of the cohorts had already undergone quality control before inclusion in GISCOME. Complementary quality control steps performed for all cohorts at entry into the study were, 1) exclusion of all variants and individuals with >10% missing data rate, 2) exclusion of variants significantly ($P < 10^{-7}$) out of Hardy-Weinberg Equilibrium, and 3) exclusion of individuals with an outlying heterozygozity rate (>3 SDs from mean). Individuals were also checked for discordant sex information. Further, five principal components were calculated for each cohort.

Imputation to the 1000 Genomes Phase 3 reference panel was performed for all cohorts individually. After imputation, duplicate markers, non-biallelic markers, indels and variants with low imputation accuracy were removed. As a post-imputation quality control, the imputed allele frequencies in each cohort were plotted against those from 1000 Genomes and visually inspected.

Investigation of eQTLs

We explored associations of the markers with P values $<1\times10^{-5}$ and proxy SNPs ($r^2 > 0.8$ in 1000G Phase 1 EUR population) with eQTLs in publicly available datasets encompassing numerous tissues: GTEx V6 [166] , GRASP2 [167, 168], HGVD [169], and BIOS [170].

Statistical analyses

In paper I-III, baseline characteristics were compared using the χ^2-test for proportions and Student's *t*-test or Mann-Whitney *U*-test for numeric variables. Correlations were assessed by Pearson or Spearman correlation coefficients. For the statistical analyses all biomarker concentrations were log transformed to reduce skewness. Data were analyzed using SPSS 20.0 and R 3.3.1. Two-tailed $P < 0.05$ was considered significant.

In paper I, uni- and multivariable Cox regression models were used to assess associations between biomarkers and recurrent stroke, coronary events, and vascular death, whichever occurred first. Patients were censored at time of death if death was due to a cause other than the predefined endpoints. Biomarker concentrations were standardized such that reported hazard ratios (HRs) represent an increase by one standard deviation (SD). We also assessed the combined effect of the biomarkers that were significantly associated with any of the outcome measures. To this end, the mean value of the standardized biomarker levels was calculated for each individual to make a combined biomarker score that was used in the Cox regression. In this way, the contribution from each biomarker will be equal. The proportional hazards assumption was tested by the correlation of the Schoenfeld residuals vs. log time [171]. The accuracy of the prediction was estimated by the area under the ROC curve (AUC)/c-statistics [172].

In paper II, multivariable linear regression was used to assess the associations between biomarker concentrations and cognitive outcomes. To be able to compare the relative importance between the biomarkers standardized betas (β_{std}) were used. The standardized betas refer to how many standard deviations the dependent variable (BNIS) is estimated to change per standard deviation increase in the predictor variable.

In paper III, linear regression was used to analyze associations between sNfL and stroke severity and neurological outcome (NIHSS at 3 months and 7 years). Associations between sNfL and functional outcome (dichotomized mRS score, 3-6 vs 0-2, at 3 months, 2 years and 7 years) were analyzed by logistic regression. To assess the diagnostic accuracy of sNfL for discriminating good and poor outcomes we calculated the AUC/c-statistics. sNfL in cases and controls were compared by Student's paired *t*-test and in multivariable analyses by mixed models using the paired sNfL as a repeated measure. Differences in sNfL according to OCSP subtypes were analyzed with ANOVA (post hoc Tukey). Differences in sNfL between the etiological subtypes were analyzed with ANCOVA adjusting for age. The decline in sNfL

from 3 months to 7 years was compared between subtypes by ANOVA using age-adjusted concentrations for the measurement at 7 years.

The statistical analyses of paper IV were performed by Erik Lorentzen at the Bioinformatics Core Facility, University of Gothenburg. Multivariable models were used for analyses of each outcome variable, under an additive genetic model. In a primary model results were adjusted for age, sex, ancestry (up to the five first principal components), and baseline stroke severity as assessed by NIHSS at 0-10 days post stroke onset, with preference to as close to day 0-1 as possible [164]. Additionally, models without adjustment for baseline stroke severity were performed for each outcome for comparison. All dichotomized analyses were performed with logistic regression. The full mRS was analyzed with ordinal logistic regression under a cumulative logit model. Inverse variance weighted fixed effects meta-analysis was performed. Variants with minor allele frequency <0.01 or that were missing in >50% of cohorts were excluded. After filtering, approximately 8.5 million SNPs were included in each of the final meta-analyses. Heterogeneity of SNP effects between studies was tested in the meta-analysis. Markers with P values $< 5\times10^{-8}$ were considered significant for association with outcome, while markers with P values $< 1\times10^{-5}$ were considered suggestive. To facilitate comparison of the results from the dichotomized and ordinal analyses, we presented all effect sizes as odds ratios (ORs) per copy of the minor allele; an OR > 1 indicates a higher mRS (worse outcome) per copy of the minor allele and an OR < 1 indicates a lower mRS. Gene-based tests were performed for each meta-analysis using VEGAS2 with LD structure based on the European (EUR) population 1000 Genomes reference panel [173]. The number of genes included was approximately 23,000 which corresponds to a Bonferroni corrected significance threshold of $P < 2.2\times10^{-6}$.

For more details regarding the statistical methods, please refer to the respective paper.

Ethical approvals

Written informed consent was obtained from all participants prior to enrolment. For participants who were unable to communicate, consent was obtained from their next-of-kin. The SAHLSIS was approved by the Ethics Committee of the University of Gothenburg. For the GISCOME GWAS local ethics committees approved the individual studies.

RESULTS

Characteristics of the SAHLSIS cohort

Table 1 shows baseline characteristics of the SAHLSIS cohort, including cardiovascular risk factors in controls, in overall ischemic stroke and in the 4 main etiological subtypes, as well as stroke severity and 3-month functional outcome in overall ischemic stroke and the 4 subtypes.

Hemostatic biomarkers are associated with long-term recurrent vascular events after ischemic stroke (Paper I)

In paper I, we investigated concentrations of the hemostatic proteins t-PA antigen, VWF antigen, fibrinogen and TAFI-AP measured at the 3-month follow-up for associations with recurrent vascular events and death after ischemic stroke. Of the 600 patients included, 7 died before the 3-month follow-up. To avoid influence of an acute vascular event on convalescent biomarker measures, patients suffering from a recurrent vascular event before follow-up were excluded (n=29). Of the remaining 564 individuals, 16 declined participation at 3-month follow-up, leaving a study sample of 548 individuals.

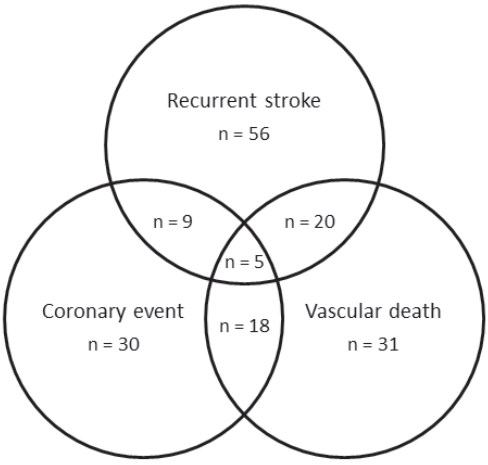
We followed cases for a median of 10.8 years (interquartile range (IQR) 9.6-11.9), representing a total of 5637 person-years. Five patients emigrated, two after five years, two after seven years, and one after nine years of follow-up. During follow-up, 115 patients died, among whom 74 were classified as vascular death. In total, 90 patients had a recurrent stroke and 62 a coronary event. In 71 patients, we registered ≥2 different events. The relationships between the outcome measures are schematically depicted in Figure 9.

Table 1. Baseline characteristics of the Sahlgrenska Academy Study on Ischemic Stroke (SAHLSIS) cohort including cardiovascular risk factors, stroke severity and 3-month functional outcome.

	Controls	Cases	Large vessel disease	Small vessel disease	Cardioembolic stroke	Cryptogenic stroke
	(n=600)	(n=600)	(n=73)	(n=124)	(n=98)	(n=162)
Age, median (IQR)	59 (52-65)	59 (52-65)	60 (57-65)	60 (54-64)	61 (54-66)	56 (48-62)***
Male sex, n (%)	385 (64)	385 (64)	54 (74)	77 (62)	66 (67)	95 (59)
Hypertension, n (%)	224 (37)	354 (59)***	44 (60)***	89 (72)***	50 (51)**	87 (54)***
Diabetes mellitus, n (%)	33 (6)	114 (19)***	25 (34)***	26 (21)***	19 (19)***	23 (14)***
Hyperlipidemia, n (%)	403 (67)	413 (76)**	53 (82)*	77 (71)	73 (82)**	107 (71)
Current smoker, n (%)	109 (18)	233 (39)***	39 (53)***	54 (44)***	34 (35)***	60 (37)***
Previous history of stroke, n (%)	0 (0)	114 (19)***	21 (29)***	25 (20)***	22 (22)***	18 (11)***
NIHSS score baseline, median (IQR)	NA	3.1 (1.6-7.2)	3.3 (1.2-11.5)	2.5 (1.6-4.1)	3.8 (1.4-10.6)	2.5 (1.2-6.4)
mRS score 3 months, median (IQR)	NA	2 (1-2)	2 (1-3)	1 (1-2)	2 (1-3)	2 (1-2)

*Data are shown as median and interquartile range (IQR) or number (n) and percentage for controls and for the whole group of ischemic stroke cases as well as for the 4 major ischemic stroke subtypes. Differences compared to the control group were examined using the χ^2-test for proportions and Mann-Whitney U-test for continuous variables. NIHSS, National Institutes of Health Stroke Scale; mRS, modified Rankin Scale. *P < 0.05, **P < 0.01, ***P < 0.001.*

Figure 9. Venn diagram illustrating the relationships between outcome measures. Numbers represent individuals.



In univariable Cox regression analyses, we found statistically significant associations between plasma concentrations of t-PA, VWF, and fibrinogen and the long-term risk of a coronary event and vascular death. We found no significant associations between TAFI-AP concentrations and any of the outcome measures. When combining the values of t-PA, VWF, and fibrinogen we found significant associations with coronary events and vascular death, with substantially higher hazard ratios for all outcomes compared to the individual biomarkers. We saw no statistically significant associations between recurrent stroke and any of the biomarkers. After adjustment for age, diabetes mellitus, and TOAST subtype, the hazard ratios were generally attenuated but remained significant for the association between t-PA antigen and vascular death, as well as the associations between the combined biomarkers and coronary events, Table 2.

For the biomarkers t-PA, VWF, and fibrinogen, we performed further univariable analyses stratified for age, sex, diabetes and atherosclerosis, respectively. These analyses revealed a statistically significant difference ($P < 0.01$) between diabetic and non-diabetic patients with respect to the association between VWF levels and coronary events (HR per 1 SD increase in plasma level 2.26 (95% CI 1.56-3.28, $P < 0.0001$) for non-diabetic patients and 0.88 (0.62-1.26, $P = 0.49$) for patients with diabetes. After adjustment for age and TOAST subtype, the HR for non-diabetic patients was 2.23 (1.45-3.43, $P < 0.001$).

Table 2. Results from Cox regression analyses showing Hazard Ratios (HR) and 95% Confidence Intervals (CI) for the outcomes coronary event and vascular death.

	Coronary event		Vascular death	
	HR (95%CI)	***P* value**	**HR (95%CI)**	***P* value**
t-PA antigen	1.39 (1.09-1.77)	0.01	1.46 (1.17-1.81)	<0.001
VWF antigen	1.56 (1.19-2.06)	<0.01	1.43 (1.12-1.84)	<0.01
Fibrinogen	1.44 (1.12-1.84)	<0.01	1.47 (1.17-1.85)	<0.001
TAFI-AP	1.21 (0.92-1.58)	0.17	1.23 (0.96-1.57)	0.10
Combined	1.74 (1.33-2.27)	<0.001	1.66 (1.31-2.12)	<0.001
Biomarkers with significant associations in multivariable models				
t-PA antigen*			1.27 (1.00-1.61)	0.047
Combined*	1.35 (1.02-1.80)	0.04		

*Cox regression was used for calculation of HRs. *adjusted for age, diabetes mellitus and TOAST subtype. Combined corresponds to the mean of the standardized values of t-PA antigen, VWF antigen and fibrinogen for each individual. t-PA, tissue-type plasminogen activator; VWF, von Willebrand factor; TAFI-AP, Thrombin Activatable Fibrinolysis Inhibitor Activation Peptide.*

We also evaluated the diagnostic accuracy for predicting coronary events and vascular death by assessing AUC. For a model with age and diabetes the AUC for predicting coronary events was 0.69, which increased to 0.73 when adding the combined biomarker score and to 0.74 when additionally adding TOAST subtype. For vascular death we found the corresponding figures 0.69 for age and diabetes, 0.70 when adding the combined biomarker score, and 0.76 when adding TOAST subtype.

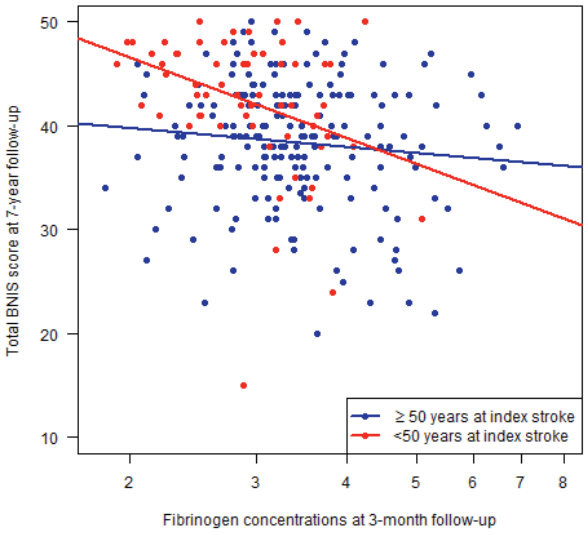
Fibrinogen concentrations predict long-term cognitive outcome in young ischemic stroke patients (Paper II)

In paper II, we investigated concentrations of the hemostatic proteins t-PA antigen, VWF antigen and fibrinogen measured at the 3-month follow-up for associations with long-term cognitive function after stroke. At 7-year follow-up, 332 ischemic stroke cases were eligible for cognitive testing and 268 of those were included in the analyses of this study. Among eligible study participants, the 268 included cases did not differ significantly with regards to age, sex, stroke severity or plasma concentrations of fibrinogen or t-PA compared to those who did not participate. However, the investigated group had significantly lower 3-month VWF concentrations compared to the group that did not participate ($P < 0.01$). Median time from index stroke to follow-up was 7.4 years (IQR 7.3-7.5). Forty-five patients experienced a recurrent stroke during the follow-up period from the 3-month follow-up to the 7-year visit. Median time from this recurrent stroke to cognitive testing was 5.0 years (IQR 3.2-7.2).

We assessed cognitive function by the BNIS. Participants with stroke <50 years of age were also examined by the Trailmaking Test A (n=44) and B (n=41). The median total BNIS score was 40 (IQR 35-43) for the whole sample, and 43 (IQR 39-47) for the group of participants with stroke <50 years of age.

Associations between biomarker concentrations and cognitive scales were assessed in the whole group and in participants with stroke <50 years of age, specifically. In the whole sample the hemostatic biomarkers fibrinogen, VWF and t-PA, were all correlated to total BNIS score, but these associations did not withstand adjustment for confounding factors (age, dichotomized age <65 years, hypertension, diabetes mellitus, educational level, SSS at baseline and time since last stroke). In the subgroup of participants who were <50 years of age at index stroke, a similar pattern of correlations between biomarker concentrations and the total BNIS score was observed as in the whole group. However, in this group the association for fibrinogen was independent of the included covariates (β_{std} = -0.27, 95% CI -0.47 to -0.07)) and significantly stronger than in the participants with stroke ≥50 years of age, Figure 10. With respect to the different BNIS subscales, high plasma concentrations of fibrinogen were associated to a low score for all subscales, and this association did not significantly differ between subscales.

Figure 10. Scatterplot with regression lines for plasma concentrations of fibrinogen at baseline in relation to total BNIS score at 7-year follow-up.



BNIS, Barrow Neurological Institute Screen for Higher Cerebral Functions.

Of the 67 participants <50 years of age that underwent cognitive testing with BNIS, 44 participants completed the Trailmaking Test A, and 41 completed Trailmaking Test B. The mean total time for Trailmaking Test A was 34 (SD 12) seconds, and for Trailmaking Test B 75 (SD 30) seconds. High fibrinogen concentrations were significantly correlated to worse performance (measured as total time) on Trailmaking Test A, r = 0.37 (95% CI, 0.07 – 0.61). In the linear regression model this association was retained, β_{std} = 0.31 (95% CI, 0.03 – 0.58). No significant association was detected for fibrinogen and Trailmaking Test B, β_{std} = 0.09 (95% CI, -0.17 – 0.35).

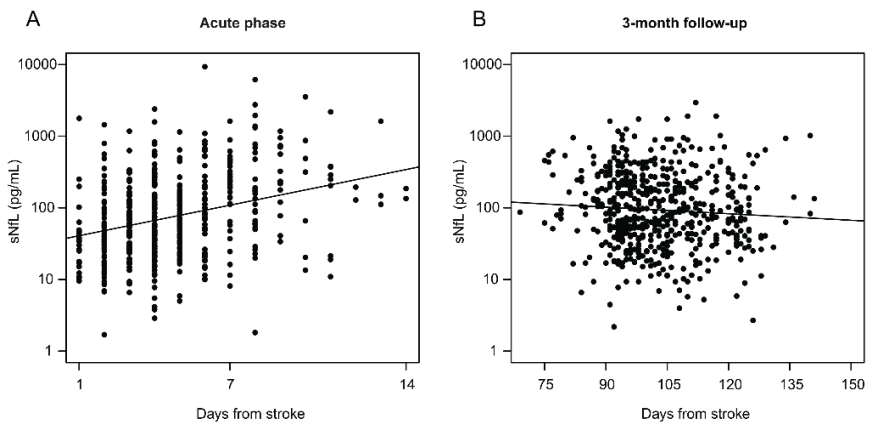
Circulating neurofilament light in ischemic stroke: Temporal profile and outcome prediction (Paper III)

In paper III, we investigated concentrations of circulating NfL as predictors of functional (mRS) and neurological (NIHSS) outcomes at 3 months, mRS at 2 years and mRS and NIHSS at 7 years post-stroke. We further assessed sNfL in relation to initial stroke severity, etiologic subtypes, and in cases compared to controls. In five patients, serum was unavailable from all time points, leading to a study cohort of 595 patients with serum sample available from at least one time point and their corresponding matched controls.

Serum samples were available from 489 cases in the acute phase (median 4 days post-stroke), 546 at 3 months post-stroke, and 595 controls. 221 cases had serum samples from the 7-year follow-up. Median sNfL concentrations were 60.2 pg/mL (IQR 28.3–190) in the acute phase, 90.6 (41.8–230) at 3 months, 18.1 (11.5–35.5) at 7 years, and 14.2 (9.0–21.0) in controls.

Acute phase sNfL increased with time to the blood draw (r=0.32), Figure 11, panel A. This association remained after adjustment for age and stroke severity in a linear regression model ($P < 0.001$). Contrary, at 3 months, sNfL decreased with the time to blood sampling (r=-0.08), Figure 11, panel B.

Figure 11. Correlations between sNfL and day of blood sampling during the acute phase (r=0.32, panel A) and 3 months post-stroke (r=-0.08, Panel B).



sNfL, serum neurofilament light chain

Three-month sNfL showed strongest correlation to baseline NIHSS (r=0.56) and was associated to both outcomes at all time-points, after adjusting for age, previous stroke, stroke severity, and day of blood draw ($P < 0.01$ throughout). Uni- and multivariable ORs for poor functional outcome, i.e. mRS>2, at 2 years are displayed in Figure 12. There were also significant associations between acute phase sNfL and outcomes, but they were generally weaker than those for 3-month sNfL. We next evaluated the diagnostic accuracy of sNfL for predicting functional outcome (mRS 0-2 vs 3-6) at 2 years post-stroke. Acute phase and 3-month sNfL yielded an AUC of 0.68 (95% CI, 0.63-0.74) and 0.79 (95 % CI, 0.74-0.84), respectively. For comparison, the corresponding AUC for stroke severity (baseline NIHSS) was 0.84 (95% CI, 0.80-0.88). Including age, sex, hypertension, diabetes mellitus, and smoking and stroke severity in the same model yielded an AUC of 0.86 (95% CI, 0.82-0.89). When adding 3-month sNfL to the latter model, the AUC increased slightly to 0.87 (95% CI, 0.84-0.90).

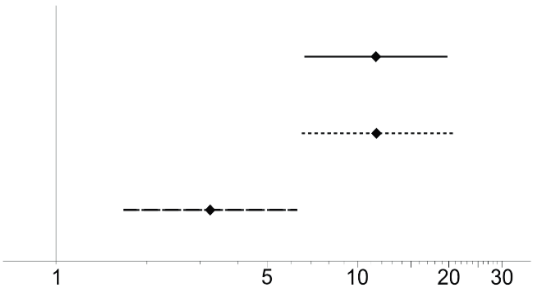


Figure 12. Forest plot showing odds ratios and 95% confidence intervals for poor 2-year functional outcome (mRS > 2) per one unit increase in 3-month log sNfL. Solid line represents univariable analysis, dotted line multivariable analysis with age, previous history of stroke and day for blood sampling at 3-month follow-up as covariates, and dashed line multivariable analysis additionally including initial stroke severity (baseline NIHSS).

Both acute phase and 3-month sNfL in cases were significantly higher than in controls. After adjustment for cardiovascular risk factors, the difference between cases and controls was retained for both time points (the ratio of sNfL in cases vs controls was 5.0 (95% CI 4.4–5.9) for the acute phase and 6.5 (6.0–7.2) for 3-month concentrations). After 7 years sNfL had declined, but was not significantly higher in cases than in controls when adjusting for age and cardiovascular risk factors (ratio of sNfL in cases vs controls, 1.1, 95% CI 1.0–1.3).

Finally, we observed that the dynamics of sNfL differ by etiologic subtype of ischemic stroke, with LVD and CE stroke showing highest concentrations in the acute phase and at 3 months, whilst there was a trend towards higher concentrations in SVD at 7 years post-stroke.

Genome-wide association meta-analysis of functional outcome after ischemic stroke (Paper IV)

In paper IV, we performed a GWAS including 6,165 patients with ischemic stroke from studies in Europe, the United States and Australia, Table 3.

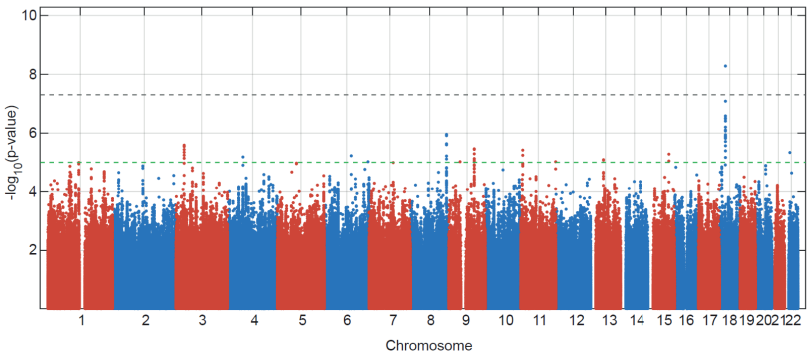
Table 3. Characteristics of the GISCOME study population by study site.

	Sex	Age	NIHSS baseline	mRS 3 months						
Study location	**Male/ Female**	**Mean (SD)**	**Median (IQR)**	**0**	**1**	**2**	**3**	**4**	**5**	**6**
Australia	270/232	71.5 (13.1)	4 (2-8)	24	190	132	64	39	16	37
Barcelona a1	68/62	71.2 (10.5)	15 (9-18)	22	32	17	20	24	15	0
Barcelona a2	78/27	68.2 (11.4)	3 (1.75-8)	30	25	16	12	12	4	6
Barcelona b	483/423	75.1 (11.2)	5 (3-12)	153	151	149	125	138	38	152
Boston 1	28/21	69.6 (16.5)	4 (2-10.5)	10	11	4	5	5	0	14
Boston 2	48/28	66.9 (13.1)	3 (1-7)	22	24	10	10	5	0	5
Boston 3	143/82	64.9 (15.6)	3 (1-8)	55	70	38	23	16	2	21
Cincinnati	194/160	69.5 (13.1)	4 (2-8)	26	71	73	71	67	18	28
Gothenburg 1	487/273	55.8 (10.7)	2 (1-7)	97	174	330	94	52	2	11
Gothenburg 2	258/140	56.1 (10.7)	2 (0-6)	-	317	-	49	21	8	3
Helsinki	212/130	63.9 (12.6)	5 (2-10)	63	97	77	47	43	3	12
Leuven	269/189	67.5 (14.5)	4 (2-8)	114	130	98	53	39	6	18
Lund 1	256/233	74.3 (12.8)	3 (2-7.25)	-	260	-	100	43	38	48
Lund 2	263/215	71.7 (13.0)	4 (2-8)	-	259	-	73	64	31	51
Malmö	170/178	78.4 (6.1)	2 (0-7)	-	136	-	104	47	28	33
Oxford	218/235	74.6 (12.3)	2 (0-4)	67	127	73	89	35	19	43
Washington	52/40	67.2 (14.3)	8 (4-12)	23	24	9	14	9	8	5
TOTAL	3497/2668	68.7 (14.1)	4 (2-8)	706	2098	1026	953	659	236	487

NIHSS, National Institutes of Health Stroke Scale; mRS, modified Rankin Scale: SD, standard deviation; IQR, interquartile range

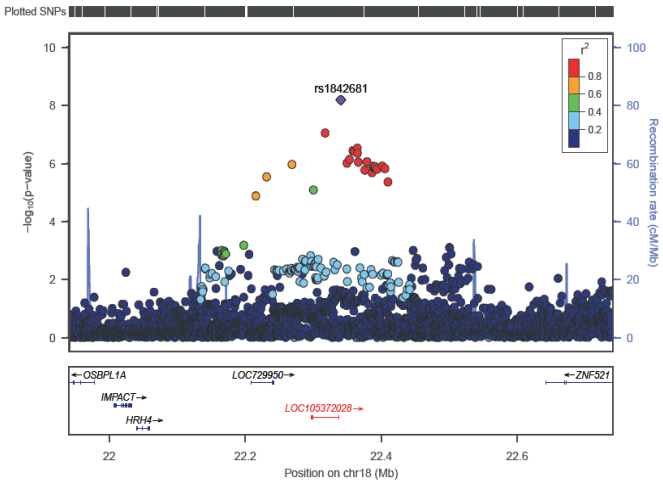
We identified one common variant on chromosome 18q11.2 (rs1842681, minor allele frequency: 0.23) that was associated at genome-wide significance with outcome defined as mRS 0-2 vs. 3-6 (OR for minor allele [A] 1.40, $P = 5.3\times10^{-9}$), Figure 13 and 14. The effect was similar with and without adjustment for stroke severity, and the association was observed in the same direction, but with a lower effect size, for ordinal mRS (OR 1.17, $P = 1.5\times10^{-4}$) and mRS 0-1 vs. 2-6 (OR 1.12, $P = 7.4\times10^{-2}$). The variant is located in an intron of the gene *LOC105372028* (long non-coding RNA (lncRNA) synonymous to *RP11-449D8.5*, [Genome Reference Consortium Human Build 38/hg38]), and is in a putative binding site of several regulatory proteins (HaploReg, 03/21/18). In HGVD this variant has a *trans*-eQTL for *KLRAQ1* (also known as *PPP1R21*, $P_{HGVD} = 1.67\times10^{-7}$). *PPP1R21* is expressed in the brain (GTEx, 06/01/18).

Figure 13. Manhattan plot of analysis for associations with dichotomized mRS at 3 months.



We measured outcome as mRS 0-2 vs. 3-6 at 3 months after ischemic stroke onset. Dotted lines show genome-wide significance (black, $P < 5\times10^{-8}$) and suggestive association level (green, $P < 10^{-5}$). Results are adjusted for age, sex, ancestry and baseline NIHSS. mRS, modified Rankin Scale; NIHSS, National Institutes of Health Stroke Scale.

Figure 14. Regional association plots for outcome at 3 months after ischemic stroke onset.



Significant locus (rs1842681) showing association with mRS 0-2 vs. 3-6, after adjustment for age, sex, principal components and baseline NIHSS. LOC105372028 (indicated in red) has been overlayed from the Genome Reference Consortium Human Build (GRCh) 38/hg38, as it was missing from GRCh37/hg19. The rs1842681 variant is intronic of the LOC105372028 gene (chromosome 18: 24725781-24766645). Position for rs1842681 in GRCh38/hg38, 18:24761199. mRS, modified Rankin Scale; NIHSS, National Institutes of Health Stroke Scale

We found thirty-three SNPs in 12 different loci (with at least 1 Mbp distance) that were suggestively ($P < 10^{-5}$) associated with mRS 0-2 vs. 3-6 (excluding the SNPs in the significant locus on chromosome 18q11.2), and 75 SNPs in 17 different loci with ordinal mRS (Figure 1-2). Of these 29 independent loci, the top SNPs of 16 loci have either significant eQTLs for and/or are located within or near (<100 kbp) genes that are expressed in the brain (GTEx, 06/01/18). Five genes that were not reported to be expressed in brain tissue were expressed in arteries or lymphocytes (GTEx 06/01/18).

Among the suggestive associations three are linked to genes with experimental evidence of influence on outcome from animal models of stroke [174-176]. Firstly, rs2236406, an intron variant in the *PTCH1* gene, was identified in the mRS 0-2 vs 3-6 analysis. Rs2236406 is a eQTL for lncRNA RP11-435O5.5 ($P_{GTEx} = 5.7\times10^{-7}$), which overlaps the *PTCH1* gene. Secondly, in the ordinal analysis, a suggestive association with mRS was found for rs13299556, an intron variant

in the *PLAA* gene. This variant, and variants in high LD with it, are reported as eQTLs for the nearby genes *TEK* and *LRRC19* (P_{GTEx}=1.6×10^{-6} for both). The variant is also predicted to alter a putative regulatory motif sequence (HaploReg, 03/21/18). Thirdly, we found suggestive association with ordinal mRS for an intron variant in *NTN4* (rs78734480). This variant was also associated with dichotomous mRS at a low *P* value, although not below the cut-off for suggestive association.

We also made an explorative effort to investigate potential associations with infarcts in subcortical and cortical locations separately. Since we did not have available information on infarct location we divided the cases into SVD (referred to as lacunar stroke) and other subtypes (referred to as non-lacunar stroke), according to subtype classifications by TOAST. A total of 992 patients had lacunar stroke, 3991 patients had non-lacunar stroke and for 1182 individuals this information was missing. The five top findings differed between these subgroups, but we did not detect any genome-wide significant associations (all $P > 6\times10^{-7}$).

DISCUSSION

General discussion

The focus of this thesis was to investigate protein and genetic biomarkers in relation to outcomes after ischemic stroke. The longitudinal design of SAHLSIS allowed us to study several different outcome measures both in the short- and long-term. The SAHLSIS cohort is well-characterized with deep phenotypic data on stroke subtypes, stroke severity and cardiovascular risk factors including social and lifestyle factors. Repeated blood sampling in cases has enabled analyses of circulating biomarkers from different time points in relation to index stroke. This thesis has a main focus on long-term outcomes with investigations of convalescent levels of proteins involved in hemostasis for associations to recurrent vascular events and vascular death, as well as cognitive function. Further, we evaluated repeated measures of a marker of neuronal damage in relation to functional and neurological outcomes. Finally, through the collaborative effort of the GISCOME network, we were able to perform the first international GWAS on outcome after overall ischemic stroke.

Hemostatic biomarkers are associated with long-term recurrent vascular events after ischemic stroke (Paper I)

As described in previous sections, thrombus formation is a key mechanistic event in cardiovascular diseases such as ischemic stroke and myocardial infarction, which constitutes the rationale for investigating proteins involved in hemostasis in relation to these diseases. Accordingly, t-PA, VWF and fibrinogen have been identified as predictors of incident myocardial infarction and stroke [84, 105, 110, 177, 178]. However, the risk factors and pathophysiological mechanisms are not necessarily the same for incident ischemic stroke and the long-term risk of recurrent vascular events after an ischemic stroke, and knowledge concerning biomarkers predictive of the long-term outcomes is yet limited. Therefore, we set out to investigate associations between circulating concentrations of t-PA antigen, VWF antigen and fibrinogen and risk of recurrent vascular events after ischemic stroke. In line with our hypothesis, our group has previously shown that levels of TAFI-AP predict death and recurrent vascular events within the first two years after ischemic stroke [118]. Therefore,

we also included TAFI-AP in our analyses on the more long-term risk of recurrent vascular events and vascular death.

In summary, we found that plasma levels of t-PA, VWF and fibrinogen were significantly associated with vascular death and coronary events. After adjustment, an independent association between t-PA and vascular death remained. The combined effect of t-PA, VWF and fibrinogen was independently associated with coronary events. In non-diabetic patients, we found an independent association for VWF levels with coronary events.

The independent associations that we found were for the outcomes vascular death and coronary events. In contrast, we found no biomarker level that was significantly associated with the risk of recurrent stroke. Although contradicting results exist [107, 179], our findings are in line with a previous study that measured convalescent plasma levels of t-PA antigen and found no association with recurrent stroke during a mean follow-up of 3.9 years [113]. Of note, in studies that did detect independent associations, the outcome of recurrent stroke was restricted to ischemic events [94 107, 179]. A similar pattern as in our study, with weaker associations for recurrent stroke compared to coronary events, was observed in a study pooling data from three different cohorts of TIA and minor stroke patients and investigating fibrinogen levels and long-term risk of recurrent vascular events [94]. A plausible explanation could be that while most coronary events are due to mechanisms involving atherosclerosis, recurrent stroke is mechanistically more heterogeneous.

We found no association between TAFI-AP and recurrent vascular events, which contrasts to the previous results from our group showing that levels of TAFI-AP predicted death and recurrent vascular events within the first two years after incident stroke. A possible explanation is that the mechanisms involved in increasing the risk of recurrent events during the first years after an ischemic stroke differ from those that are of importance in the long-term. It could also be that individuals who remain free from recurrent events during the first years are less prone to new events or are in possession of a protective factor reducing the effect of high TAFI-AP levels.

Finally, for outcome prediction, combining the effect of several biomarkers into a score can be a more powerful approach as compared to single biomarkers. In line with this reasoning, we found that the combined effect of t-PA, VWF, and fibrinogen was associated with the future risk of coronary events. Importantly however, when it comes to different outcome measures, our results suggest that different biomarkers are predictive of different types of vascular events. Furthermore, the significance of biomarker levels may differ between subgroups of patients. In our study population, levels of VWF had a significant effect regarding future risk of coronary events, but only in

non-diabetic patients. This indicates that biomarkers can add prognostic information, specifically in patients with fewer established risk factors for vascular disease. Our results suggest that for the construction of a risk score based on hemostatic biomarkers, a careful selection of the included biomarkers is essential since different biomarkers are likely to be of relevance for different outcomes as well as for different subgroups of patients.

To conclude, the results from paper I show that hemostatic proteins are associated with long-term risk of recurrent vascular events and vascular death after ischemic stroke. However, the effect sizes are small and our results, in line with previous data, show that individual hemostatic biomarkers provide little prognostic value in addition to conventional risk factors [84]. Moreover, it is important to note that our study design does not allow any conclusions on causation. The causal relationship between hemostatic biomarkers and cardiovascular disease is not clear as illustrated by a large Mendelian randomization study that suggested that elevated fibrinogen may be a consequence of rather than causal for atherosclerosis [111, 180]. Nevertheless, our results indicate that prothrombotic factors may be associated with recurrent vascular events post-stroke, and that these associations may vary depending on comorbidities and outcomes.

Fibrinogen concentrations predict long-term cognitive outcome in young ischemic stroke patients (Paper II)

Reduced cognitive function is a common impairment after stroke with long-term consequences. As described in previous sections, there is support for a role of hemostatic proteins in vascular dementia, as well as cognitive decline in individuals with atherosclerosis [81, 82, 100, 101]. To explore whether this pathway is involved also in post-stroke cognitive impairment, we analyzed concentrations of the hemostatic proteins fibrinogen, VWF antigen and t-PA at three months post-stroke for associations with long-term cognitive outcomes after ischemic stroke.

We found that fibrinogen, VWF and t-PA, were all correlated to cognitive function measured as the total BNIS score, but these associations did not withstand adjustment for confounding factors in the whole group. However, in patients <50 years, we found an independent association between fibrinogen concentrations and the total BNIS score and to performance on the Trailmaking Test A, a measure of processing speed. No such association was seen for the Trailmaking Test B, a more complex test also requiring working memory and attentional shifting, nor did we see any differences in associations to the cognitive domains covered by the BNIS subscales.

The number of previous studies on circulating biomarkers and long-term cognitive outcomes after stroke is limited. Interestingly though, a study on cases with lacunar infarctions used a proteomic discovery approach to analyze microvesicle-enriched fractions of plasma pools. They found that fibrinogen alpha-, beta-, and gamma chains as well as VWF were among the proteins that were upregulated in patients with cognitive decline at 5-year follow-up, and fibrinogen alpha was the protein that was most upregulated in this group [181].

Several potential mechanisms could explain the observed association between fibrinogen concentrations and cognitive outcome. First, a prothrombotic state and progressive cerebral micro-infarction is a plausible explanation that is supported by the fact that the correlation between hemostatic biomarkers and vascular dementia is generally stronger compared to all dementia [82, 100]. Alternatively, the main effect could be through inflammation rather than hemostasis. Fibrinogen is involved in inflammation and associations between markers of inflammation and cognitive function after ischemic stroke have previously been reported [182]. Finally, although we do not know if peripheral fibrinogen concentrations correlate to those in the CNS, as described in the Background section, there is experimental data supporting pathological mechanisms for fibrinogen in the brain.

Similar to paper I, our study design does not allow any conclusion on causation. In the case of fibrinogen and cognitive function, a previous study suggested a reverse causation such that cognitive ability in early life may predict later change in fibrinogen, possibly mediated through lifestyle [183].

To conclude, the results from paper II indicate that fibrinogen concentrations predict long-term cognitive function in the young, but not all, ischemic stroke cases. This could be due to multiple contributing factors that lead to reduced cognitive performance in older individuals; hence, the relative importance of fibrinogen would be less. Based on our findings, future studies on hemostatic biomarkers and long-term cognitive outcomes after stroke should focus on stroke in the young.

Circulating neurofilament light in ischemic stroke: Temporal profile and outcome prediction (Paper III)

There are several reasons why sNfL has potential as a predictor of stroke outcomes. In contrast to the hemostatic biomarkers in paper I and II, as well as most previously studied biomarkers that have mainly focused on inflammatory and hormonal pathways, the neurofilament proteins are exclusively expressed

in neurons and thereby represent markers specific for neuroaxonal injury. This means they are independent of the underlying causal pathways and provide a measure of brain injury with the advantage of being much more readily available compared to brain imaging techniques. However, results from previous studies suggest that time-point of measurement is of great importance when evaluating sNfL post-stroke. Previous data suggests that sNfL increases with the time from symptom onset to blood draw [128-130], and that sNfL remains elevated at 3 months post-stroke [128, 129]. Recent studies have found that sNfL measured at day 7 after symptom onset, but not as early as within 24 hours, is independently associated with 3-month functional outcome [128, 131]. Therefore, we aimed to further explore the temporal profile of sNfL post-stroke and to investigate sNfL measured at different time-points in relation to long- and short-term post-stroke outcomes. We additionally assessed differences in sNfL concentrations between cases and controls and the relation between sNfL concentrations and initial stroke severity and etiologic subtypes.

In summary, we found that 3-month sNfL showed strongest correlation to baseline NIHSS. Acute phase and 3-month sNfL was independently associated to NIHSS and mRS at 3 months, mRS at 2 years and NIHSS and mRS at 7 years. Both acute phase and 3-month sNfL in cases were significantly higher than in controls. Finally, we observed that the time profile of sNfL differs by etiologic subtype of ischemic stroke.

Associations between 3-month sNfL and outcomes measured as mRS and NIHSS at 3 months and 2 years post-stroke were independent of age, stroke severity and previous history of stroke. Similar findings were made for 3-month sNfL and more long-term outcomes for the subgroup that participated in the 7-year follow-up. Our findings support the potential use of sNfL as a predictor of post-stroke outcomes. Previous literature on sNfL and ischemic stroke outcome is very limited, especially regarding studies that have measured sNfL with the recently developed SiMoA assay. However, as stated above, a recent but smaller study (n=110) with data on sNfL measured at day 7 from admission also demonstrated an independent association to 3-month functional outcome [128].

Our results show that the magnitude of the sNfL increment within 3 months post-stroke was associated to stroke severity, i.e. baseline NIHSS, which is in line with previous data [130, 131] and expected since both are measures related to the magnitude of brain injury. Also in line with previous data, our results indicate that sNfL continues to increase during the first period after injury [129, 130, 184]. A recent study on a subgroup of patients with acute ischemic stroke (n=89) who had measurements of sNfL on day 1, 2, 3 and 7 after hospital admission observed highest concentrations at day 7 [128] and a study on phosphorylated neurofilament heavy protein observed highest concentrations

at 3 weeks post-stroke [185]. Interestingly, a small recent study on 30 ischemic stroke patients with repeated measurements of sNfL on day 0-1, 2-3, 7-9, 3 weeks and 3-5 months confirmed these findings with increasing sNfL during the acute phase and highest concentrations at 3 weeks [186]. In our study, sNfL 3 months post-stroke was significantly higher compared to the acute phase, but contrary to the acute phase, the 3-months concentrations were decreasing with the day of blood draw. Neurofilaments are abundant in large myelinated axons susceptible to Wallerian post-stroke degeneration and the time profile of immunohistochemical as well as radiological findings after stroke show a delayed and longstanding degenerative process fitting the release pattern of sNfL in our study [187, 188]. Taken together, these data point towards a peak in sNfL somewhere between the acute phase and 3 months post-stroke. However, it is reasonable to assume that this time point may also vary with, for instance, infarct size. Thus, future studies need to define the temporal profile of sNfL post-stroke in more detail in order to define the optimal time(s) for measurement, but clearly the very first days after stroke are too early. Speculatively, peak levels could add predictive ability of sNfL. Moreover, repeated measurements capturing the rates of increase and decrease in sNfL might be even better.

With respect to etiologic subtypes, we found highest acute phase and 3-month concentrations in LVD and CE stroke, which is expected since those etiologies in general cause larger strokes. At 7 years, SVD showed highest sNfL, although after adjustment the difference between the subtypes was not statistically significant at this time point in our sample. However, it is plausible that the concentrations are constantly elevated in some patients with SVD, due to a progressive disease course. In line with this, a study on patients with small subcortical infarcts (<25 mm) showed that sNfL was elevated in patients with new, but clinically silent cerebral small vessel disease-related lesions on MRI at follow-up [129]. Moreover, a recent study found that ischemic stroke patients with recurrent ischemic lesions on MRI at 6 months had higher sNfL compared to those without new lesions [128].

Our study is the largest on sNfL in ischemic stroke to date and the first on overall ischemic stroke including analyses of sNfL in etiologic subtypes. The longitudinal design with repeated blood draws and assessments of both functional and neurological outcomes in short and long-term provide novel information on the dynamics of sNfL, its relation to ischemic stroke subtypes, and its potential as a predictor of ischemic stroke outcomes. The results add to the scarce previous literature on sNfL in stroke and these results will inform the stroke research community in order to optimize the design of future studies on sNfL as a predictor of ischemic stroke outcomes.

Genome-wide association meta-analysis of functional outcome after ischemic stroke (Paper IV)

Functional outcomes after ischemic stroke have a wide range of inter-individual variability, from complete recovery to persistent severe disability. Genetic factors may account for part of this variability and studies in humans and animals support a genetic influence on recovery after brain injury [143-145]. Increased knowledge on the genetic contributors to a trait, such as post-stroke outcome, can provide information on individual risk that may enable targeted intervention, prevention or treatment. Furthermore, insights into pathophysiological mechanisms might lead to the identification of new therapeutic targets. Therefore, within the GISCOME network we performed a GWAS on functional outcome after stroke with the aim to identify genetic variants that may influence functional outcome after ischemic stroke.

We identified one intronic variant (rs1842681) in the *LOC105372028* gene, which was associated with outcome defined as mRS 0-2 vs 3-6 at the genome-wide significance level. The OR for a poor outcome was 1.40 for the minor allele, which in the context of a GWAS is a relatively large effect size, and the minor allele frequency was relatively high, i.e. 23%. We also detected several variants suggestively associated with outcome ($P < 10^{-5}$), of which some are related to genes with experimental evidence of influence on ischemic stroke volume and brain recovery.

The genome-wide significant SNP (rs1842681) for good vs poor functional outcome (mRS 0-2 vs. 3-6) is an intronic variant in the gene *LOC105372028*, lncRNA. LncRNAs represent a heterogeneous class of non-coding RNAs which are not translated into protein and are thought to modulate gene expression through multiple distinct mechanisms, e.g. by participating in both transcriptional and post-transcriptional stages [135]. The function of *LOC105372028* is not known, but expression analyses show highest gene expression in brain tissue (NCBI, 03/18). Concerning the putative functional role of rs1842681, it is predicted to alter several transcription factor regulatory motifs (HaploReg, 03/21/18), and it is a trans-eQTL for *KLRAQ1*, also known as *PPP1R21*, which encodes a regulatory subunit of protein phosphatase 1 (PP1). PP1 is a ubiquitous phosphatase implicated in many brain functions including learning and memory formation [189, 190]. In addition, PP1 is a key regulator of Ca^{2+}/calmodulin (CaM)-dependent protein kinase II (CaMKII) signaling which is crucial for Ca^{2+} mediated neuronal plasticity in the brain [191]. Speculatively, rs1842681 may thus modulate expression of PPP1R21 which in turn could affect brain plasticity and thereby outcome post-stroke, a hypothesis that requires validation through functional experiments.

We also identified 29 independent loci that were associated with ischemic stroke outcome at the pre-defined suggestive association level. Three of these loci are linked to genes with experimental evidence of influence on outcome from animal models of stroke [174-176].

In addition to our study, there is only one published GWAS on stroke outcome to date. This recent study restricted their inclusion to ischemic stroke cases with cortical infarcts of the anterior circulation causing moderate to severe stroke (NIHSS ≥5). In addition, individuals who were dependent on others (mRS score of >2) prior to index stroke were excluded. In a discovery cohort of 1225 individuals, and following joint analyses with a replication cohort (total n=2482), they found genome-wide significant associations for intronic variants in the *PATJ* gene and functional outcome at 3 months post-stroke [150]. This locus was not among the significant or suggestively associated variants in our study. Thus, *PATJ* might be of relevance for functional outcome after severe cortical infarcts of the anterior circulation in individuals without a previous severe functional impairment. This approach, using restricted inclusion criteria to study a specific subgroup of cases can provide valuable information given that molecular mechanisms and pathways have different relative impact on outcome in different subgroups of cases. The drawback, however, is the decrease in sample size. In samples limited to a few thousand individuals, it is only possible to detect variants with relatively large effect sizes. Since effect sizes of common genetic variants are generally small, as was concluded in our study, even larger samples are needed to identify additional variants associated with stroke outcome and to enable subgroup analyses. Thus, inclusion of all available cases to increase sample size for studying mechanisms that are common to overall ischemic stroke, in contrast to inclusion restricted to specific subgroups, represent complementary approaches. Another complementary approach is the study of other sorts of common genetic variation than SNPs. A recent study investigated CNVs for associations to 3-month functional outcome in around 3300 ischemic stroke cases. They found that increased genetic imbalance, i.e. the total number of protein coding genes affected by CNVs in an individual, was associated with unfavourable outcome [151].

To conclude, present studies on genetic variation and ischemic stroke outcomes have found support for some common variants with influence on post-stroke outcomes. However, to unravel the presumed heritability, i.e. the genetic component of outcome, more studies including larger sample sizes as well as complementary study designs are warranted. For instance, rare variants not detected with the GWAS design or gene – gene and gene – environmental interactions might also be at play [192].

Methodological considerations

The results from the studies included in this thesis are presented as effect estimates. The ambition is for these estimates to appropriately reflect a true relationship, which means that they are valid also outside the individual study situation. However, any effect estimate is likely to be affected by systematically (bias) or randomly (chance) induced errors. The internal validity of a study includes the quality of the data; the total bias, and the statistical power; the likelihood of randomly induced errors. Therefore, an appropriate study design, conduct and analysis, as well as a sufficient sample size is critical to reduce errors. On the other hand, the perfect study exists only in theory. In reality, additional factors, such as feasibility and costs need to be taken into account in order to optimize the use of available resources. Nevertheless, the results of a study should always be interpreted in light of the internal validity of the study. In the following sections, study design and potential sources of bias in the papers of this thesis are discussed. Finally, the external validity is considered, which means to which extent the results can be generalized to a broader population.

Study design

The conclusions that can be drawn from a study are naturally linked to the study design. In observational studies, such as the papers of this thesis, associations can be detected, but causation cannot be proven. However, different study designs are complementary and cost-effective observational studies are valuable to detect associations that merit further investigation in other types of studies, such as labor-intensive laboratory experimental studies or randomized controlled trials (RCTs).

Specific aspects of study design that limit the conclusions that can be drawn from this thesis include the sample sizes of all papers which were insufficient for adequate analyses of relevant subgroups, in particular the etiologic subgroups of ischemic stroke given their diversity regarding pathophysiological mechanisms. Further, we did not assess outcomes in the control group. Regarding cognitive impairment in paper II, this would have enabled evaluation of the relative importance of hemostatic biomarkers in stroke cases compared to a general population. In paper III, repeated blood sampling in narrower time intervals over the first weeks and months post-stroke would have been valuable to determine the trajectories of sNfL after stroke. Finally, the GWAS approach in paper IV is by design limited to the

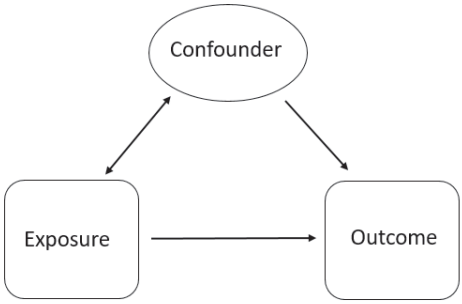
detection of genetic variants fitting the common disease – common variants hypothesis, described in the Background section.

From an ethical point of view, an optimal use of available resources is important. Study participants invest their time and might expose themselves to potential risks, medical or psychological, and donors, as well as tax-payers through government funding, invest their money in order to contribute to scientific progress. It is the responsibility of researchers to carefully use these resources in a way that is beneficial for patients. In this context, it is also important to include the patient's perspective in the design and conduction of a study. As an example, patient reported outcome measures (PROMs), implicates outcome measures directly reported by the individual who experienced it. In some instances, this can be more relevant compared to outcome measures reported by someone else, such as a physician or a nurse. In this thesis, PROMs have not been assessed as outcome measures, although these types of measures are also available in SAHLSIS (e.g. the Stroke Impact Scale, the Daily Fatigue Impacts Scale, and measures of quality of life with for instance EQ5D) and are of interest for future studies. Moreover, with respect to post-stroke outcomes, the diversity of impairments means that there are a variety of different outcome metrics that are of potential relevance. Here as well, the patients' perspective should be considered to accurately also address the most relevant aspects from a patient's, in addition to a healthcare or research, point of view.

Confounding

Confounding is a systematic error that occurs when the association between an exposure and outcome is affected by additional factors (confounders). Confounders affect outcome, and are associated to the exposure that is aimed to study, Figure 15.

Figure 15. A confounder affects the outcome, and is associated to the exposure that is aimed to study.



To reduce errors induced by confounding we have adjusted for potential confounders in the statistical analyses. However, the selection of potential confounders is not always straightforward since knowledge on all underlying causal pathways is sometimes limited. With respect to circulating concentrations of hemostatic biomarkers there are many potential confounders, especially for measurements in the acute phase of ischemic stroke. This is one of the reasons why we in paper I and II chose to focus on convalescent plasma concentrations that are less influenced by the acute phase response resulting from tissue injury as well as acute pharmacological treatments. The well-characterized study sample in SAHLSIS has enabled adjustment for several possible confounders. However, there is still a possibility that uncontrolled confounding, such as pharmaceutical treatment or comorbidities has influenced our results.

Selection bias

Selection bias occurs when the selected study subjects do not represent the population that it aims to describe. This can arise at inclusion to a study or be due to loss to follow-up (attrition bias). SAHLSIS is based on hospitalized consecutive ischemic stroke cases. Of 645 eligible cases between 1998 and 2003, 45 were not included, 29 because of unwillingness to participate and 16 because they died before informed consent could be given. Moreover, only cases admitted to a stroke unit were invited to participate. The controls in a case-control study should be representative of the same population as the cases. The controls in SAHLSIS were randomly selected from the general population in the same geographical area as cases. However, a significant proportion of invited controls did not participate (208 did not respond, 191 were unwilling

to participate, and 108 were excluded because of a history of cardiovascular disease). To conclude, we cannot fully exclude the possibility of selection bias.

In paper I, results were based on time to an event, and cases were included until an event occurred, which lowers the risk of introducing selection bias. However, with respect to the long-term follow-ups in paper II and III that are not based on register data, there was a loss to follow-up that could have influenced our results. Of the 411 initially included cases from the main center at Sahlgrenska, 358 were alive at 7 years post-stroke and 268 and 272 of those were included in the outcome analyses of paper II and III, respectively. Thus, the drop out rate was around 25%. However, to assess long-term outcomes that requires follow-up visits and extensive testing this is probably inevitable. Importantly, this needs to be kept in mind when interpreting the results on long-term follow-up in paper II and III.

In paper IV, we can also not exclude the possibility of selection bias since data entry also requested availability of functional status follow-up data and there might have been loss to follow-up.

Information bias

Information bias can result from misclassification or measurement errors. To reduce measurement errors of biomarker concentrations in SAHLSIS, blood sampling has been strictly standardized and performed at 8:30 AM - 10:30 AM after overnight fasting and all samples have been aliquoted and stored at -80 °C pending analysis. Furthermore, for all protein measurements individuals performing the laboratory analyses have been blinded to clinical information. ELISA was the method of choice used for protein measurements of t-PA, VWF and TAFI in papers I and II. This method relies on antibodies to detect a specific protein. As genes are multi-exonic structures, alternative splicing can occur which produces mRNA and protein isoforms with differing roles. Further, underlying genetic variation within a gene's coding region can result in an alternative amino acid sequence (missense mutation) within a protein, contributing further to protein variation. All of these can affect the affinity of antibodies used in such immunological assays. For TAFI, a common SNP leading to a missense mutation has been demonstrated to greatly affect immunoreactivity in several commercially available ELISAs that measure total TAFI-antigen [193]. Thus for TAFI measurement, we instead used a well-characterized assay for TAFI-AP, which only includes amino acids 1-92 and is intrinsically isoform-independent to minimize potential experimental artefacts [194]. Apart from different isoforms, proteins can also circulate in different molecular forms or complexes. For instance, as previously mentioned

in the Background section, t-PA circulates both as a free form and in complex with different inhibitors, mainly PAI-1, and VWF circulates in large multimers of differing sizes. These complexes can, similarly, affect antibody binding affinity. For measurements of t-PA antigen we used an ELISA with antibodies with similar affinities to all molecular forms [195]. While we tried to minimize any possible experimental artefacts (measurement errors), we cannot rule out that differential reactivities of the ELISAs could exist. With respect to paper III, sNfL was measured by the most sensitive method presently available [122].

Another important aspect is the time-point of blood sampling. This is illustrated by the sNfL concentrations that are highly correlated to day of blood sampling during the acute phase and where the time-point of measurement seems to be crucial for the predictive ability of sNfL on post-stroke outcomes. In this case, since concentrations diverge, the measurement errors are likely to be larger in cases with early, i.e. day 1, compared to later, i.e. day 6-7 blood draw. Ideally, to fully capture the dynamics of sNfL post-stroke, repeated measurements in each individual should have been performed. Also in the case of hemostatic biomarkers the time-point of measurement is relevant. We chose to analyze convalescent concentrations measured approximately 3 months after stroke. However, we do not know if these concentrations reflect the individual pre-stroke baseline concentrations or if they remain influenced by the stroke event.

To obtain information on recurrent vascular events, survival rates and causes of death for paper 1, multiple overlapping methods were used, including national registers with high coverage. Since the high coverage in registers might be at the expense of detailed and accurate information complementary interviews and review of medical records were performed. Nevertheless, we cannot exclude that some misclassification has occurred. Moreover, with respect to recurrent stroke, the exclusion of TIAs and the inclusion of both ischemic and hemorrhagic strokes might have reduced our ability to detect associations between hemostatic biomarkers and recurrent cerebral ischemia.

We used validated scales and tests to assess functional, neurological and cognitive outcomes and raters were blinded to biomarker results. A limitation is that we did not have information on individual baseline or pre-morbid scores. Within SAHLSIS a restricted number of trained individuals performed the assessments. Thus, the inter-rater reliability is unlikely to be a source of measurement error. However, in paper IV that is a multi-center study we do not know if the mRS scoring was performed uniformly between the different studies. Moreover, the use of the mRS as primary outcome measure in paper IV is a limitation since it represents a crude measurement of outcome that could be affected by other factors than the recovery mechanisms that we are searching. Ideally, several more detailed outcome metrics would have been

used assessing different domains of recovery such as motor impairment, aphasia, neglect, and cognitive impairment. Moreover, aiming to detect genetic variants involved in recovery mechanisms we would ideally have had information on pre-morbid mRS to be able to calculate the change in functional ability. Finally, the mRS scores were assessed at different time points ranging from 60-190 days, potentially diluting any detectable associations as functional outcome status may vary over time. However, the majority of included studies assessed mRS less than two weeks before or after 90 days after stroke onset.

Statistical aspects

The choice of statistical methods is of course fundamental, and this is why statisticians have been involved in the analytic process of all papers. In the analytical process, some of the errors due to, for instance, confounding and misclassification can be reduced. Through regression modelling, we adjusted the analyses of all papers for potential confounders. To reduce errors due to misclassification we performed sensitivity analyses excluding individuals with neurological comorbidities and recurrent stroke during follow-up in paper III. In paper II, we chose a different approach by adjusting for time since last stroke in the regression models. SAHLSIS has deep phenotypic data, which has enabled adjustment for several potential confounders. It follows that accurate measures and classification of confounders is crucial for this purpose. As an example, in paper IV, analyses were adjusted for stroke severity as assessed by the NIHSS. The NIHSS varies over time post-stroke and the time-points of scoring varied between the included studies, which could have been a source of bias. However, in a majority of cases NIHSS was scored early and in only 160 individuals later than day 3 after admission, which is why this probably did not have a major impact on our results.

With respect to GWAS (paper IV), there are several specific aspects that merit consideration.

Multiple testing, Sample size and Power:

In GWAS a large number of genetic variants, typically around 1 million, are genotyped. In addition, imputation is usually performed, which means that a dataset such as the 1000 Genomes Project is used to estimate genotypes of untyped SNPs by establishing the pattern of LD between nearby SNPs and apply to the sampled data. This results in an even larger number of variants. It follows that a large number of statistical tests need to be performed, which will inevitably lead to false positive results. It is therefore necessary to compensate for multiple testing. A traditional method is Bonferroni correction, in which the *P* value is multiplied with the number of tests performed. However, in

genetic studies such as GWAS this approach is considered too conservative because of the correlation between SNPs (please refer to the description of haplotypes, page 22). In practice, a significance threshold of $P < 5\times10^{-8}$, corresponding to a Bonferroni correction of one million tests, has been generally accepted as a threshold for genome-wide significance. Since the aim is to detect variants with moderate to low effects sizes, this means that large samples sizes are needed to obtain adequate power given this strict significance threshold.

In paper IV, more than six thousand ischemic stroke cases were included, but only one significant locus was identified. Thus, the findings do not explain the assumed genetic variation for ischemic stroke outcome. This implies that the effect sizes of individual SNPs on outcome of ischemic stroke are small and that our sample size might have been insufficient even to detect common genetic variants.

Population stratification:

Population stratification, or geographical confounding, occurs when a sample consists of subpopulations (e.g. countries of a continent) between which there has been less mating, so that allele frequencies of common variants also differ between the groups. If the proportions of each subpopulation are unevenly distributed between the categories of the outcome measure, variants that differ in frequency between these subpopulations will appear to be associated with the outcome. There are different approaches to identify this kind of population stratification, such as performing principal component analysis (PCA) and subsequently exclude outliers and adjust for principal components in the regression analyses, which we did in paper IV. The rationale behind this approach is that individuals who are geographically close are likely to be more closely related and thus correlated in terms of genotypes. The first principal component gives the linear combination of genotypes that best capture the variation in the data. The second principal component is the orthogonal combination that best captures the remaining variation in the data, and so forth. As an example, in Europe, PCA can distinguish individuals from different countries.

Replication:

Given the hypothesis-free approach of GWAS and the large number of tests performed, any detected variant or locus is usually not established as associated with a disease or outcome until it has been replicated, which means that an association is demonstrated also in a second independent sample. Moreover, due to the winner's curse phenomenon effect sizes from the first discovery sample are likely to be overestimated and thus the replication sample

should preferably be larger. By using a replication sample more accurate and unbiased risk estimates can be obtained.

In paper IV, we did not have any available cohorts for replication of our findings, which is a limitation. Thus, future replication in independent cohorts is warranted as they become available.

Generalizability

SAHLSIS includes adult ischemic stroke cases younger than 70 years of age, which is why the results of papers I-III are only representative for this age group. For instance, since women generally get stroke at a higher age compared to men [196], in the general ischemic stroke population a larger proportion of the cases are women compared to in our sample. It is also of note that the proportion of cases with cardiovascular risk factors and comorbidities vary with age, as does the distribution of etiological subtypes of ischemic stroke. Moreover, SAHLSIS is based on hospitalized cases. However, in Sweden more than 95% of patients with suspected stroke before 74 years of age are hospitalized [197] , therefore our results are likely to be generalizable to Swedish ischemic stroke before 70 years of age.

In paper IV, the study population comprised of ischemic stroke cases from different studies in Europe, USA and Australia. Participants were of European ancestry, and the results are thus not generalizable to stroke populations of other ethnicities. In addition, the study sample mainly reflected milder strokes (median NIHSS of 4) and detected variants might not have the same impact on more severe strokes.

CONCLUSIONS TO GIVEN AIMS

We found a significant association between circulating convalescent (3-month) concentrations of t-PA and the long-term risk of vascular death after ischemic stroke, and this association remained when adjusting for age, diabetes mellitus and etiologic subtype. The combined effect of t-PA, VWF, and fibrinogen, was independently associated with the risk of coronary events. Analyses stratified for diabetes revealed a larger effect of VWF levels on the risk of coronary events in non-diabetic patients.

In participants younger than 50 years at index stroke, convalescent concentrations of fibrinogen were independently associated to overall cognitive function as assessed by the BNIS, and processing speed as assessed by the Trailmaking Test A 7 years post-stroke. In the whole group, fibrinogen, VWF, and t-PA concentrations were correlated to 7-year cognitive function, but these associations were not independent of confounding factors.

sNfL increased with the day of blood draw during the acute phase of ischemic stroke and concentrations were highest after 3 months. High sNfL associated to poor outcomes evaluated by NIHSS and mRS both in the short and long term, and all associations were strongest for 3-month sNfL and independent of age and stroke severity. The predictive value of sNfL for discriminating good vs poor outcome (mRS 0-2 vs 3-6) was of similar magnitude as baseline NIHSS. We also showed that the dynamics of sNfL differs by subtype of ischemic stroke.

We identified one genetic variant associated with 3-month functional outcome with genome-wide significance (mRS 0-2 vs 3-6). This intronic variant (rs1842681) in the *LOC105372028* gene, is a previously reported trans-eQTL for *PPP1R21*, which encodes a regulatory subunit of PP1. We also detected several variants with suggestive association with outcome ($P < 10^{-5}$), some of which are within or near genes with experimental evidence of influence on ischemic stroke volume and/or brain recovery (e.g. *NTN4*, *TEK* and *PTCH1*).

FUTURE PERSPECTIVES

The results from this thesis demonstrate associations between protein and genetic biomarkers and outcomes after ischemic stroke. They represent small steps towards an increased understanding of mechanisms underlying the individual variability in post-stroke outcomes as well as an improved individual prognostic prediction. Such achievements could ultimately contribute to an improved personalized post-stroke management. However, this is a long-term vision and much remains for its realization.

Circulating protein biomarkers have the potential of capturing several aspects of an individual's response to injury, by the magnitude of its increase or decrease, but also by its temporal profile. As confirmed by the results of this thesis, the time-point of measurement is crucial. Hypothetically, genetic or other factors could explain why some individuals have a more unfavourable response to injury with a long-standing or chronic course as opposed to a more rapid healing. If this could be quantified or predicted by biomarker measurements, those individuals could be identified and possibly provided with targeted interventions. Moreover, identifying the molecular pathways involved might lead to new targets for interventions or drug development. To be able to capture the dynamics of injury and recovery processes, repeated measures of both biomarkers and outcomes will be required in future studies. Such efforts are currently ongoing within our and other research groups. Another generally important aspect, also illustrated by the results of this thesis, is the need for stratified analyses examining specific subgroups including, for instance, etiologic subtypes, sex, age, and comorbidities. These types of analyses will require larger sample sizes.

In order to increase knowledge on mechanisms underlying post-stroke outcomes and recovery, a major challenge is the identification of causal relationships, especially for proteins involved in complex biological pathways, and under potential influence of numerous confounders, such as the hemostatic proteins studied in this thesis. Observational biomarker studies can identify associations with different outcomes, but to establish causality complementary approaches are needed. One way is through Mendelian randomization studies, in which genetic variants with known impact on an exposure, for instance protein biomarker concentrations, are assessed for associations to an outcome measure. If the protein biomarker is causally related to the outcome, then the genetic variants will be associated with the outcome to the extent predicted by

their effect on the protein biomarker concentration. As knowledge on effects of common genetic variants on different traits, including protein biomarker concentrations, is increasing, this approach has gained attention in recent years. Thus for the future, the combination of biomarker studies and Mendelian randomization studies can, for instance, lead to the identification of potential drug targets that subsequently can be passed on or dismissed after confirming or ruling out causality.

Based on previous knowledge, it is plausible to assume that the hemostatic pathway is involved in mechanisms underlying several post-stroke outcomes. In this thesis, we assessed some hemostatic biomarkers for associations to outcomes, and provide some insights on the association between prothrombotic factors and poor outcomes. However, the assessed hemostatic proteins are part of a complex interplay with a large number of enzymes. For future studies, a different, and potentially more informative approach might be the assessment of a more global measure of the thrombotic state. Alternatively, panels integrating several individual biomarkers could, for instance, improve prognostic prediction. For this purpose, however, a combination with many different parameters, in addition to prothrombotic biomarkers, is likely advantageous and could provide a way to improve the predictive accuracy as knowledge increases and new parameters can be added.

By its specificity for neuroaxonal injury, sNfL provides a measure of brain injury with the advantage of being more readily available compared to brain imaging techniques. It is therefore a promising candidate for post-stroke outcome prediction, a postulation that was supported by the results of paper III. Speculatively, repeated measurements capturing the rates of increase and decrease in sNfL could further add to its predictive value. Thus, repeated measures are the warranted next step to define the post-stroke dynamics of sNfL, also in specific subgroups.

With respect to the genetic contribution to post-stroke outcomes, the results of paper IV confirm the usefulness of a GWAS approach. Future studies should include larger sample sizes and several ethnicities in order to replicate our findings, to detect additional variants, and to investigate specific subgroups defined by, for instance, infarct location (cortical vs non-cortical stroke), sex, certain comorbidities and ethnicity. We aim to cover some of these aspects in an ongoing second phase of the GISCOME project. In order to elucidate different aspects of recovery, future studies should ideally also include several, more specific outcome measures, such as motor impairment, aphasia, neglect and cognitive impairment. To obtain sufficient sample sizes, such studies will require further international collaborative efforts, which entail challenges such

as harmonization of phenotype data. For these studies to be successful, the demand for large sample sizes should not be at the cost of well-characterized samples. Importantly at last, in order to increase knowledge on biological mechanisms underlying post-stroke outcomes, the essential next step following detection of associated genetic loci in GWAS is functional characterization through experimental studies in order to unravel the underlying molecular pathways. Another future application of variants detected in GWAS is their integration into polygenic risk scores in order to provide prognostic information at the individual level.

ACKNOWLEDGEMENTS – TACK

Jag vill rikta ett stort och varmt tack till alla som bidragit till denna avhandling. Särskilt vill jag tacka:

Min handledare, *Christina Jern,* för att du gav mig chansen att introduceras till forskning och för att du stöttat mig sedan dess. Du inspirerar genom din kunskap, din entusiasm, din förmåga att snabbt omvandla problem till nya möjligheter och ditt sätt att fylla samtal och möten med nya idéer. Dessutom, när jag verkligen behövt det har du alltid funnits där, oavsett när, eller var du befunnit dig.

Min bihandledare, *Tara Stanne,* för att du är en förebild genom hur du kombinerar kvalitet med effektivitet och samtidigt alltid är hjälpsam och generös med din tid. Du har tålmodigt läst i stort sett allt jag skrivit. Tack för vänskap, positiv energi och pepp! Vi kom in i forskargruppen vid samma tid och du har gjort jobbet roligare – från första dagen.

Mina bihandledare, *Staffan Nilsson* och *Katarina Jood*. Staffan, din statistiska kompetens har varit oumbärlig. Jag uppskattar också att du ibland kan ha en befriande distans till alltihop. Katarina, din kliniska kunskap och dina kloka synpunkter har varit värdefulla längs vägen.

Ingrid Eriksson, Martina Olsson, Cecilia Lagging och *Sofia Klasson* för att ni alltid är så stöttande och hjälpsamma. Ingrid, du tar hand om patienterna på ett fantastiskt sätt och håller koll på stort som smått i gruppen. Cecilia, tack för korrekturläsning, men framför allt diskussioner och rumssällskap under det senaste året, det har verkligen varit uppskattat! Sofia, tack för all värdefull hjälp, inte minst de sista veckorna.

Gamla och nuvarande medarbetare i SAHLSIS-gruppen. Ett särskilt tack till *Christian Blomstrand,* för att du delar med dig av din stora kunskap om stroke och för värdefull återkoppling på delar av ramen, och till *Hans Samuelsson,* för att du bidragit med din kunskap om kognition till manus och ram. Mina medförfattare, *Petra Redfors*, *Lukas Holmegaard, Maja Olsson* och *Jo Viken.* Jag vill också tacka *Sven-Öjvind Swahn* för praktisk hjälp genom åren.

I would also like to express my gratitude to *the GISCOME network*, especially *Martin Söderholm, Erik Lorentzen, Arne Lindgren* and *Jane Maguire*. I am so grateful for being invited to join the collaborative work of GISCOME. Thank you for the intense, inspiring and fruitful collaboration throughout the manuscript writing process.

Medförfattare från andra grupper och samarbeten. *Lars Rosengren, Henrik Zetterberg* och *Kaj Blennow* för samarbetet med NfL och *Ann Gils* och *Paul DeClerk* för samarbetet med TAFI.

Sektionschef *Lovisa Lovmar* och enhetschef *Mirja Henricsson Marcher* på Klinisk Genetik och Genomik, för att ni har gjort det lätt att kombinera forskning med klinisk tjänstgöring. Tack också till er och alla övriga kollegor på Klinisk Genetik och plan 3 för att ni tillsammans skapar en så fin arbetsmiljö. Särskilt tack till mina närmaste medarbetare i mottagningsgruppen och ST-gruppen.

Tidigare och nuvarande institutionsledning vid Institutionen för Biomedicin, prefekterna *Anders Oldfors* och *Sven Enerbäck* med medarbetare, för att ni skapar en god forskningsmiljö.

Maria Lagging för illustrationer till denna bok.

Lärare och klasskamrater på Kliniska epidemiologiska forskarskolan.

Alla patienter och kontrollpersoner, för ert ovärderliga bidrag genom deltagande i studierna, ett stort och hjärtligt tack.

Mina vänner, ni som kommit in under olika perioder av mitt liv, och finns kvar. Tack för att ni delar, förstår och gör livet roligare!

Mamma och *pappa. Erik, Natalie* och *Arvid*. Tack för aldrig sinande stöd och uppmuntran, inspiration och glädje!

Morgan, min motvikt och kärlek. Jag är evigt tacksam för allt du gör och är. Älskade *Stella* och *Anna*. Tack för perspektivet ni ger, varje dag en påminnelse om vad som är viktigt, på riktigt.

Arbetet i denna avhandling har genomförts med stöd av Vetenskapsrådet, Hjärt-Lungfonden, Svenska staten genom Avtal om Läkarutbildning och Forskning, ALF, Sroke-Riksförbundet och Insamlingsstiftelsen för Neurologisk Forskning.

The work presented in this thesis was supported by the Swedish Research Council, the Swedish Heart and Lung Foundation, grants from the Swedish state under the agreement between the Swedish government and the county councils, the ALF-agreement, The Swedish Stroke Association and the Gothenburg Foundation for Neurological research.

REFERENCES

1. Quest DO. Stroke: A Selective History. *Neurosurgery*. 1990;27:440-445
2. Tatu L, Moulin T, Monnier G. The Discovery of Encephalic Arteries. From Johann Jacob Wepfer to Charles Foix. *Cerebrovascular Diseases*. 2005;20:427-432
3. Focus Medica. *Landmarks in stroke*. Focus Medica Pte Limited; 2011.
4. Wardlaw JM, Murray V, Berge E, del Zoppo G, Sandercock P, Lindley RL, et al. Recombinant tissue plasminogen activator for acute ischaemic stroke: An updated systematic review and meta-analysis. *Lancet*. 2012;379:2364-2372
5. Lambrinos A, Schaink AK, Dhalla I, Krings T, Casaubon LK, Sikich N, et al. Mechanical Thrombectomy in Acute Ischemic Stroke: A Systematic Review. *The Canadian Journal of Neurological Sciences. Le journal canadien des sciences neurologiques*. 2016;43:455-460
6. Organised inpatient (stroke unit) care for stroke. *The Cochrane database of systematic reviews*. 2013:Cd000197
7. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the global burden of disease study 2010. *Lancet*. 2012;380:2095-2128
8. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380:2197-2223
9. Feigin VL, Norrving B, Mensah GA. Global Burden of Stroke. *Circulation Research*. 2017;120:439-448
10. Kissela BM, Khoury JC, Alwell K, Moomaw CJ, Woo D, Adeoye O, et al. Age at stroke: Temporal trends in stroke incidence in a large, biracial population. *Neurology*. 2012;79:1781-1787
11. Rosengren A, Giang KW, Lappas G, Jern C, Toren K, Bjorck L. Twenty-Four-Year Trends in the Incidence of Ischemic Stroke in Sweden from 1987 to 2010. *Stroke.* 2013;44:2388-2393
12. Https://www.World-stroke.Org/images/wso_global_stroke_fact_sheet.Pdf. 2019:WSO Global Stroke Fact Sheet
13. Stroke--1989. Recommendations on Stroke Prevention, Diagnosis, and Therapy. Report of the who task force on stroke and other cerebrovascular disorders. *Stroke*. 1989;20:1407-1431

14. Appelros P, Jonsson F, Asberg S, Asplund K, Glader EL, Asberg KH, et al. Trends in Stroke Treatment and Outcome between 1995 and 2010: Observations from Riks-Stroke, the Swedish Stroke Register. *Cerebrovascular diseases*. 2014;37:22-29
15. Adams HP, Jr., Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke*. 1993;24:35-41
16. Bamford J, Sandercock P, Dennis M, Burn J, Warlow C. Classification and natural history of clinically identifiable subtypes of cerebral infarction. *Lancet*. 1991;337:1521-1526
17. Redfors P, Jood K, Holmegaard L, Rosengren A, Blomstrand C, Jern C. Stroke subtype predicts outcome in young and middle-aged stroke sufferers. *Acta Neurologica Scandinavica*. 2012
18. Ntaios G, Papavasileiou V, Makaritsis K, Milionis H, Michel P, Vemmos K. Association of ischaemic stroke subtype with long-term cardiovascular events. *European Journal of Neurology*. 2014;21:1108-1114
19. Petty GW, Brown RD, Jr., Whisnant JP, Sicks JD, O'Fallon WM, Wiebers DO. Ischemic Stroke Subtypes : A Population-Based Study of Functional Outcome, Survival, and Recurrence. *Stroke.* 2000;31:1062-1068
20. Ornello R, Degan D, Tiseo C, Di Carmine C, Perciballi L, Pistoia F, et al. Distribution and Temporal Trends From 1993 to 2015 of Ischemic Stroke Subtypes: A Systematic Review and Meta-Analysis. *Stroke*. 2018;49:814-819
21. Kirshner HS. Differentiating ischemic stroke subtypes: Risk factors and secondary prevention. *Journal of the Neurological Sciences*. 2009;279:1-8
22. Rovira A, Grive E, Rovira A, Alvarez-Sabin J. Distribution territories and causative mechanisms of ischemic stroke. *European Radiology*. 2005;15:416-426
23. Bailey EL, Smith C, Sudlow CL, Wardlaw JM. Pathology of lacunar ischemic stroke in humans--a systematic review. *Brain pathology*. 2012;22:583-591
24. Bamford J, Sandercock P, Jones L, Warlow C. The Natural History of Lacunar Infarction: The Oxfordshire Community Stroke Project. *Stroke.* 1987;18:545-551
25. Arboix A, Alio J. Acute cardioembolic stroke: An update. *Expert review of cardiovascular therapy*. 2011;9:367-379
26. Grau AJ, Weimar C, Buggle F, Heinrich A, Goertler M, Neumaier S, et al. Risk factors, outcome, and treatment in subtypes of ischemic stroke: The German Stroke Data Bank. *Stroke*. 2001;32:2559-2566

27. Hart RG, Catanese L, Perera KS, Ntaios G, Connolly SJ. Embolic Stroke of Undetermined Source: A systematic review and clinical update. *Stroke.* 2017;48:867-872
28. Kent DM, Dahabreh IJ, Ruthazer R, Furlan AJ, Reisman M, Carroll JD, et al. Device closure of patent foramen ovale after stroke: Pooled analysis of completed randomized trials. *Journal of the American College of Cardiology*. 2016;67:907-917
29. Jood K, Ladenvall C, Rosengren A, Blomstrand C, Jern C. Family History in Ischemic Stroke Before 70 Years of Age: The Sahlgrenska Academy Study on Ischemic Stroke. *Stroke*. 2005;36:1383-1387
30. Rolfs A, Fazekas F, Grittner U, Dichgans M, Martus P, Holzhausen M, et al. Acute Cerebrovascular Disease in the Young: The Stroke in Young Fabry Patients Study. *Stroke*. 2013;44:340-349
31. Putaala J, Metso AJ, Metso TM, Konkola N, Kraemer Y, Haapaniemi E, et al. Analysis of 1008 consecutive patients aged 15 to 49 with first-ever ischemic stroke: The Helsinki Young Stroke Registry. *Stroke*. 2009;40:1195-1203
32. Ganesh A, Gutnikov SA, Rothwell PM. Late functional improvement after lacunar stroke: A population-based study. *Journal of Neurology, Neurosurgery, and Psychiatry*. 2018;89:1301-1307
33. Duncan PW, Lai SM, Keighley J. Defining post-stroke recovery: Implications for design and interpretation of drug trials. *Neuropharmacology*. 2000;39:835-841
34. Rutten-Jacobs LC, Maaijwee NA, Arntz RM, Schoonderwaldt HC, Dorresteijn LD, van der Vlugt MJ, et al. Long-term risk of recurrent vascular events after young stroke: The FUTURE study. *Annals of Neurology*. 2013;74:592-601
35. Rutten-Jacobs LC, Arntz RM, Maaijwee NA, Schoonderwaldt HC, Dorresteijn LD, van Dijk EJ, et al. Cardiovascular disease is the main cause of long-term excess mortality after ischemic stroke in young adults. *Hypertension*. 2015;65:670-675
36. Planton M, Peiffer S, Albucher JF, Barbeau EJ, Tardy J, Pastor J, et al. Neuropsychological outcome after a first symptomatic ischaemic stroke with 'good recovery'. *European Journal of Neurology*. 2012;19:212-219
37. Mercier L, Audet T, Hebert R, Rochette A, Dubois MF. Impact of motor, cognitive, and perceptual disorders on ability to perform activities of daily living after stroke. *Stroke*. 2001;32:2602-2608
38. Ullberg T, Zia E, Petersson J, Norrving B. Changes in functional outcome over the first year after stroke: An observational study from the Swedish Stroke Register. *Stroke*. 2015;46:389-394
39. Sennfalt S, Norrving B, Petersson J, Ullberg T. Long-term survival and function after stroke. *Stroke*. 2018:Strokeaha118022913
40. Mahoney FI, Barthel DW. Functional evaluation: The barthel index. *Maryland state medical journal*. 1965;14:61-65

41. Quinn TJ, Dawson J, Walters MR, Lees KR. Functional outcome measures in contemporary stroke trials. *International Journal of Stroke*. 2009;4:200-205
42. Banks JL, Marotta CA. Outcomes validity and reliability of the modified Rankin Scale: Implications for stroke clinical trials: A literature review and synthesis. *Stroke*. 2007;38:1091-1096
43. Wilson JT, Hareendran A, Hendry A, Potter J, Bone I, Muir KW. Reliability of the modified Rankin Scale across multiple raters: Benefits of a structured interview. *Stroke*. 2005;36:777-781
44. Shinohara Y, Minematsu K, Amano T, Ohashi Y. Modified Rankin Scale with expanded guidance scheme and interview questionnaire: Interrater agreement and reproducibility of assessment. *Cerebrovascular Diseases*. 2006;21:271-278
45. Woo D, Broderick JP, Kothari RU, Lu M, Brott T, Lyden PD, et al. Does the National Institutes of Health Stroke Scale favor left hemisphere strokes? NINDS t-PA stroke study group. *Stroke*. 1999;30:2355-2359
46. Fink JN, Selim MH, Kumar S, Silver B, Linfante I, Caplan LR, et al. Is the Association of National Institutes of Health Stroke Scale scores and acute Magnetic Resonance Imaging Stroke Volume Equal For Patients With Right- and Left-Hemisphere Ischemic Stroke? *Stroke*. 2002;33:954-958
47. Weimar C, Konig IR, Kraywinkel K, Ziegler A, Diener HC. Age and National Institutes of Health Stroke Scale score within 6 hours after onset are accurate predictors of outcome after cerebral ischemia: Development and external validation of prognostic models. *Stroke*. 2004;35:158-162
48. Askim T, Bernhardt J, Churilov L, Indredavik B. The Scandinavian Stroke Scale is equally as good as the National Institutes of Health Stroke Scale in identifying 3-month outcome. *Journal of Rehabilitation Medicine*. 2016;48:909-912
49. Kauranen T, Turunen K, Laari S, Mustanoja S, Baumann P, Poutiainen E. The severity of cognitive deficits predicts return to work after a first-ever ischaemic stroke. *Journal of Neurology, Neurosurgery, and Psychiatry*. 2013;84:316-321
50. Verhoeven CL, Post MW, Schiemanck SK, van Zandvoort MJ, Vrancken PH, van Heugten CM. Is cognitive functioning 1 year poststroke related to quality of life domain? *Journal of Stroke and Cerebrovascular Diseases*. 2011;20:450-458
51. Schaapsmeerders P, Maaijwee NA, van Dijk EJ, Rutten-Jacobs LC, Arntz RM, Schoonderwaldt HC, et al. Long-term cognitive impairment after first-ever ischemic stroke in young adults. *Stroke.* 2013;44:1621-1628

52. Redfors P, Hofgren C, Eriksson I, Holmegaard L, Samuelsson H, Jood K. The Barrow Neurological Institute Screen for Higher Cerebral Functions in Cognitive Screening after Stroke. *Journal of Stroke and Cerebrovascular Diseases*. 2014;23:349-355
53. Hofgren C, Esbjornsson E, Aniansson H, Sunnerhagen KS. Application and validation of the Barrow Neurological Institute Screen for higher cerebral functions in a control population and in patient groups commonly seen in neurorehabilitation. *Journal of Rehabilitation Medicine*. 2007;39:547-553
54. Boosman H, Visser-Meily JM, Post MW, Duits A, van Heugten CM. Validity of the Barrow Neurological Institute (BNI) Screen for Higher Cerebral Functions in stroke patients with good functional outcome. *The Clinical Neuropsychologist*. 2013;27:667-680
55. Reitan RM. The relation of the Trail Making Test to organic brain damage. *Journal of Consulting Psychology*. 1955;19:393-394
56. Putaala J, Haapaniemi E, Metso AJ, Metso TM, Artto V, Kaste M, et al. Recurrent ischemic events in young adults after first-ever ischemic stroke. *Annals of Neurology*. 2010;68:661-671
57. Appelros P, Gunnarsson KE, Terent A. Ten-year risk for myocardial infarction in patients with first-ever stroke: A community-based study. *Acta Neurologica Scandinavica*. 2011;124:383-389
58. Dhamoon MS, Tai W, Boden-Albala B, Rundek T, Paik MC, Sacco RL, et al. Risk of myocardial infarction or vascular death after first ischemic stroke: The Northern Manhattan Study. *Stroke*. 2007;38:1752-1758
59. Schmidt M, Jacobsen JB, Johnsen SP, Botker HE, Sorensen HT. Eighteen-year trends in stroke mortality and the prognostic influence of comorbidity. *Neurology*. 2014;82:340-350
60. Kammersgaard LP, Olsen TS. Cardiovascular risk factors and 5-year mortality in the copenhagen stroke study. *Cerebrovascular Diseases*. 2006;21:187-193
61. Rutten-Jacobs LC, Arntz RM, Maaijwee NA, Schoonderwaldt HC, Dorresteijn LD, van Dijk EJ, et al. Long-term mortality after stroke among adults aged 18 to 50 years. *Jama*. 2013;309:1136-1144
62. Lopez-Espuela F, Pedrera-Zamorano JD, Jimenez-Caballero PE, Ramirez-Moreno JM, Portilla-Cuenca JC, Lavado-Garcia JM, et al. Functional status and disability in patients after acute stroke: A longitudinal study. *American journal of critical care.* 2016;25:144-151
63. Chang WH, Sohn MK, Lee J, Kim DY, Lee SG, Shin YI, et al. Predictors of functional level and quality of life at 6 months after a first-ever stroke: The KOSCO study. *Journal of Neurology*. 2016;263:1166-1177

64. Synhaeve NE, Arntz RM, van Alebeek ME, van Pamelen J, Maaijwee NA, Rutten-Jacobs LC, et al. Women have a poorer very long-term functional outcome after stroke among adults aged 18-50 years: The FUTURE study. *Journal of Neurology*. 2016;263:1099-1105
65. Willers C, Lekander I, Ekstrand E, Lilja M, Pessah-Rasmussen H, Sunnerhagen KS, et al. Sex as predictor for achieved health outcomes and received care in ischemic stroke and intracerebral hemorrhage: A register-based study. *Biology of sex differences*. 2018;9:11
66. Pohjasvaara T, Vataja R, Leppavuori A, Kaste M, Erkinjuntti T. Depression is an independent predictor of poor long-term functional outcome post-stroke. *European Journal of Neurology*. 2001;8:315-319
67. Fischer U, Arnold M, Nedeltchev K, Schoenenberger RA, Kappeler L, Hollinger P, et al. Impact of comorbidity on ischemic stroke outcome. *Acta Neurologica Scandinavica*. 2006;113:108-113
68. Jimenez Caballero PE, Lopez Espuela F, Portilla Cuenca JC, Ramirez Moreno JM, Pedrera Zamorano JD, Casado Naranjo I. Charlson comorbidity index in ischemic stroke and intracerebral hemorrhage as predictor of mortality and functional outcome after 6 months. *Journal of Stroke and Cerebrovascular Diseases*. 2013;22:e214-218
69. Redfors P, Isaksen D, Lappas G, Blomstrand C, Rosengren A, Jood K, et al. Living alone predicts mortality in patients with ischemic stroke before 70 years of age: A long-term prospective follow-up study. *BMC Neurology*. 2016;16:80
70. Hankey GJ. Stroke. *Lancet*. 2017;389:641-654
71. Garcia-Ptacek S, Contreras Escamez B, Zupanic E, Religa D, von Koch L, Johnell K, et al. Prestroke Mobility and Dementia as Predictors of Stroke Outcomes in Patients Over 65 Years of Age: A Cohort Study From The Swedish Dementia and Stroke Registries. *Journal of the American Medical Directors Association*. 2018;19:154-161
72. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics*. 2001;69:89-95
73. De Marchis GM, Katan M, Weck A, Fluri F, Foerch C, Findling O, et al. Copeptin adds prognostic information after ischemic stroke: Results from the CoRisk study. *Neurology*. 2013;80:1278-1286
74. Rost NS, Biffi A, Cloonan L, Chorba J, Kelly P, Greer D, et al. Brain natriuretic peptide predicts functional outcome in ischemic stroke. *Stroke*. 2012;43:441-445
75. Winbeck K, Poppert H, Etgen T, Conrad B, Sander D. Prognostic relevance of early serial C-reactive protein measurements after first ischemic stroke. *Stroke*. 2002;33:2459-2464

76. Tian Y, Jia H, Li S, Wu Y, Guo L, Tan G, et al. The associations of stroke, transient ischemic attack, and/or stroke-related recurrent vascular events with lipoprotein-associated phospholipase A2: A systematic review and meta-analysis. *Medicine*. 2017;96:e9413
77. Bustamante A, Simats A, Vilar-Bergua A, Garcia-Berrocoso T, Montaner J. Blood/brain biomarkers of inflammation after stroke and their association with outcome: From C-reactive protein to damage-associated molecular patterns. *Neurotherapeutics*. 2016;13:671-684
78. Samson AL, Medcalf RL. Tissue-type plasminogen activator: A multifaceted modulator of neurotransmission and synaptic plasticity. *Neuron*. 2006;50:673-678
79. Stott DJ, Robertson M, Rumley A, Welsh P, Sattar N, Packard CJ, et al. Activation of hemostasis and decline in cognitive function in older people. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 2010;30:605-611
80. Gillett SR, McClure LA, Callas PW, Thacker EL, Unverzagt FW, Wadley VG, et al. Hemostasis biomarkers and incident cognitive impairment: The REGARDS study. *Journal of Thrombosis and Haemostasis*. 2018;16:1259-1267
81. Gallacher J, Bayer A, Lowe G, Fish M, Pickering J, Pedro S, et al. Is sticky blood bad for the brain?: Hemostatic and inflammatory systems and dementia in the caerphilly prospective study. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 2010;30:599-604
82. Quinn TJ, Gallacher J, Deary IJ, Lowe GD, Fenton C, Stott DJ. Association between circulating hemostatic measures and dementia or cognitive impairment: Systematic review and meta-analyzes. *Journal of Thrombosis and Haemostasis*. 2011;9:1475-1482
83. Danesh J, Lewington S, Thompson SG, Lowe GD, Collins R, Kostis JB, et al. Plasma fibrinogen level and the risk of major cardiovascular diseases and nonvascular mortality: An individual participant meta-analysis. *Jama*. 2005;294:1799-1809
84. Tzoulaki I, Murray GD, Lee AJ, Rumley A, Lowe GD, Fowkes FG. Relative value of inflammatory, hemostatic, and rheological factors for incident myocardial infarction and stroke: The Edinburgh artery study. *Circulation*. 2007;115:2119-2127
85. Whiteley W, Jackson C, Lewis S, Lowe G, Rumley A, Sandercock P, et al. Inflammatory markers and poor outcome after stroke: A prospective cohort study and systematic review of interleukin-6. *PLoS medicine*. 2009;6:e1000145
86. Rist PM, Buring JE, Kase CS, Ridker PM, Kurth T. Biomarkers and functional outcomes from ischaemic cerebral events in women: A prospective cohort study. *European Journal of Neurology*. 2013;20:375-381

87. Welsh P, Barber M, Langhorne P, Rumley A, Lowe GD, Stott DJ. Associations of inflammatory and haemostatic biomarkers with poor outcome in acute ischaemic stroke. *Cerebrovascular Diseases*. 2009;27:247-253
88. del Zoppo GJ, Levy DE, Wasiewski WW, Pancioli AM, Demchuk AM, Trammel J, et al. Hyperfibrinogenemia and functional outcome from acute ischemic stroke. *Stroke.* 2009;40:1687-1691
89. Whiteley W, Wardlaw J, Dennis M, Lowe G, Rumley A, Sattar N, et al. The use of blood biomarkers to predict poor outcome after acute transient ischemic attack or ischemic stroke. *Stroke*. 2012;43:86-91
90. Turaj W, Slowik A, Dziedzic T, Pulyk R, Adamski M, Strojny J, et al. Increased plasma fibrinogen predicts one-year mortality in patients with acute ischemic stroke. *Journal of the Neurological Sciences*. 2006;246:13-19
91. Whiteley W, Jackson C, Lewis S, Lowe G, Rumley A, Sandercock P, et al. Association of circulating inflammatory markers with recurrent vascular events after stroke: A prospective cohort study. *Stroke*. 2011;42:10-16
92. Di Napoli M, Papa F. Inflammation, hemostatic markers, and antithrombotic agents in relation to long-term risk of new cardiovascular events in first-ever ischemic stroke patients. *Stroke*. 2002;33:1763-1771
93. Carter AM, Catto AJ, Mansfield MW, Bamford JM, Grant PJ. Predictive variables for mortality after acute ischemic stroke. *Stroke.* 2007;38:1873-1880
94. Rothwell PM, Howard SC, Power DA, Gutnikov SA, Algra A, van Gijn J, et al. Fibrinogen concentration and risk of ischemic stroke and acute coronary events in 5113 patients with transient ischemic attack and minor ischemic stroke. *Stroke*. 2004;35:2300-2305
95. Resch KL, Ernst E, Matrai A, Paulsen HF. Fibrinogen and viscosity as risk factors for subsequent cardiovascular events in stroke survivors. *Annals of Internal Medicine*. 1992;117:371-375
96. Ryu JK, Davalos D, Akassoglou K. Fibrinogen signal transduction in the nervous system. *Journal of Thrombosis and Haemostasis*. 2009;7 Suppl 1:151-154
97. Schachtrup C, Ryu JK, Helmrick MJ, Vagena E, Galanakis DK, Degen JL, et al. Fibrinogen triggers astrocyte scar formation by promoting the availability of active tgf-beta after vascular damage. *The Journal of neuroscience*. 2010;30:5843-5854
98. Petersen MA, Ryu JK, Chang KJ, Etxeberria A, Bardehle S, Mendiola AS, et al. Fibrinogen activates BMP signaling in oligodendrocyte progenitor cells and inhibits remyelination after vascular damage. *Neuron*. 2017;96:1003-1012.e1007

99. Ahn HJ, Zamolodchikov D, Cortes-Canteli M, Norris EH, Glickman JF, Strickland S. Alzheimer's disease peptide beta-amyloid interacts with fibrinogen and induces its oligomerization. *Proceedings of the National Academy of Sciences of the United States of America*. 2010;107:21812-21817
100. van Oijen M, Witteman JC, Hofman A, Koudstaal PJ, Breteler MM. Fibrinogen is associated with an increased risk of Alzheimer disease and vascular dementia. *Stroke*. 2005;36:2637-2641
101. Marioni RE, Stewart MC, Murray GD, Deary IJ, Fowkes FG, Lowe GD, et al. Peripheral levels of fibrinogen, C-reactive protein, and plasma viscosity predict future cognitive decline in individuals without dementia. *Psychosomatic Medicine*. 2009;71:901-906
102. Bowie EJ, Solberg LA, Jr., Fass DN, Johnson CM, Knutson GJ, Stewart ML, et al. Transplantation of normal bone marrow into a pig with severe von Willebrand's disease. *The Journal of Clinical Investigation*. 1986;78:26-30
103. Wu MD, Atkinson TM, Lindner JR. Platelets and von Willebrand factor in atherogenesis. *Blood*. 2017;129:1415-1419
104. Spiel AO, Gilbert JC, Jilma B. Von Willebrand factor in cardiovascular disease: Focus on acute coronary syndromes. *Circulation*. 2008;117:1449-1459
105. Wennberg P, Wensley F, Di Angelantonio E, Johansson L, Boman K, Rumley A, et al. Haemostatic and inflammatory markers are independently associated with myocardial infarction in men and women. *Thrombosis Research*. 2012;129:68-73
106. Greisenegger S, Segal HC, Burgess AI, Poole DL, Mehta Z, Rothwell PM. Biomarkers and mortality after transient ischemic attack and minor ischemic stroke: Population-based study. *Stroke*. 2015;46:659-666
107. Williams SR, Hsu FC, Keene KL, Chen WM, Dzhivhuho G, Rowles JL, 3rd, et al. Genetic drivers of von Willebrand factor levels in an ischemic stroke population and association with risk for recurrent stroke. *Stroke*. 2017;48:1444-1450
108. Zhao BQ, Chauhan AK, Canault M, Patten IS, Yang JJ, Dockal M, et al. Von Willebrand factor-cleaving protease ADAMTS13 reduces ischemic brain injury in experimental stroke. *Blood*. 2009;114:3329-3334
109. Zhu X, Cao Y, Wei L, Cai P, Xu H, Luo H, et al. Von Willebrand factor contributes to poor outcome in a mouse model of intracerebral haemorrhage. *Scientific Reports*. 2016;6:35901
110. Ridker PM, Hennekens CH, Stampfer MJ, Manson JE, Vaughan DE. Prospective study of endogenous tissue plasminogen activator and risk of stroke. *Lancet*. 1994;343:940-943

111. Willeit P, Thompson A, Aspelund T, Rumley A, Eiriksdottir G, Lowe G, et al. Hemostatic factors and risk of coronary heart disease in general populations: New prospective study and updated meta-analyses. *PloS one*. 2013;8:e55175
112. Zeng L, Liu J, Wang Y, Wang L, Weng S, Chen S, et al. Cocktail blood biomarkers: Prediction of clinical outcomes in patients with acute ischemic stroke. *European Neurology*. 2013;69:68-75
113. Woodward M, Lowe GD, Campbell DJ, Colman S, Rumley A, Chalmers J, et al. Associations of inflammatory and hemostatic variables with the risk of recurrent stroke. *Stroke*. 2005;36:2143-2147
114. Medcalf RL. Fibrinolysis: From blood to the brain. *Journal of Thrombosis and Haemostasis*. 2017;15:2089-2098
115. Leebeek FW, Goor MP, Guimaraes AH, Brouwers GJ, Maat MP, Dippel DW, et al. High functional levels of thrombin-activatable fibrinolysis inhibitor are associated with an increased risk of first ischemic stroke. *Journal of Thrombosis and Haemostasis*. 2005;3:2211-2218
116. Brouns R, Heylen E, Willemse JL, Sheorajpanday R, De Surgeloose D, Verkerk R, et al. The decrease in procarboxypeptidase U (TAFI) concentration in acute ischemic stroke correlates with stroke severity, evolution and outcome. *Journal of Thrombosis and Haemostasis*. 2010;8:75-80
117. Denorme F, Wyseure T, Peeters M, Vandeputte N, Gils A, Deckmyn H, et al. Inhibition of thrombin-activatable fibrinolysis inhibitor and plasminogen activator inhibitor-1 reduces ischemic brain damage in mice. *Stroke*. 2016;47:2419-2422
118. Jood K, Redfors P, Gils A, Blomstrand C, Declerck PJ, Jern C. Convalescent plasma levels of TAFI activation peptide predict death and recurrent vascular events in ischemic stroke survivors. *Journal of Thrombosis and Haemostasis*. 2012;10:725-727
119. Khalil M, Teunissen CE, Otto M, Piehl F, Sormani MP, Gattringer T, et al. Neurofilaments as biomarkers in neurological disorders. *Nature Reviews. Neurology*. 2018
120. Yilmaz A, Blennow K, Hagberg L, Nilsson S, Price RW, Schouten J, et al. Neurofilament light chain protein as a marker of neuronal injury: Review of its use in HIV-1 infection and reference values for HIV-negative controls. *Expert Review of Molecular Diagnostics*. 2017;17:761-770
121. Disanto G, Barro C, Benkert P, Naegelin Y, Schadelin S, Giardiello A, et al. Serum neurofilament light: A biomarker of neuronal damage in multiple sclerosis. *Annals of Neurology*. 2017;81:857-870

122. Kuhle J, Barro C, Andreasson U, Derfuss T, Lindberg R, Sandelius A, et al. Comparison of three analytical platforms for quantification of the neurofilament light chain in blood samples: ELISA, electrochemiluminescence immunoassay and Simoa. *Clinical Chemistry and Laboratory Medicine*. 2016;54:1655-1661
123. Gisslen M, Price RW, Andreasson U, Norgren N, Nilsson S, Hagberg L, et al. Plasma Concentration of the Neurofilament Light Protein (NFL) is a Biomarker of CNS injury in HIV infection: A Cross-Sectional Study. *EBioMedicine*. 2016;3:135-140
124. Lu CH, Macdonald-Wallis C, Gray E, Pearce N, Petzold A, Norgren N, et al. Neurofilament light chain: A prognostic biomarker in amyotrophic lateral sclerosis. *Neurology*. 2015;84:2247-2257
125. Rohrer JD, Woollacott IO, Dick KM, Brotherhood E, Gordon E, Fellows A, et al. Serum neurofilament light chain protein is a measure of disease intensity in frontotemporal dementia. *Neurology*. 2016;87:1329-1336
126. Steinacker P, Huss A, Mayer B, Grehl T, Grosskreutz J, Borck G, et al. Diagnostic and prognostic significance of neurofilament light chain NF-L, but not progranulin and S100B, in the course of amyotrophic lateral sclerosis: Data from the German MND-net. *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration*. 2017;18:112-119
127. Shahim P, Zetterberg H, Tegner Y, Blennow K. Serum neurofilament light as a biomarker for mild traumatic brain injury in contact sports. *Neurology*. 2017;88:1788-1794
128. Tiedt S, Duering M, Barro C, Kaya AG, Boeck J, Bode FJ, et al. Serum neurofilament light: A biomarker of neuroaxonal injury after ischemic stroke. *Neurology*. 2018;91:e1338-e1347
129. Gattringer T, Pinter D, Enzinger C, Seifert-Held T, Kneihsl M, Fandler S, et al. Serum neurofilament light is sensitive to active cerebral small vessel disease. *Neurology*. 2017
130. Traenka C, Disanto G, Seiffge DJ, Gensicke H, Hert L, Grond-Ginsbach C, et al. Serum neurofilament light chain levels are associated with clinical characteristics and outcome in patients with cervical artery dissection. *Cerebrovascular Diseases*. 2015;40:222-227
131. De Marchis GM, Katan M, Barro C, Fladt J, Traenka C, Seiffge DJ, et al. Serum neurofilament light chain in patients with acute cerebrovascular events. *European Journal of Neurology*. 2017
132. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*. 1953;171:737-738
133. Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, Harrow J, et al. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Human Molecular Genetics*. 2014;23:5866-5878

134. Finishing the euchromatic sequence of the human genome. *Nature*. 2004;431:931-945
135. Mattick JS, Taft RJ, Faulkner GJ. A global view of genomic information--moving beyond the gene and the master regulator. *Trends in Genetics*. 2010;26:21-28
136. Maurano MT, Haugen E, Sandstrom R, Vierstra J, Shafer A, Kaul R, et al. Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. *Nature Genetics*. 2015;47:1393-1401
137. Bevan S, Traylor M, Adib-Samii P, Malik R, Paul NLM, Jackson C, et al. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. *Stroke*. 2012;43:3161-3167.
138. Flossmann E, Schulz UG, Rothwell PM. Systematic review of methods and results of studies of the genetic epidemiology of ischemic stroke. *Stroke*. 2004;35:212-227
139. Seshadri S, Beiser A, Pikula A, Himali JJ, Kelly-Hayes M, Debette S, et al. Parental occurrence of stroke and risk of stroke in their children: The Framingham study. *Circulation*. 2010;121:1304-1312
140. Kasiman K, Lundholm C, Sandin S, Malki N, Sparen P, Ingelsson E. Familial effects on ischemic stroke: The role of sibling kinship, sex, and age of onset. *Circulation. Cardiovascular Genetics*. 2012;5:226-233
141. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nature Genetics*. 2018;50:524-537
142. Sim J, Teece L, Dennis MS, Roffe C. Validation and recalibration of two multivariable prognostic models for survival and independence in acute stroke. *PloS one*. 2016;11:e0153527
143. McAllister TW. Genetic factors modulating outcome after neurotrauma. *PM & R*. 2010;2:S241-252
144. Pearson-Fuhrhop KM, Burke E, Cramer SC. The influence of genetic factors on brain plasticity and recovery after neural injury. *Curr Opin Neurol*. 2012;25:682-688
145. Goldberg A, Curtis CL, Kleim JA. Linking genes to neurological clinical practice: The genomic basis for neurorehabilitation. *Journal of Neurologic Physical Therapy*. 2015;39:52-61
146. Zhao J, Wu H, Zheng L, Weng Y, Mo Y. Brain-derived neurotrophic factor G196A polymorphism predicts 90-day outcome of ischemic stroke in Chinese: A novel finding. *Brain Research*. 2013;1537:312-318
147. Stanne TM, Tjarnlund-Wolf A, Olsson S, Jood K, Blomstrand C, Jern C. Genetic variation at the BDNF locus: Evidence for association with long-term outcome after ischemic stroke. *PloS one*. 2014;9:e114156

148. Maguire J, Thakkinstian A, Levi C, Lincz L, Bisset L, Sturm J, et al. Impact of COX-2 rs5275 and rs20417 and GPIIIA rs5918 polymorphisms on 90-day ischemic stroke functional outcome: A novel finding. *Journal of Stroke and Cerebrovascular Diseases.* 2011;20:134-144
149. Lindgren A, Maguire J. Stroke recovery genetics. *Stroke*. 2016;47:2427-2434
150. Mola-Caminal M, Carrera C, Soriano-Tarraga C, Giralt-Steinhauer E, Diaz-Navarro RM, Tur S, et al. PATJ low frequency variants are associated with worse ischemic stroke functional outcome. *Circulation Research*. 2019;124:114-120
151. Pfeiffer D, Chen B, Schlicht K, Ginsbach P, Abboud S, Bersano A, et al. Genetic imbalance is associated with functional outcome after ischemic stroke. *Stroke*. 2019:Strokeaha118021856
152. Wilhelmsen L, Johansson S, Rosengren A, Wallin I, Dotevall A, Lappas G. Risk factors for cardiovascular disease during the period 1985-1995 in Goteborg, Sweden. The GOT-MONICA project. *Journal of Internal Medicine*. 1997;242:199-211
153. Multicenter trial of hemodilution in ischemic stroke--background and study protocol. Scandinavian stroke study group. *Stroke*. 1985;16:885-890
154. Williams LS, Yilmaz EY, Lopez-Yunez AM. Retrospective assessment of initial stroke severity with the NIH Stroke Scale. *Stroke*. 2000;31:858-862
155. Gray LJ, Ali M, Lyden PD, Bath PM. Interconversion of the National Institutes of Health Stroke Scale and Scandinavian Stroke Scale in acute stroke. *Journal of Stroke and Cerebrovascular Diseases*. 2009;18:466-468
156. Olsson S, Holmegaard L, Jood K, Sjogren M, Engstrom G, Lovkvist H, et al. Genetic variation within the interleukin-1 gene cluster and ischemic stroke. *Stroke*. 2012;43:2278-2282
157. Arsava EM, Ballabio E, Benner T, Cole JW, Delgado-Martinez MP, Dichgans M, et al. The Causative Classification of Stroke system: An international reliability and optimization study. *Neurology*. 2010;75:1277-1284
158. Prigatano GP. Bni screen for higher cerebral functions: Rationale and initial validation. *BNI Quart*. 1991:2-9
159. Jern C, Blomstrand C, Westerlind A. Evidence of a net release of tissue-type plasminogen activator across the human cerebral vasculature. *Thrombosis and Haemostasis*. 2004;91:1019-1025
160. Hanson E, Jood K, Karlsson S, Nilsson S, Blomstrand C, Jern C. Plasma levels of von Willebrand factor in the etiologic subtypes of ischemic stroke. *Journal of Thrombosis and Haemostasis*. 2011;9:275-281

161. Ladenvall C, Gils A, Jood K, Blomstrand C, Declerck PJ, Jern C. Thrombin activatable fibrinolysis inhibitor activation peptide shows association with all major subtypes of ischemic stroke and with TAFI gene variation. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 2007;27:955-962
162. Jood K, Danielson J, Ladenvall C, Blomstrand C, Jern C. Fibrinogen gene variation and ischemic stroke. *Journal of Thrombosis and Haemostasis*. 2008;6:897-904
163. Rissin DM, Kan CW, Campbell TG, Howes SC, Fournier DR, Song L, et al. Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations. *Nature Biotechnology*. 2010;28:595-599
164. Maguire JM, Bevan S, Stanne TM, Lorenzen E, Fernandez-Cadenas I, Hankey GJ, et al. GISCOME – Genetics of Ischaemic Stroke Functional Outcome network: A protocol for an international multicentre genetic association study. *European Stroke Journal*. 2017;2:229-237
165. Eriksson M, Appelros P, Norrving B, Terent A, Stegmayr B. Assessment of functional outcome in a national quality register for acute stroke: Can simple self-reported items be transformed into the modified Rankin Scale? *Stroke*. 2007;38:1384-1386
166. Human genomics. The genotype-tissue expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*. 2015;348:648-660
167. Eicher JD, Landowski C, Stackhouse B, Sloan A, Chen W, Jensen N, et al. GRASP v2.0: An update on the genome-wide repository of associations between snps and phenotypes. *Nucleic Acids Research*. 2015;43:D799-804
168. Leslie R, O'Donnell CJ, Johnson AD. Grasp: Analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. *Bioinformatics*. 2014;30:i185-194
169. Higasa K, Miyake N, Yoshimura J, Okamura K, Niihori T, Saitsu H, et al. Human genetic variation database, a reference database of genetic variations in the Japanese population. *Journal of Human Genetics*. 2016;61:547-553
170. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nature Genetics*. 2017;49:131-138
171. Grambsch P, Therneau T. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. 1994;81:515-526
172. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the c-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine*. 2011;30:1105-1117

173. Mishra A, Macgregor S. VEGAS2: Software for more flexible gene-based testing. *Twin Research and Human Genetics*. 2015;18:86-91
174. Zheng Q, Zhu D, Bai Y, Wu Y, Jia J, Hu Y. Exercise improves recovery after ischemic brain injury by inducing the expression of Angiopoietin-1 and Tie-2 in rats. *The Tohoku Journal of Experimental Medicine*. 2011;224:221-228
175. Ji H, Miao J, Zhang X, Du Y, Liu H, Li S, et al. Inhibition of sonic hedgehog signaling aggravates brain damage associated with the down-regulation of Gli1, Ptch1 and SOD1 expression in acute ischemic stroke. *Neuroscience Letters*. 2012;506:1-6
176. Hoang S, Liauw J, Choi M, Choi M, Guzman RG, Steinberg GK. Netrin-4 enhances angiogenesis and neurologic outcome after cerebral ischemia. *Journal of Cerebral Blood Flow and Metabolism.* 2009;29:385-397
177. Ernst E, Koenig W. Fibrinogen and cardiovascular risk. *Vascular medicine*. 1997;2:115-125
178. Johansson L, Jansson JH, Boman K, Nilsson TK, Stegmayr B, Hallmans G. Tissue Plasminogen Activator, Plasminogen Activator Inhibitor-1, and Tissue Plasminogen Activator/Plasminogen Activator Inhibitor-1 Complex as Risk Factors for the Development of a first Stroke. *Stroke*. 2000;31:26-32
179. Williams SR, Hsu FC, Keene KL, Chen WM, Nelson S, Southerland AM, et al. Shared genetic susceptibility of vascular-related biomarkers with ischemic and recurrent stroke. *Neurology*. 2016;86:351-359
180. Sabater-Lleal M, Huang J, Chasman D, Naitza S, Dehghan A, Johnson AD, et al. Multiethnic meta-analysis of genome-wide association studies in >100 000 subjects identifies 23 fibrinogen-associated loci but no strong evidence of a causal association between circulating fibrinogen and cardiovascular disease. *Circulation*. 2013;128:1310-1324
181. Datta A, Chen CP, Sze SK. Discovery of prognostic biomarker candidates of lacunar infarction by quantitative proteomics of microvesicles enriched plasma. *PloS one*. 2014;9:e94663
182. Narasimhalu K, Lee J, Leong YL, Ma L, De Silva DA, Wong MC, et al. Inflammatory markers and their association with post stroke cognitive decline. *International Journal of Stroke*. 2015;10:513-518
183. Luciano M, Marioni RE, Gow AJ, Starr JM, Deary IJ. Reverse causation in the association between C-reactive protein and fibrinogen levels and cognitive abilities in an aging sample. *Psychosomatic Medicine*. 2009;71:404-409
184. Kuhle J, Gaiottino J, Leppert D, Petzold A, Bestwick JP, Malaspina A, et al. Serum neurofilament light chain is a biomarker of human spinal cord injury severity and outcome. *Journal of Neurology, Neurosurgery, and Psychiatry*. 2015;86:273-279

185. Singh P, Yan J, Hull R, Read S, O'Sullivan J, Henderson RD, et al. Levels of phosphorylated axonal neurofilament subunit h (pnfh) are increased in acute ischemic stroke. *Journal of the Neurological Sciences*. 2011;304:117-121
186. Pujol-Calderon F, Portelius E, Zetterberg H, Blennow K, Rosengren LE, Hoglund K. Neurofilament changes in serum and cerebrospinal fluid after acute ischemic stroke. *Neuroscience Letters*. 2018;698:58-63
187. Buss A, Brook GA, Kakulas B, Martin D, Franzen R, Schoenen J, et al. Gradual loss of myelin and formation of an astrocytic scar during Wallerian degeneration in the human spinal cord. *Brain*. 2004;127:34-44
188. Thomalla G, Glauche V, Koch MA, Beaulieu C, Weiller C, Rother J. Diffusion tensor imaging detects early Wallerian degeneration of the pyramidal tract after ischemic stroke. *NeuroImage*. 2004;22:1767-1774
189. Heroes E, Lesage B, Gornemann J, Beullens M, Van Meervelt L, Bollen M. The PP1 binding code: A molecular-lego strategy that governs specificity. *The FEBS journal*. 2013;280:584-595
190. Genoux D, Haditsch U, Knobloch M, Michalon A, Storm D, Mansuy IM. Protein phosphatase 1 is a molecular constraint on learning and memory. *Nature*. 2002;418:970-975
191. Shioda N, Fukunaga K. Physiological and pathological roles of CaMKII-PP1 signaling in the brain. *International Journal of Molecular Sciences*. 2017;19
192. Gibson G. Rare and common variants: Twenty arguments. *Nature Reviews. Genetics*. 2012;13:135-145
193. Guimaraes AH, van Tilburg NH, Vos HL, Bertina RM, Rijken DC. Association between thrombin activatable fibrinolysis inhibitor genotype and levels in plasma: Comparison of different assays. *British Journal of Haematology*. 2004;124:659-665
194. Ceresa E, Brouwers E, Peeters M, Jern C, Declerck PJ, Gils A. Development of ELISAs measuring the extent of TAFI activation. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 2006;26:423-428
195. Ranby M, Nguyen G, Scarabin PY, Samama M. Immunoreactivity of tissue plasminogen activator and of its inhibitor complexes. Biochemical and multicenter validation of a two site immunosorbent assay. *Thrombosis and Haemostasis*. 1989;61:409-414
196. Appelros P, Stegmayr B, Terent A. Sex differences in stroke epidemiology: A systematic review. *Stroke*. 2009;40:1082-1090
197. Stegmayr B, Asplund K. Measuring stroke in the population: Quality of routine statistics in comparison with a population-based stroke registry. *Neuroepidemiology*. 1992;11:204-213