Aus dem Centrum für Schlaganfallforschung
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

The role of hypercoagulability in neurovascular disorders: Insights from clinical epidemiology

Die Rolle der Hyperkoagulabilität bei neurovaskulären Erkrankungen: Einblicke aus der klinischen Epidemiologie

zur Erlangung des akademischen Grades
Doctor of Philosophy (PhD)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Jessica L. Rohmann

aus Milwaukee, Wisconsin, USA

Datum der Promotion: 25.11.2022

Table of Contents

1. List of abbreviations

3MSE	Modified Mini-Mental State Examination
ANOVA	Analysis of Variance
CHS	Cardiovascular Health Study
CI	Confidence Interval
cOR	Common (ordinal) odds ratio
DSST	Digit Symbol Substitution Test
F11	Gene encoding Factor XI
F12	Gene encoding Factor XII
FVIII	Factor VIII
FXI	Factor XI
FXII	Factor XII
HR	Hazard ratio
HMWK	High molecular weight kininogen
IQR	Interquartile Range
KLKB1	Gene encoding prekallikrein
KNG1	Gene encoding high molecular weight kininogen
MRI	Magnetic resonance imaging
OR	Odds ratio
PK	Prekallikrein
PROSCIS-B	Prospective Cohort with Incident Stroke - Berlin (Study)
RATIO	Risk on Arterial Thrombosis In Oral contraception (Study)
SD	Standard deviation
SE	Standard error
SNP	Single nucleotide polymorphism
TIA	Transient ischemic attack
WMH	White Matter Hyperintensities

2. Abstract

Given their crucial role in preserving the hemostatic balance, coagulation factors are of great interest in the context of both venous and arterial thrombosis. Recent laboratory work has indicated that a hypercoagulable state increases risk for ischemic events. Furthermore, first clinical trials have found that temporarily lowering Factor XI levels prevents clots without increasing bleeds in knee arthroplasty patients, which may make these factors attractive therapeutic targets in secondary prevention of ischemic stroke.

My dissertation subprojects aimed to study the role of coagulation factors in neurovascular disease using existing observational datasets. As such, these clinical epidemiologic investigations provided important first insights in a cost-effective way. My work describes diverse aspects of the processes by which coagulation factors are implicated in neurovascular diseases, with each subproject characterized by a different study design and population.

We identified genetic determinants of contact factor levels (High Molecular Weight Kininogen (HMWK), Prekallikrein (PK), FXI and FXII) and probed their associations with vascular disease phenotypes in a case-control study. In addition to replicating known associations between single genetic variants and contact system factor levels, we identified two novel loci; one for PK antigen levels (*KLKB1* rs4253243; $\beta_{conditional}$=-12.38; 95% confidence interval (CI), -20.07 to -4.69) and one for HMWK antigen levels (*KNG1* rs5029980; $\beta_{conditional}$=5.86; 95%CI: 2.40 to 9.32).

We estimated the effects of hypercoagulability on post-stroke outcomes in a cohort of 576 ischemic stroke patients. After controlling for confounding, compared with having low or normal levels, having high (>75th-percentile) FXI activity levels increased the hazard for the combined endpoint (recurrent stroke, myocardial infarction, or all-cause mortality) within three years of first ischemic stroke (Hazard Ratio (HR)=1.80, 95%CI: 1.09–2.98). High FVIII activity was also linked to worse outcomes (HR=2.05, 95%CI: 1.28–3.29), whereas high FXII activity was not (HR=0.86, 95%CI: 0.49–1.51).

Finally, in a general population cohort of older persons, we found no evidence of a relevant contribution of factor VIII activity to the presence or worsening of white matter hyperintensities or cognitive performance over time.

Each of these subprojects provided a substantial contribution to filling knowledge gaps in the field of hypercoagulability research. Taken together, this dissertation work contributes to a better understanding of genetic influences on coagulation factor levels as well as the longer-term effects of expressed hypercoagulability on outcomes among stroke patients and in a general population sample of older persons. My work concludes with a set of suggestions for future coagulation factor research, especially in the context of stroke, based on lessons learned during the process of synthesizing and critically evaluating my results.

3. Deutsche Zusammenfassung

Aufgrund ihrer entscheidenden Rolle bei der Aufrechterhaltung der hämostatischen Balance, sind Gerinnungsfaktoren im Zusammenhang mit venösen sowie arteriellen Thrombosen von großem Interesse. Neuere Laborergebnisse zeigten ein erhöhtes Risiko für ischämische Ereignisse durch einen hyperkoagulablen Zustand. Erste klinische Studien zeigten eine vielversprechende Wirkung gegen Thrombosen durch die temporäre Senkung des Gerinnungsfaktor-XI-Spiegels – ohne Erhöhung des Blutungsrisikos bei Knieendoprothese-Patient*innen. Diese Erkenntnis ist auch im Kontext der Sekundärprävention von Schlaganfällen von hoher Relevanz.

Das Ziel meiner Dissertation war, in drei Teilprojekten die Rolle bestimmter Gerinnungsfaktoren zu untersuchen, die zur Hyperkoagulabilität bei neurovaskulären Erkrankungen beitragen. Die auf klinisch-epidemiologischen Fragestellungen basierenden Sekundäranalysen von Beobachtungsdatensätzen liefern wichtige Erkenntnisse für die Zielpopulationen in kosten-effektiver Weise.

Im Rahmen meiner Dissertation wurden genetische Determinanten von Gerinnungsfaktoren des Contact-Activation-Systems (hochmolekulares Kininogen (HMWK), Präkallikrein (PK), Faktor XI und Faktor XII) identifiziert. Außerdem untersuchten wir ihre Assoziationen mit Phänotypen von vaskulären Erkrankungen in einer Fall-Kontroll-Studie. Zusätzlich zur Replikation bekannter Assoziationen zwischen einzelnen genetischen Varianten und den Spiegeln der Gerinnungsfaktoren identifizierten wir zwei neue Genloci: einen für den PK-Antigenspiegel (*KLKB1* rs4253243; $\beta_{adjustiert}$=-12,38; 95% Konfidenzintervall (KI): -20,07 bis -4,69) und einen für den HMWK-Antigenspiegel (*KNG1* rs5029980; $\beta_{adjustiert}$=5,86; 95%KI, 2,40 bis 9,32).

Des Weiteren schätzten wir die Effekte von hohen Faktor-XI-, Faktor-XII- und Faktor-VIII-Aktivitätsspiegeln auf Langzeit-Outcomes in einer Kohorte von 576 Patient*innen mit ischämischem Schlaganfall. Im Vergleich zu niedrigen oder normalen Werten hatten Schlaganfallpatient*innen mit einer hohen FXI-Aktivität (>75. Perzentil) ein höheres Risiko für den kombinierten Endpunkt (rezidivierender Schlaganfall, Myokardinfarkt oder Gesamtmortalität) innerhalb von drei Jahren (HR=1,80; 95%KI: 1,09 bis 2,98) nach Confounding-Adjustierung. Eine hohe FVIII-Aktivität war ebenfalls mit einem

schlechteren Outcome verbunden (HR=2,05; 95%KI: 1,28 bis 3,29), eine hohe FXII-Aktivität hingegen nicht (HR=0,86; 95%KI: 0,49 bis 1,51).

Schließlich fanden wir in einer Allgemeinbevölkerungskohorte älterer Personen keine Hinweise auf einen relevanten Beitrag des Faktor-VIII-Aktivitätsspiegels zur Präsenz von White Matter Hyperintensities oder verminderten kognitiven Funktionen und deren Verschlechterung im Zeitverlauf.

Meine Arbeit schließt mit praktischen und methodischen Vorschlägen für zukünftige Forschung zu Gerinnungsfaktoren, insbesondere im Kontext des ischämischen Schlaganfalls, basierend auf den Erkenntnissen aus der Synthese und kritischen Bewertung meiner Ergebnisse.

4. Introduction

4.1. Stroke, white matter hyperintensities and post-stroke outcomes

Stroke is a neurovascular disease with severe morbidity, high mortality, and substantial cost burden on healthcare systems.[1] Globally, 5.5 million deaths and 116.4 million disability-adjusted life years (DALYs) are attributed to stroke, according to 2016 Global Burden of Disease Study estimates.[1] As the second most common cause of death and the second-highest contributor to global DALYs,[1] gaining a better understanding not only of the underlying causes of stroke, but also its consequences is of utmost public health importance. Thanks to improved acute stroke care worldwide and perhaps most importantly, the introduction of specialized stroke-units, a steady decline in short-term acute stroke fatality has been observed in recent years, especially in high-income countries.[2,3]

Ischemic stroke is the most common stroke type, accounting for 84.4% of the global prevalent strokes; however, its DALYs are lower compared with the rarer hemorrhagic stroke type.[1] In addition to overt stroke events, undetected ischemic events may contribute to the development of covert white matter hyperintensities (WMH), which are highly prevalent among aging individuals and become more common and severe with increasing age.[4] Over the life course, the accumulation of small microbleeds and lesions, visible on magnetic resonance imaging (MRI) as WMH, contributes to vascular risk as well as risk for cognitive decline and dementia.[5]

Following the initial ischemic stroke event, the risk of having a further, overt vascular event, such as a secondary stroke or myocardial infarction, is elevated and sustained even in the longer-term.[6,7] Despite this increased risk, little is known about factors which may causally contribute to secondary events, and whether these differ from known causal risk factors for primary ischemic stroke. One known risk factor for first stroke and possible risk factor for second stroke is hypercoagulability, which is detailed in the next section.

4.2. Hypercoagulability and the hemostatic balance

An excess of prothrombotic factors and/or thrombophillic activities contribute to a so-called hypercoagulable state, or more generally termed, hypercoagulability.[8] Hypercoagulability is attributable to either genetic variation, an acquired condition (e.g., autoimmune diseases), or a combination of both factors.[9] Persons with hypercoagulable states are predisposed to developing several forms of arterial thrombotic disease including ischemic stroke.[8–10]

As the hemostatic balance to maintain fluidity while preventing bleeds is a delicate one, it is well known that even small shifts in this balance in the procoagulant or anticoagulant direction can result in serious pathologies.[11,12] Both of the "waterfall-like" enzymatic extrinsic and intrinsic pathways can activate Factor X to Xa, which is the first step in the "common" pathway that ends with the formation of a clot through the generation of fibrin strands that hold together the platelet plug.[13,14] The resulting clot can be life-saving, in the case of tissue repair, or life-threatening, in the case of thrombosis; Rudolf Virchow appreciated this delicate balance as early as the mid-1800s, and described the concept of hypercoagulability as one of three causal contributors to thrombosis in eponym now know as "Virchow's triad."[15]

Even within "normal" levels of coagulation factors, having above-average antigen and/or activity levels of certain coagulation factors may cause a shift of the hemostatic balance in the thrombotic direction, increasing risk for thrombotic events.[8,10] Though much of the clinical research investigating the relationship between elevated levels of coagulation factors and thrombotic events has focused on outcomes affecting venous circulation (e.g., deep vein thrombosis and pulmonary embolism), specific coagulation system factors, mostly from the intrinsic pathway, have also been implicated in thrombotic events impacting arterial circulation; the current state of scientific knowledge is summarized in the following section.

4.3. Coagulation factors: promising targets and first insights in the context of vascular disease

In this section, I summarize key scientific findings pertaining to several promising coagulation factors in the context of vascular disease, which I chose to focus on as

exposures of interest across my three dissertation subprojects: Factor XII (FXII), Factor XI (FXI), and Factor VIII (FVIII).

The coagulation system pathways are characterized by a cascade-like structure with many feedback loops and interdependencies. The activation of FXII, also known as the Hageman factor, kicks off the appropriately named “contact activation system” upon contact with a negatively charged surface.[16] Once activated, this first and most upstream plasma protein activates FXI, which subsequently sets a series of procoagulant processes of the intrinsic coagulation cascade into motion.[17,18]

In parallel to its activation of FXI, activated FXII also cleaves Prekallikrein (PK) to its activated form, Kallikrein, in a first step of the pro-inflammatory Kinin-Kallikrein peptide cascade.[19] As a part of yet another feedback loop, activated PK is involved in reciprocal transactivation with FXII,[20] before PK ultimately cleaves High Molecular Weight Kininogen (HMWK) into bradykinin.[19] The interconnectedness of this contact activation system is further evidenced by HMWK’s role as a membrane-anchoring cofactor necessary for the activation of both FXI and PK.[17,21]

Further downstream, the intrinsic system connects with the extrinsic (common) coagulation pathway via Factor VIII. FVIII is activated by thrombin, and is itself involved in the activation of Factor X, which results in downstream activation of more thrombin.[13,14] Ultimately, the generation of fibrin, through the cleavage of its precursor fibrinogen by thrombin, results in blood clot formation through polymerization and crosslinking.[13,14]

In the early 2000s, FXII was a particularly promising treatment target for investigation in animal studies, especially because the absence of FXII was observed to be protective against cerebral ischemia in the mouse model without jeopardizing the hemostatic balance in terms of any relevant increase in bleeding risk.[22,23] Like in the knock-out mouse model, humans with a rare inherited condition causing extreme FXII-deficiency were found to be protected against both venous thrombosis and ischemic stroke.[24,25] However, the initial excitement in the research community with respect to FXII as a promising future therapeutic target for vascular disease[26,27] was somewhat attenuated after observational studies in humans yielded inconsistent results, outlined below.

Though it was postulated that low FXII levels would be protective against thrombotic events, seminal research from the Study of Myocardial Infarctions Leiden (SMILE) found that in middle-aged men, low levels of FXII activity actually increased the risk of myocardial infarction.[28] Moreover, in a nested case-control study embedded in the Northwick Park Heart Study, having low levels of inhibitory complexes of FXIIa was associated with an increased risk for coronary heart disease and stroke (of ischemic or hemorrhagic nature) in middle-aged men.[29] Contrarily, an investigation of the same relationship found that high levels of FXII activation was associated with an increase in stroke risk (OR=2.1, 95%CI: 1.3-3.5) but not myocardial infarction (OR=0.82; 95%CI: 0.46 to 1.47) among young women in the Netherlands.[30] A more recent Swedish study found an association between FXII levels and hemorrhagic stroke risk but no relationship between FXII and ischemic stroke or myocardial infarction risk.[31] Taken together, these findings suggest the relationship between FXII levels and thrombosis is more complex than initially postulated, likely due to the numerous complex interdependencies of the contact activation system, and warrants further investigation.

Just downstream of FXII, FXI presented as another potential treatment target. FXI was particularly appealing because laboratory and animal studies showed that FXI's role in hemostasis is not critical; specifically, reducing FXI seemed to curb thrombotic clotting without increasing bleeding risk.[32–35] The question remained whether these findings from *in vitro* and *in vivo* (animal model) experiments were translatable to humans. First observational studies found a moderate association between having high FXI levels and risk for first ischemic stroke (OR=2.65, 95%CI: 1.27-5.56).[36] A systematic review by Maino et al. later clarified that hypercoagulability appears to be a stronger risk factor for ischemic stroke than for myocardial infarction.[10] In 2015, results from a first clinical trial confirmed that it was possible to reduce FXI levels using FXI antisense oligonucleotides in humans, and that this reduction prevented postoperative venous thromboembolism in knee arthroplasty patients.[37] Importantly, this intervention was deemed safe in terms of bleeding risk, showing that selectively inhibiting FXI does not impact hemostasis in the same way as conventional prophylactic treatments targeting Factor Xa or thrombin.[37]

High levels of FVIII, further downstream in the cascade, have been linked with an increased risk of first ischemic stroke and mortality in the general population.[38,39] Strong

associations between high FVIII and deep vein thrombosis (OR=4.8, 95%CI: 2.3-10.0) and recurrent venous thromboembolism (RR=6.7, 95%CI: 3.0-14.8) as well as a dose-response relationship have been observed,[40,41], but it is yet to be established to what extent higher levels of FVIII may contribute to the covert ischemia and the development of WMH.

Following an ischemic stroke event, FVIII levels are elevated. However, these increased levels are thought to persist beyond the initial acute phase, at least in some individuals.[42,43]. Though it remains unknown whether this increase may have a clinically-relevant impact on long-term post-stroke outcomes, a recent study suggested that concurrent elevation of FVIII and the von Willebrand Factor in the acute phase of ischemic stroke was linked to poorer outcomes in the very short-term, including worse functional outcome at hospital discharge (OR=2.87, 95%CI: 1.16-7.06), as well as higher odds of inpatient complications (OR=8.6, 95%CI: 1.58-46.85) and neuroworsening (OR=3.2, 95%CI: 1.18-8.73).[44]

To date, much of our knowledge about the potential role of coagulation factors in the context of ischemic disease is based on laboratory knowledge. Especially in the context of secondary events after first stroke and covert white matter hyperintensities development, a large knowledge gap exists in the literature. Therefore, in my dissertation work, I sought to generate clinically relevant, epidemiological research questions from laboratory insights and preliminary research to better understand causes and consequences of hypercoagulability in the context of neurovascular outcomes.

5. Objectives

The overarching goal of my PhD research was to investigate both determinants and consequences of a set of coagulation factors known to contribute to hypercoagulability, which are also relevant treatment targets in the context of ischemic vascular disease. The cumulative dissertation subprojects described in detail in the next chapters were designed to address three specific aims:

(1) To explore potential associations between genetic variants and activity and antigen levels of contact system coagulation factors as well as vascular outcomes (Publication 1).[45]

(2) To investigate the role of coagulation factors known to contribute to hypercoagulability measured in blood samples taken from patients shortly after first stroke on the occurrence of post-stroke endpoints (Publication 2).[46]

(3) To determine whether FVIII levels, a risk factor for overt thrombotic events, also contribute to the presence and worsening of small areas of tissue death in the brain (WMH) and cognitive decline (Publication 3).[47]

6. Methods

To address the objectives outlined in the previous chapter, my PhD research project involved the secondary use of existing data supplemented with additional laboratory measurements performed in stored blood samples from three large human studies. In this chapter, I describe these datasets, outline key study design elements, and highlight the analytical approaches that were employed to address the aforementioned objectives of this dissertation.

6.1.1. Dataset: Population-based Dutch case-control study, RATIO

To explore the potential role of genetic variation in determining expression of antigen and activity levels of the contact system factors (XII, XI, PK and HMWK) as well as vascular disease phenotypes, we used data from the Dutch Risk on Arterial Thrombosis

In Oral contraception (RATIO) study.[45] This population-based case-control study was originally designed to determine the relationship between oral contraceptive use and vascular outcomes among females aged 18 to 49 recruited from 16 locations in the Netherlands.[48,49] In this study, adult women aged less than 50 years who were hospitalized for an ischemic stroke or myocardial infarction between January 1990 and October 1995 were recruited from the participating centers as cases.[48,49] Data collected from the ischemic stroke and myocardial infarction patients in the RATIO study were relevant for our analyses of associations between single genetic variants and disease phenotypes.[45]

In the RATIO study, female participants from the areas surrounding the participating centers, representing the underlying target population that gave rise to the cases, were recruited using random digit dialing as controls.[48] The final control group reflected a population sample that was frequency-matched to the cases by index year, place of residence, and age group,[48] and it was this group that comprised our sample for the analysis of inherited genetic variants and contact system factor levels (traits).[45]

A standardized questionnaire was used to collect relevant demographic information from all participants upon enrollment.[48] Blood samples and buccal swabs were collected and stored during the second phase of the RATIO study, which occurred a median of 69 months after myocardial infarction or 95 months after ischemic stroke among patients, and at a comparable interval among the frequency-matched control participants.[36] Ethics committees of all collaborating hospitals approved the study, also allowing secondary data analysis projects, such as this one, and written informed consent was obtained from all RATIO study participants.[48]

6.1.2. RATIO: Single nucleotide polymorphism selection, genotyping and laboratory measurements

We were interested in capturing the common genetic variation across the genes encoding the contact activation system proteins of interest; namely, FXII (encoded by the *F12* gene on chromosome 5), FXI (*F11* on chromosome 4), PK (*KLKB1*, *kallikrein B1* on chromosome 4), and HMWK (*KNG1* on chromosome 3). Therefore, we used Haploview software (version 4.2) to select tag SNPs (single nucleotide polymorphisms)

having a minor allele frequency of at least 5% and an R^2 larger than 0.80 according to data from the HapMap Central European reference population.[45,50] Ultimately, we selected two *F12*, eleven *F11*, ten *KLKB1,* and twenty *KNG1* tagging SNPs for genotyping and further analysis[45]. The assay design failed for one SNP (rs5030072) in two different assays, so we were forced to exclude this SNP.[45]

In the RATIO study, antigen levels of FXII, FXI, PK and HMWK, relevant for our primary analysis of common genetic variation and coagulation factor levels (quantitative trait loci analyses), were quantified using sandwich-ELISA-based assays and are expressed as percentages of normal pooled plasma.[36,45,51] For one factor, FXI, activity measurements were available in addition to antigen levels. These activity levels were ascertained using a one-stage clotting assay and FXI-deficient plasma and are reported as percentages of activated pooled plasma.[45,52]

6.1.3. RATIO: Statistical analysis

For each SNP, we computed the major and minor allele frequencies among control participants.[45] We further determined the proportion of individuals for whom the corresponding SNP information is available, reported as the "call rates".[45]

The single SNP quantitative trait loci analyses were performed using age-adjusted linear regression models using data from control participants only.[45] In these models, we assumed that inheritance was additive; therefore, the reported β estimates (with corresponding 95% confidence intervals (CI)) can be interpreted as the change in the level of the corresponding trait per copy of the minor allele of that SNP.[45]

Since we employed statistical testing for discovery purposes, and we tested for statistical associations between the individual SNPs of the four genes (N=43) across the five traits, correction for multiple statistical hypothesis testing was warranted. Therefore, we imposed a global significance level of 0.00116 for all statistical tests, reflecting a conservative Bonferroni correction.[45] Sometimes, we obtained multiple statistically significant associations between SNPs within one gene and a given trait. In these instances, we performed a "conditional analysis", in which we further adjusted these regression models for the so-called lead SNP (i.e., the SNP whose coefficient had the smallest p-value across the single-SNP regression models for that gene).[45] Within each

gene, SNPs with associations that remained statistically significant were then included in a single model, and ultimately, only those still remaining statistically significant were carried forward and reported as “secondary signals”.[45]

To assess across-trait associations between a SNP from one gene and a different trait (i.e., not encoded by the gene of the SNP), we followed the same aforementioned procedure with additional adjustment for the levels of the trait encoded by the SNP’s gene.[45]

In secondary analyses, associations between all lead and secondary signal SNPs and the two disease phenotypes (myocardial infarction and ischemic stroke) were estimated using logistic regression models in the full RATIO study population including both cases and control participants.[45] Since both cases and controls were included, these models were conditioned on the set of variables used to frequency-match the participants, including age, place of residence, and index year.[45] All aforementioned single SNP analyses were conducted using IBM SPSS statistical software (version 23).

To quantify associations between inherited combinations of these variants and coagulation factor levels, we built “haplotypes” for each gene using the lead SNP and all secondary signals using the haplo.em function of the haplo.stats package using the open-source statistical software, R.[45,53] For all haplotypes occurring in more than 1% of control participants, we ran age-adjusted regression models for each of the five expressed outcome traits (factor levels) using the haplo.glm function,[45,53] with the haplotype containing major allele copies as the reference. For these analyses, we obtained β estimates, corresponding standard errors (SE), and p-values from the generalized linear models.[45] We employed logistic regression models to estimate the associations between these haplotypes and the two disease phenotypes in the full study population, adjusting for matching variables.[45]

6.2. Coagulation factor levels and post-stroke outcomes

6.2.1. Dataset: Berlin-based stroke patient cohort with long-term follow up; PROSCIS-B

To estimate the effects of high coagulation factor FXII, FXI and FVIII activity levels on vascular post-stroke outcomes, we used data from the longitudinal Prospective Cohort with Incident Stroke Berlin (PROSCIS-B) study (N=699).[46] Full details regarding eligibility criteria and participant recruitment in this hospital-based, longitudinal cohort study, which was originally conceptualized to develop a clinical risk prediction model, are detailed elsewhere.[54] To summarize, this patient cohort consisted of individuals aged 18 or older diagnosed with first-ever stroke at one of the three stroke units at the Charité - Universitätsmedizin Berlin between January 2010 and February 2013.[46,54] After participants or their legal representatives provided written informed consent, the study team conducted a detailed interview, performed a clinical examination, and collected blood samples within one week of the index event.[46,54]

Trained study nurses contacted participants in PROSCIS-B for structured telephone interviews (or sent a survey by postal mail, if the participant was not reachable by phone) once per year during the three-year follow-up period after stroke.[46,54] Among other parameters, information was obtained about vital status as well as the date of occurrence and type of any incident vascular events.[46,54] This information was volunteered by the participants themselves, with the assistance of a family member, or by their legal guardian.[46,54] We verified these endpoints and obtained information about additional events meeting the endpoint criteria through review of Charité hospital medical records or information provided by participants' general practitioners or treating hospital (if outside the Charité).[46,54] For participants who could not be reached, vital status information was obtained from the Berlin city registration office.[46,54] The Charité – Universitätsmedizin Berlin ethics committee granted approval for the PROSCIS-B study (EA1/218/09), which includes the research activities of this subproject.[46]

6.2.2. PROSCIS-B: Relevant outcome variables and eligibility criteria

In this study, our outcome of interest was a combined endpoint, defined as the first occurrence of one of the following events after participant inclusion: (1) stroke, (2) myocardial infarction, or (3) death due to any cause during follow-up.[46] An endpoint committee consisting of two independent vascular neurologists blinded to exposure status validated all vascular events occurring during the PROSCIS-B follow-up.[46,54] In addition to the eligibility criteria of the original PROSCIS-B,[54] for the purposes of this subproject, we excluded participants with index strokes not of ischemic nature (i.e., hemorrhagic stroke and sinus venous thrombosis), since these stroke types have distinct underlying etiologies, as well as severe strokes (participants with a National Institute of Health Stroke Scale score >15), which were very uncommon in the cohort, to limit heterogeneity of the study sample.[46]

6.2.3. PROSCIS-B: Coagulation factor activity measurements

In collaboration with the Leiden University Medical Center, we assayed FXII, FXI and FVIII activity levels using a one-stage clotting protocol in thawed citrate-buffered blood plasma samples.[46] These samples were originally harvested shortly after participant inclusion, within one week of the index stroke (median: 4 days, and were stored uninterruptedly at -80°C until the activity measurements were performed after a single thaw.[46] All laboratory assays were conducted without knowledge of endpoint status.[46] In a few instances, the plasma aliquots were of insufficient volume to measure all three factors of interest; in these cases, FXI then FXII were prioritized over FVIII.[46]

6.2.4. PROSCIS-B: Statistical analysis

In the primary analysis, we created two categories for the coagulation factor activity levels. Participants with activity levels in the top fourth (>75th percentile) of all available measurements for a given factor were considered to have “high” levels of that factor. Participants with normal or low levels (less than or equal to the 75th-percentile) comprised the reference group.[46] We additionally analyzed the exposure variables continuously by dividing the raw activity measurements of each factor by the standard

deviation (SD) of that factor's measurements to facilitate comparison of the effect size estimates across the different factors.[46]

Among those experiencing the combined endpoint during the follow-up period, contributed person-time was defined from the index stroke date until the date of the endpoint event.[46] Person-time for individuals not experiencing the combined endpoint was computed as the interval between the index event date and the end of the study period, or, in instances of loss-to-follow-up, the date last reachable.[46]

We computed unadjusted Kaplan-Meier estimates of cumulative probabilities of the combined endpoint among the categorical exposure groups for each coagulation factor and created three graphical representations of the corresponding curves.[46] For each of the three exposures of interest, after visually confirming no strong violations of the proportional hazards assumption (i.e., no crossing of the Kaplan-Meier curves), we conducted a two-tailed log-rank test with a significance level of 5% to assess whether there was a crude, statistically significant difference in the combined endpoint among those having high versus reference levels.[46]

Next, we ran three Cox proportional hazards models, each to estimate the effect of having high levels of a given coagulation factor on the combined endpoint outcome, which were quantified as Hazard Ratios (HR) and corresponding 95% CIs.[46] These models were run as complete case analyses adjusted for an *a priori*-defined set of potential confounding factors, including age, sex, body mass index, high-density lipoprotein levels, low-density lipoprotein levels, smoking status, regular alcohol consumption, hypertension, diabetes mellitus, and acute coronary syndrome, each thought to influence both the coagulation factor levels and the risk for the combined endpoint.[46] We used Stata IC version 14.2 for all reported analyses in this subproject.[46]

6.3. Coagulation factor VIII activity, white matter hyperintensities and cognitive function in the older general population

6.3.1. Dataset: Population-based U.S. Cardiovascular Health Study (CHS)

Originally designed to determine risk factors for cardiovascular disease among the older general population, the large, longitudinal, Cardiovascular Health Study (CHS) conducted in the United States is well-known for its rigorous and comprehensive data collection.[55] This section provides a short summary of CHS design aspects relevant for this subproject's secondary data analysis. In the first CHS study year (1989-1990), a cohort of individuals from four U.S. cities were recruited from Medicare (federal health insurance for people aged 65 and older) eligibility lists (N=5,201).[55] Upon obtaining informed written consent, a detailed baseline examination was conducted; then, each year thereafter for the next nine years, annual clinic visits (or telephone interviews) took place.[55] Ethics committees at each participating CHS center approved the original study, and the Charité - Universitätsmedizin Berlin's ethics committee approved this secondary data analysis.[47]

6.3.2. CHS: Relevant variables and eligibility criteria

CHS investigators measured FVIII activity levels with the Coag-a-mate X2 instrument calibrated to WHO standards alongside several other biomarkers in blood samples collected during the baseline study visit.[47,56] As in the prior subprojects, all activity measurements were quantified as percentage units of normal pooled plasma.[47,56] We categorized FVIII activity levels as "high" (>75th percentile), "low" (less than or equal to the 25th percentile), or "normal" (reference, measurements between the first and third quartiles) levels.[47] To reflect the continuous nature of the activity level measurements, in a secondary analysis, we used the normalized variable obtained by dividing each FVIII activity measurement by the SD of all activity measurements as the exposure.[47] We also requested relevant data about demographic and additional cardiovascular risk factors documented during the baseline visit to describe the study sample and for confounding control.[47]

During the CHS follow-up period, two magnetic resonance imaging (MRI) scans of the brain were performed on consenting participants without contraindications; the first scan during the 2nd, 3rd or 4th follow-up visit, and the second scan during the 8th or 9th follow-up visit.[47,57,58] Experienced neuroradiologists used a 10-point scale to quantify the burden of white matter hyperintensities (WMH) visible on the standardized sagittal axial-spin density/T2-weighted cranial MRI images.[59] This scale was based on training template images, and ranged from 0 (no hyperintensities visible) to 9 (maximum amount of hyperintensities).[58] In these analyses, we operationalized WMH burden as "low" (scores of 0 or 1), "medium" (scores of 2 or 3) and "high" (scores of 4 or higher).[47].

To determine longitudinal worsening in WMH burden, we relied on ratings of both scans made at the time of the second scan, at which point a side-by-side assessment was possible.[47,57] We operationalized the WMH burden worsening between the two scans in three categories: (1) no worsening, (2) a worsening of 1 grade and (3) a worsening of 2 or more grades.[47]

The secondary outcome of this subproject was cognitive performance. We made use of the annual cognitive function measurements available starting with the first CHS follow-up, which included the 100-point Modified Mini-Mental State Examination (3MSE)[60] and the 90-point Digit Symbol Substitution Test (DSST).[61] In an effort to counteract missing values, we were able to impute missing 3MSE scores with estimates derived from Telephone Interview for Cognitive Status (TICS),[62] which were available for the sixth follow-up and in subsequence years for those who opted for a phone-based follow-up instead of an in-person visit.[47]

On top of the eligibility criteria imposed in the original CHS, we further excluded participants who reported a prior overt clinical stroke or transient ischemic attack (TIA) at baseline, those with confirmed dementia, and those with low baseline cognitive performance (defined as scoring less than 78 on the first 3MSE or within the lowest tenth of DSST scores).[47]

6.3.3. CHS: Statistical analysis

In this subproject, Stata IC software (version 14) was used to perform all analyses. First, to determine whether mean FVIII activity levels differed across covert WMH burden severity levels measured on the first cranial MRI, we first performed one-way Analysis of Variance (ANOVA).[47] Then, to estimate the effect of FVIII activity at baseline on this outcome, we used ordinal logistic regression models adjusted for a set of pre-specified demographic, socioeconomic, and cardiovascular factors thought to contribute to confounding.[47] These included: age, sex, education level, ethnicity, smoking status, frequency of alcohol use, body mass index, hypertension, diabetes, high and low density lipoprotein cholesterol levels, fibriongen, C-reactive protein, maximum common and internal carotid intima-media thickness, and the occurrence of a TIA or stroke after the baseline visit but before the first MRI scan.[47]

We obtained common odds ratios (cOR) with corresponding 95% CIs from crude and fully-adjusted models using the *gologit2* package in Stata,[63,64] which fits a partial proportional odds model to accommodate the ordinal outcome variable and can selectively relax the proportional odds assumption in case of violations.[47] The fully-adjusted models included a variable for the occurrence of an overt cerebrovascular event (clinical stroke or TIA) during the follow-up period but before the first MRI scan; however, since this variable is both a proxy for confounding as well as a potential causal mediator on the path between the exposure and outcome variables, we further performed a sensitivity analysis in which we omitted it from the model.[47]

For the longitudinal comparisons assessing the relationship between baseline FVIII activity levels and the outcome of WMH burden worsening, only participants taking part in both cranial MRI scans could be included.[47] We again computed cORs and 95% CIs from ordinal logistic regression models estimated using *gologit2*.[47,63] In addition to the aforementioned set of confounding variables, these longitudinal analyses were additionally adjusted for the time interval between scans.[47] We did not adjust for baseline WMH burden, as such an adjustment with a change-score as an outcome has been shown to introduce bias.[47,65]

To quantify the associations between baseline FVIII activity levels and the second outcome of interest, cognitive performance (operationalized as 3MSE and DSST scores), we first obtained crude and adjusted β estimates and 95% CIs from linear regression models.[47] The same set of variables was used for confounding control as in the aforementioned analyses of the primary outcome (WMH burden).[47] To accommodate the repeated nature of the serial, annual cognitive performance measurements, in the longitudinal analyses, we used linear models with mixed effects, introducing random intercepts for each individual, to obtain the adjusted effect estimates of interest (presented as β estimates and 95% CIs).[47]

6.4. Summary of data sources

To provide an overview of the three datasets used in my dissertation subprojects, I have created a condensed summary of their designs, participants' ages, and years during which the studies recruited participants in Table 1. This table also includes the number of participants who ultimately met the criteria for inclusion, as well as the exposure and outcome variables of interest for each subproject.

Table 1. Overview of datasets, designs and study samples per subproject

Original dataset information					**Subproject study design**		
Study name	**Study population**	**Study design**	**Ages**	**Recruit - ment period**	**Included participants**	**Main Exposure**	**Outcomes of interest**
RATIO (phase 2) [48,49]	Young women living in the Netherlands	Population -based case-control study	18-50	1990-1995	755 controls, 216 myocardial infarction patients, 182 Ischemic stroke patients with available DNA samples. 630 controls also had blood samples [45]	42 SNPs capturing common variation in *F12, F11, KLKB1* and *KNG1* genes	1. Coagulation factor traits - FXII, FXI, PK and HMWK antigen levels - FXI activity levels 2. Disease - Myocardial infarction - Ischemic stroke
PROSCIS -B [54]	First stroke patients in Berlin	Hospital-based cohort study	18+	2010-2013	576 [46]	FXII, FXI and FVIII activity levels	1. Combined endpoint (stroke, myocardial infarction, death)
CHS [55]	Adults from four US communities	Population -based cohort study	65+	1989-1990	4,295 [47]	FVIII activity levels	1. First measurement - White matter hyperintensity burden - Cognitive performance (3MSE, DSST) 2. Longitudinal - White matter hyperintensity worsening - Cognitive decline

Abbreviations: SNP; single nucleotide polymorphism; FXII, coagulation factor XII; FXI, coagulation factor XI; PK, prekallikrein; HMWK, high molecular weight kininogen; 3MSE, Modified Mini-Mental State Examination; DSST, Digit Symbol Substitution Test. The information presented in this table highlights the key design elements of the three dissertation subprojects; full details can be found in resulting original scientific publications: Rohmann et al. 2019, Rohmann et al. 2020, and Rohmann et al. 2020.[45–47]

7. Main Results

7.1. Common genetic variation in the contact system: RATIO study results

7.1.1. Genotyping results

The call rates in all genotyped control participant samples (N=755) ranged from 90.5% to 98.5% (Table 2).[45] Minor allele frequencies for each included SNP (organized by chromosome and gene) are also displayed in Table 2.

Table 2.

Chr.	Gene	Tag SNPs	Major/minor alleles	Minor allele frequency	Call rate
5	*F12*	rs1801020	C/T	25.3%	96.4%
5	*F12*	rs17876032	A/G	34.9%	91.0%
4	*F11*	rs4253399	T/G	38.9%	95.6%
4	*F11*	rs2036914	C/T	46.3%	96.6%
4	*F11*	rs1593	A/T	12.2%	96.4%
4	*F11*	rs4253417	T/C	41.7%	94.6%
4	*F11*	rs4253418	G/A	4.9%	97.9%
4	*F11*	rs4253430	G/C	35.7%	95.4%
4	*F11*	rs4253429	A/G	16.0%	97.1%
4	*F11*	rs4253406	G/T	7.9%	95.2%
4	*F11*	rs3733403	C/G	10.3%	96.0%
4	*F11*	rs5966	A/G	4.8%	95.4%
4	*F11*	rs4253431	G/A	14.1%	92.6%
4	*KLKB1*	rs2304595	G/A	43.2%	92.8%
4	*KLKB1*	rs1511801	T/A	45.5%	95.4%
4	*KLKB1*	rs4253243	T/C	6.9%	95.5%
4	*KLKB1*	rs4253327	T/A	28.9%	90.9%

4	*KLKB1*	rs4253326	T/C	18.4%	91.0%
4	*KLKB1*	rs925453	C/T	31.0%	98.5%
4	*KLKB1*	rs4253292	A/G	13.3%	95.9%
4	*KLKB1*	rs4253315	C/T	10.3%	92.1%
4	*KLKB1*	rs3087505	G/A	10.7%	96.2%
4	*KLKB1*	rs4253246	A/T	13.2%	93.4%
3	*KNG1*	rs5030062	A/C	38.3%	91.9%
3	*KNG1*	rs5030039	T/C	25.6%	94.7%
3	*KNG1*	rs166479	C/T	45.3%	92.3%
3	*KNG1*	rs5030060	C/T	30.0%	94.2%
3	*KNG1*	rs1621816	T/C	28.9%	93.9%
3	*KNG1*	rs2304456	T/G	11.4%	95.8%
3	*KNG1*	rs1469859	G/A	32.1%	96.3%
3	*KNG1*	rs266723	A/C	48.2%	90.5%
3	*KNG1*	rs5029980	T/C	12.9%	92.5%
3	*KNG1*	rs1648722	C/T	39.3%	91.8%
3	*KNG1*	rs5029999	C/T	20.5%	95.4%
3	*KNG1*	rs5030091	T/C	44.8%	93.9%
3	*KNG1*	rs5030102	T/G	9.9%	91.1%
3	*KNG1*	rs1624230	C/A	42.4%	94.8%
3	*KNG1*	rs1836860	T/C	33.8%	93.8%
3	*KNG1*	rs4686799	C/T	20.7%	93.9%
3	*KNG1*	rs5030095	G/C	13.3%	95.8%
3	*KNG1*	rs266760	G/A	26.2%	91.8%
3	*KNG1*	rs5030003	T/G	47.7%	94.3%
3	*KNG1*	rs5030072	--	failed in lab*	--

Abbreviations: Chr., chromosome; SNP, single nucleotide polymorphism (variant). Minor allele frequencies and SNP call rates were calculated using data from all control participants with available DNA samples. This Table represents a modification of a published version in Rohmann et al. 2019.[45]

7.1.2. Associations between SNPs and traits (coagulation factor levels)

We included 630 control participants with both DNA and blood plasma samples available for the protein quantitative trait loci analyses. In this section, results are presented organized by gene (*F12*, *F11*, *KLKB1* then *KNG1*).

***F12* SNPs and FXII antigen levels:** In the age-adjusted single SNP analyses within the *F12* gene, two SNPs were found to have highly statistically significant associations (below the Bonferroni-corrected alpha level) of considerable magnitude with FXII antigen levels: lead SNP rs1801020 (β=-41.54, 95%CI: -45.67 to -37.41, $p=6.85 \cdot 10^{-67}$) and rs17876032 (β=-28.09, 95%CI: -32.52 to -23.66, $p=1.42 \cdot 10^{-31}$).[45] However, the secondary signal for rs17876032 disappeared after adjustment for the lead SNP in the conditional analyses.[45]

F11* SNPs, FXI antigen and FXI activity levels*: Of five *F11* SNPs showing statistically significant signals with FXI antigen levels, the strongest signal was with SNV rs2036914 (β=-11.38, 95%CI: -14.18 to -8.58, $p=7.87 \cdot 10^{-15}$).[45] After adjustment for the lead SNP and mutual adjustment, two secondary signals persisted; the associations with rs1593 ($\beta_{adjusted}$=-7.47, 95%CI: -11.99 to -2.96, $p_{adjusted}$=0.001) and rs4253399 (β=6.11, 95%CI: 2.02 to 10.20, p=0.003).[45] Five *F11* SNPs were also significantly associated with FXI activity levels in the age-adjusted single SNP regression models.[45] In addition to lead SNP rs4253399 (β=9.24, 95%CI: 6.69 to 11.79, $p=3.32 \cdot 10^{-12}$), only the additional secondary signal between rs1593 and FXI activity ($\beta_{adjusted}$=-5.65, 95%CI: -9.46 to -1.83, $p_{adjusted}$=0.004) persisted after adjustment for the lead SNP and mutual adjustment for the other SNPs with secondary signals (rs4253418 and rs2036914).[45]

***KLKB1* SNPs and PK antigen levels:** For this second gene of interest that, like *F11*, is also situated on chromosome 4, we estimated below-threshold signals for a total of six *KLKB1* SNPs in the age-adjusted single-SNP models with PK antigen levels as the outcome.[45] In addition to the lead SNP rs2304595 (β=10.33, 95%CI: 6.65 to 14.02, $p=5.58 \cdot 10^{-8}$), a conditionally independent signal with rs4253243 persisted ($\beta_{adjusted}$=-12.38, 95%CI: -20.07 to -4.69, $p_{adjusted}$=0.0017).[45] This locus had not previously been described in the literature and thus represents a novel finding.

***KNG1* SNPs and HMWK antigen levels:** Of the four contact system genes, *KNG1* had the highest number of SNPs to capture common genetic variation; we found statistically significant associations for 11 of these with HMWK antigen levels in the age-adjusted, single-SNP analyses.[45] The lead SNP was rs5030062 (β=10.23, 95%CI: 7.87 to 12.59, $p=1.63 \cdot 10^{-16}$), and upon conditioning on this SNP, associations between eight *KNG1* SNPs and HMWK antigen levels remained.[45] After mutual adjustment for the strongest secondary signals, only one association involving rs5029980 persisted ($\beta_{adjusted}$=5.86, 95%CI: 2.40 to 9.32, $p_{adjusted}$=0.001).[45] This novel locus was first identified by our work.

Across-trait associations: In addition to the aforementioned associations between single SNPs from the contact system genes and levels of the respective proteins encoded by each of these four genes, we also looked for signals "across-traits". Upon investigating any statistically significant associations between SNPs of one gene and other traits (expressed contact factor levels not encoded by that gene) in the age-adjusted single-SNP models, we found no signals below the Bonferroni-corrected statistical significance threshold for SNPs within the *F12* or *F11* genes.[45] We identified three *KLKB1* SNPs having statistically significant single-SNP associations below the multiple testing-corrected threshold with both FXI antigen and activity levels (rs1511801, rs2304595 and rs3087505).[45] In *KNG1*, four SNPs (rs5030062, rs5030060, rs1469859 and rs166479) were associated with FXI activity levels, and two of these (rs5030062, rs5030060) also with FXI antigen levels.[45] Furthermore, two *KNG1* SNPs (rs5030062 and rs5030060) were associated with PK antigen levels.[45]

7.1.3. Haplotype construction and analysis

F12: Since no secondary signals were conditionally independent of the lead SNP that we identified for the association with FXII antigen levels (rs1801020), no haplotypes could be constructed for *F12*.[45]

F11: Using the discovered *F11* SNPs, we were able to construct five haplotypes present in at least 1% of the included control participants.[45] Compared with the reference haplotype occurring in 15% of participants, which contained only major alleles of rs4253399 (T), rs2036914 (C) and rs1593 (A), a haplotype consisting of at least one copy of the rs4253399 major allele and less frequently occurring minor allele variants of

both rs2036914 (T) and rs1593 (T) occurred in 12.0% of control participants.[45] This haplotype was statistically significantly associated with 12.9-unit lower levels of FXI antigen (SE=2.8, p<0.0001).[45] No other *F11* haplotypes occurring in at least 1% of the control participants showed a statistically significant association with the expressed FXI antigen or FXI activity levels.[45]

KLKB1: The two *KLKB1* SNPs we identified with mutually independent associations with PK antigen levels did not occur together in an inherited combination; therefore, no haplotypes could be constructed in our study sample of control participants.[45]

KNG1**:** In *KNG1*, we constructed a total of five haplotypes consisting of the lead SNP and the two secondary signals.[45] The reference haplotype consisted of all major alleles only for rs5029980 (T), rs2304456 (T) and rs5030062 (A) and occurred in 43.2% of participants.[45] The alternative, rarer haplotype containing at least one copy of C alleles for both the rs5029980 and rs5030062 loci was associated with an increase in HMWK antigen levels (β=25.2, SE= 3.4; p-value<0.0001).[45]

7.1.4. Associations with disease phenotypes

We found no statistically significant associations between any of the included SNPs that were significantly associated with the protein levels encoded by their gene (N=26), and either myocardial infarction or ischemic stroke in the disease phenotype analyses that included all RATIO case and control participants with DNA samples.[45] Furthermore, upon assessing the associations between the haplotypes described in the previous section and these diseases, none were statistically significant.[45]

7.2. High levels of FVIII, FXI or FXII activity and post-stroke outcomes: results from the PROSCIS-B study

7.2.1. Study population characteristics

The majority of first-ever stroke patients in the PROSCIS-B study were middle or older age (median age: 69 years; IQR 58-76), more than half were men (61%), and the participants' median BMI was 27 kg/m^2 (IQR: 24-30).[46] A total of 576 participants met eligibility criteria and had at least one coagulation factor measurement.[46] For all

available measurements, median levels of FVIII, FXI, and FXII activities were 140 (IQR: 51), 133 (IQR: 31), and 108 (IQR: 36).[46]

7.2.2. Survival analysis results

A total of 1419.5 person-years were contributed by the included PROSCIS-B study participants.[46] During follow-up, a total of 94 sequelae meeting the definition of the combined endpoint were observed (48 deaths, 41 strokes, and 5 myocardial infarctions).[46]

Plots of the Kaplan-Meier estimates for the cumulative probability of the combined endpoint are displayed in Figure 1, in which high levels of each of the three coagulation factors were compared with low-to-normal reference levels.[46] The log-rank test results for the crude comparison of high versus reference activity levels indicated a significant difference for FVIII (p=0.0001) but not FXI (p=0.06) or FXII (p=0.48) activity levels.[46] However, these results do not reflect any confounding adjustment, which is critical in this observational cohort design, and was thus addressed in a second analytical step.

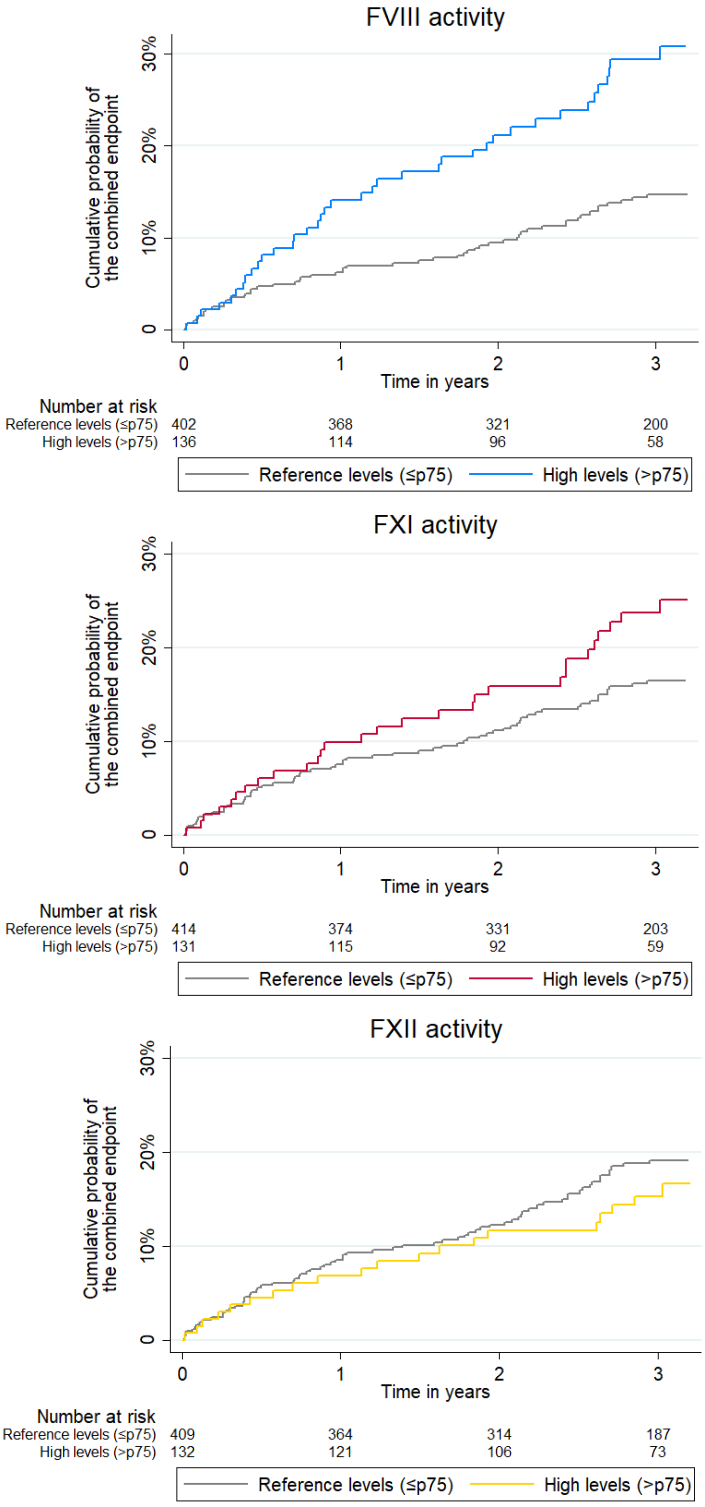

Figure 1. Kaplan-Meier estimates of cumulative probabilities of the combined endpoint by coagulation factor activity group (high versus reference levels). Activities are reported as percentages of activated normal pooled plasma. Abbreviations: FVIII, coagulation factor VIII; FXI, coagulation factor XI; FXII, coagulation factor XII; p75, 75th-percentile. This figure was redrawn from original data, and represents a modification from the original version published in Rohmann et al. 2020.[46]

As shown in Figure 2, after adjustment for the *a priori*-specified confounding variables, having high levels of FXI activity compared with reference levels increased the hazard for the combined endpoint (HR = 1.80, 95%CI: 1.09–2.98).[46] This was also observed for having high FVIII activity levels (HR = 2.05, 95%CI: 1.28–3.29).[46] Among those having high levels of FXII activity, we did not observe a statistically significant relationship with the combined endpoint (HR=0.86, 95%CI: 0.49–1.51).[46] The standardized continuous analyses yielded analogous results (Figure 2).[46]

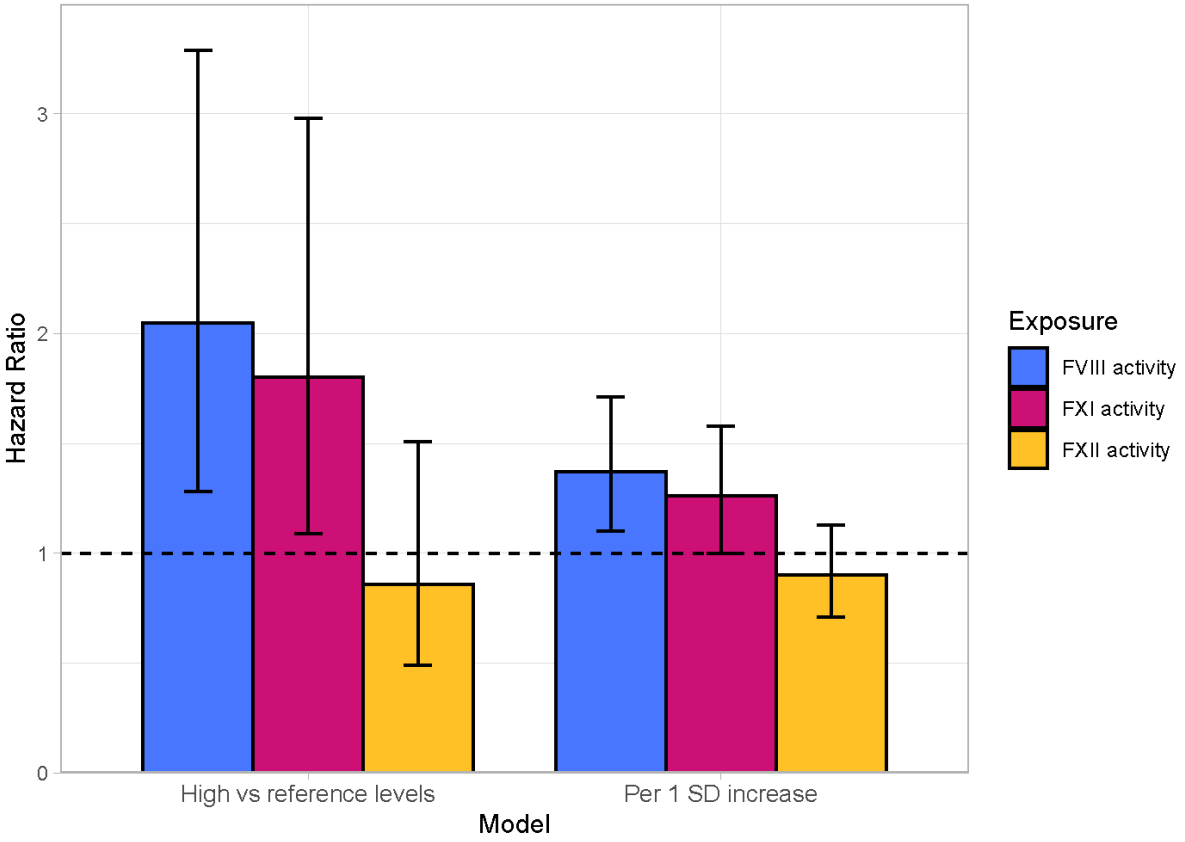

Figure 2. Hazard ratios for FVIII, FXI and FXII activity levels (high versus reference) and the combined endpoint outcome. High levels were defined as values greater than the 75th percentile for each factor's measurements. Abbreviations: FVIII, coagulation factor VIII; FXI, coagulation factor XI; FXII, coagulation factor XII; SD, standard deviation. This figure reflects effect estimates from published tables in Rohmann et al. 2020.[46]

7.3. FVIII activity, white matter hyperintensities and cognitive performance: results from the CHS

7.3.1. Baseline characteristics of included participants

Of the original 5,201 CHS participants, after excluding those with missing FVIII activity measurements, prior overt clinical stroke or TIA, adjudicated dementia, and those having very low baseline cognitive performance, the study population meeting inclusion for at least one of the analyses consisted of 4,295 individuals aged 65 or older at study start.[47] The mean age was 72.3 years (SD: 5.3), more than half of the included participants were female (59.3%), the vast majority were white (96.0%), and 76.9% had at least 12 years of education (high school diploma). In terms of traditional cardiovascular risk factors, 45.4% of participants were hypertensive, 13.9% had diagnosed or detected fasting glucose levels indicative of diabetes, 54.1% reported being a former or current smoker, and 13.8% reported consuming more than 7 alcoholic drinks per week, on average.[47] A complete summary of baseline characteristics for the included participants can be found in the original publication.[47]

7.3.2. FVIII activity and white matter hyperintensity burden

For a total of 2,735 included participants, WMH burden results from the first cranial MRI scan were available in the dataset for analysis.[47] In line with our hypotheses, the mean FVIII activity levels were highest among individuals with the high WMH burden (mean: 121.2 activity units) compared with individuals having medium (120.5) or low (115.3) WMH burden.[47] This difference was statistically significant in the crude, unadjusted comparison (ANOVA p=0.001).[47]

Compared with having “normal” FVIII activity levels (reference; between the 1st and 3rd quartiles), having “high” FVIII activity levels in the upper fourth of all measurements (>75th percentile) was weakly but not statistically significantly associated with a more severe WMH burden on the first cranial MRI scan after full adjustment for confounding (cOR=1.20, 95%CI: 0.99-1.45).[47] On the other end of the spectrum, having FVIII activity levels in the lowest fourth (≤25th percentile) was not associated with more severe WMH

burden (cOR=1.03, 95%CI: 0.85-1.23).[47] In the analysis treating FVIII activity as a continuous variable, we obtained analogous results. Per one SD increase in FVIII activity, corresponding to an increase of 36 units, our model estimated a cOR for the ordinal WMH burden outcome on the first MRI scan of 1.08 (95%CI: 0.99-1.17).[47]

In the longitudinal analyses, we quantified the association between baseline FVIII activity levels and WMH worsening observed over time on two cranial scans, separated on average by 5 years, in participants who received both scans (N=1,527).[47] After full adjustment for confounding, we found no statistically significant association between having high (cOR=1.18, 95%CI: 0.87-1.59) or low (cOR=1.01, 95%CI: 0.76-1.33) FVIII activity levels and this ordinal outcome variable, compared with normal (reference) levels.[47] Similarly, per one SD increase in FVIII activity (36.3 units), the cOR was 1.07 (95%CI: 0.94-1.22).[47]

7.3.3. FVIII activity and cognitive performance

In the fully-adjusted models, we found that having high FVIII activity levels at the first follow-up visit was not statistically significantly associated with 3MSE scores (β=-0.06, 95%CI: -0.45 to 0.32) or DSST scores (β=-0.69, 95%CI: -1.52 to 0.13) among the approximately 4,000 participants for whom these scores were available.[47] Having low FVIII activity levels showed a small, not statistically significant association with 3MSE scores (β=0.33, 95%CI: -0.04 to 0.71), and a small, statistically significant association with DSST scores (β=0.85, 95%CI: 0.06 to 1.64) in the fully-adjusted models.[47] Correspondingly, in the continuous analyses, per one SD increase in FVIII levels, we estimated a 0.14-point lower average score for the 3MSE (95%CI: -0.30 to 0.03) and 0.37-point lower score for the DSST (95%CI: -0.72 to -0.02).[47]

Our longitudinal analyses showed that high FVIII activity levels were not associated with 3MSE or DSST scores over time across the duration of study follow-up in the fully adjusted mixed models (3MSE β=-0.07, 95%CI: -0.58 to 0.44; DSST β=-0.22, 95%CI: -0.97 to 0.53).[47] Our findings treating the exposure as a continuous variable were consistent (β for 3MSE per SD of FVIII=0.15, 95%CI: -0.06 to 0.37; β for DSST per SD of FVIII=-0.11, 95%CI: -0.43 to 0.22).[47]

In the sensitivity analysis, in which we removed the variable encoding the occurrence of a TIA or stroke during the follow-up, we obtained nearly identical estimates for all relationships outlined above.[47]

8. Discussion

8.1. Summary of main findings, strengths and relevance

The overarching aim of this dissertation work was to investigate possible determinants and consequences of high levels of coagulation factors, which were selected as particularly promising targets in the context of ischemic vascular pathologies. Specifically, across the three subprojects, I sought to probe potential genetic influences on coagulation factor levels and disease phenotypes,[45] examine relationships between coagulation factors already implicated in first stroke with outcomes after stroke,[46] and determine whether FVIII, known to be a cause of overt vascular events, may also contribute to covert white matter hyperintensity development and cognitive performance among healthy older adults.[47]

My work on the first subproject contributed to the identification of two new genetic loci that influence contact system factor levels. Using data from the RATIO study, we identified novel locus rs4253243 (which tagged rs4253331) within the *KLKB1* gene as a determinant of PK antigen levels, and this relationship was conditionally independent of the *KLKB1* lead SNP rs2304595.[45] The second novel locus discovered was rs5029980 (no tagged SNPs) within the *KNG1* gene, which remained independently associated with HMWK levels, even after conditioning on both the *KNG1* lead SNP rs5030062 and a replicated secondary signal in *KNG1*, rs2304456.[45]

We further confirmed postulated across-trait influences of some *KNG1* loci on both FXI antigen and activity levels.[45] These relationships persisted after conditioning on HMWK antigen levels, and appeared to be driven by FXI activity levels, since the associations with FXI antigen disappeared after further adjustment for FXI activity.[45] Neither the

single SNPs nor haplotypes constructed from all signals within a given gene were found to be associated with ischemic stroke or myocardial infarction in the analyses using data from both case and control participants.[45]

The work of the first subproject provided a further valuable contribution to the field of coagulation factor genetics in that it replicated several previously reported single-variant associations in a single study. The idea of external replication of genetic signals in independent datasets is a critical pillar to genetic discovery studies,[66,67] especially in light of the ongoing, well-described reproducibility crisis in the published scientific literature.[68]

Although our study of coagulation factor genetic variation only captured common variation across the four contact system genes, it had a unique strength in that we had multiple coagulation factor measurements available for the same set of individuals, as well as both antigen and activity level measurements for FXI. This allowed us to probe more complex across-trait relationships that had been previously postulated in the literature. For example, we showed that the association between *KNG1* SNP rs5030062 and FXI antigen levels appears to be driven by rs5030062's (or a tagged SNP's) influence on FXI activity.[45] This finding provides relevant contextualization for a previously proposed mechanism of *KNG1*-driven modulation of FXI levels by a tagged SNP rs710446.[69] The C allele of this missense mutation is thought to alter the binding site functionality of HMWK, which impacts FXI activity (but not antigen levels), because HMWK and FXI circulate together as complexes in the blood.[69] Of course, the exact mechanism would need to be investigated in a functional study, but our corroboration of this initial proposition is an important first step in terms of understanding genetic influences on the contact system factor expression.[45]

In the second subproject, using data from the stroke patient cohort, PROSCIS-B, I was interested in quantifying the relationships between activity levels of two of the aforementioned contact system factors belonging to the intrinsic coagulation cascade (FXI and FXII) as well as the more downstream FVIII with post-stroke outcomes.[46] While all three of these factors have been implicated as causes of first ischemic stroke, evidence regarding whether they also contribute to the development of post-stroke endpoints was lacking in the literature. After adjustment for confounding, while we did

not observe a relevant relationship between high levels of FXII activity and the combined endpoint (HR=0.86, 95%CI: 0.49–1.51), the HR estimates for having high FXI and high FVIII were 1.80 (95%CI: 1.09–2.98) and 2.05 (95%CI: 1.28–3.29), respectively, suggesting the relevance of these factors in the context of secondary events after first stroke.[46]

Upon initiation of this subproject, the potential role of coagulation factors XII, XI, and VIII in contributing to vascular events after stroke was largely unexplored. This was surprising given their prior implications in first ischemic stroke and other vascular events. First results from concurrent studies have since linked FVIII to post-stroke outcomes. For instance, a study of hospitalized stroke patients in the U.S. found that ischemic stroke patients with high levels of FVIII more frequently experienced vascular events during the acute phase.[70] Among the subset of ischemic stroke patients who received thrombolysis therapy, another study found that having high FVIII activity levels (measured both pre- and post-thrombolysis time points) was associated with worse functional outcomes three months after the index event.[71]

Due to its large size and three years of follow-up, during which incident vascular events and mortality were recorded, using the PROSCIS-B dataset provided a perfect opportunity to build on these first observations and investigate whether the impact of FVIII persisted also in the longer term.[46] Furthermore, in the same study population, we could investigate the relationships between the activity levels of FXII and FXI and post-stroke outcomes, and provide first results showing that FXI seems to play a role in the occurrence of relevant post-stroke endpoints.[46]

The third subproject again focused on FVIII activity, but this time, my central interest was to estimate effects of having high FVIII levels on covert WMH and cognitive outcomes in a large, general population-based sample of older persons. The underlying motivation for this research line was our hypothesis that FVIII-induced white matter hyperintensity development might contribute to subsequent cognitive decline.[47]

Contrary to our hypothesis, we did not find that individuals with high FVIII activity levels had a more severe burden of white matter hyperintensities on the first cranial MRI scan compared with individuals with low or normal levels.[47] We further did not find evidence

that having high FVIII was linked to WMH worsening over an average period of 5 years.[47] With regard to the cognitive performance outcomes, we did not observe strong relationships between FVIII activity levels and the 3MSE or DSST scores in the general population of older persons; neither at the time point of first assessment nor in the longitudinal analyses with annual measurements across the 9-year follow-up period.[47]

These findings were consistent between both tests assessing cognitive performance (3MSE and DSST).[47] Both tests were repeated according to standardized protocols and had unique advantages; for the 3MSE, we were able to impute telephone estimations (TICS)[62] to avoid large numbers of missing values over time.[47] We opted to use DSST scores as a second outcome measure to check the robustness of our results since the DSST overcomes the well-described ceiling of the 3MSE.[47,61]

Since we observed no relationships between FVIII activity levels and either outcome of interest (WMH or cognitive test performance), a formal causal mediation analysis was not warranted. Taken together with the findings of the second subproject, it appears that FVIII is a more relevant research target for overt vascular events than for covert ischemia or cognitive decline in the general population of older persons.[47]

8.2. Critical reflection and recommendations for future coagulation factor research in observational studies

8.2.1. Timing is everything in matters of coagulation factor measurement

Research investigating the relationship between coagulation factors and specific endpoints is only interpretable with explicit consideration of the timing of the measurements, that is, when the blood samples are obtained relative to the disease process. Therefore, it is of utmost importance to choose a relevant time point for the coagulation factor measurement aligned with the operationalization of the exposure in the research question of interest. My dissertation project made use of coagulation factor measurements obtained from blood samples taken at time points specified by the study protocols of already existing studies. While I, myself, could not choose the moment at which these blood samples were taken, I emphasize that the results of my work should be interpreted and contextualized with explicit consideration of the underlying disease

process timeline. Noting that there is no definitive scientific consensus definitions for the exact length of the “acute phase” nor the meaning of “short-term” versus “long-term” in the context of post-stroke outcomes, I created a simplified timeline with one possible operationalization of the relevant time periods leading up to and following ischemic stroke onset for illustrative purposes (Figure 3) to demonstrate some lessons learned.

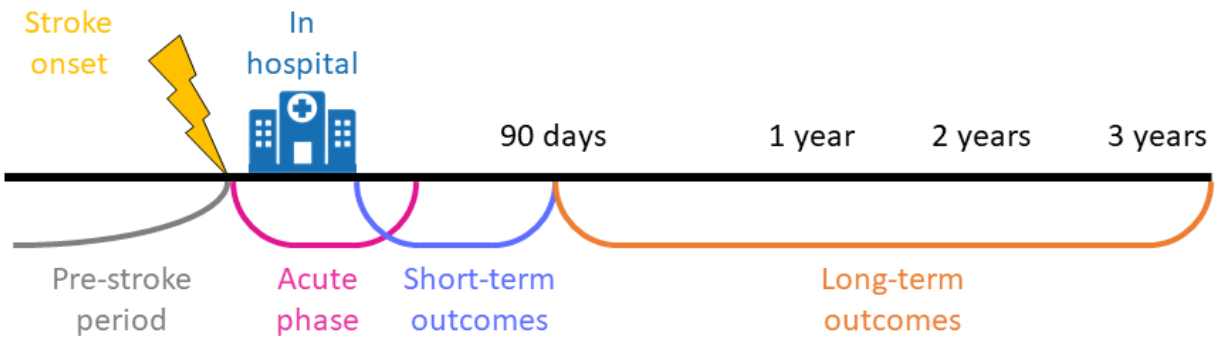

Figure 3. Timeline surrounding stroke onset relevant for blood biomarker measurements. Timeline not to scale and is for illustrative purposes. There is no consensus of how these terms (“acute phase”, “short term”, and long-term” outcomes) are operationalized in the scientific literature. This operationalization is therefore exemplary, but the concept holds also for other definitions of these terms.

To ascertain whether a given exposure of interest could be a causal risk factor for first stroke, the relevant measurement must be made in a sample taken during the pre-stroke period. Though I did not investigate causes of first stroke in this thesis, in the subproject focusing on a different set of outcomes (WMH development and cognitive performance decline) the FVIII activity measurements were technically performed during this phase because they were sampled from the general population who had not experienced ischemic stroke.

Though there is formally no consensus on its length in the literature, after ischemic stroke onset, the acute phase immediately follows (Figure 3). This time period extends for at least several days up to several weeks after symptom onset and is characterized by an inflammatory response, during which several biomarkers, called acute phase reactants, such as CRP, FVIII, and fibrinogen, become elevated for an extended period of time.[70,72,73]. This is the time period during which ischemic stroke patients should be

rapidly hospitalized and may receive treatment, such as intravenous thrombolysis and/or mechanical thrombectomy.[74] Patients remain hospitalized for various lengths of time depending on ischemic stroke subtype, severity and comorbidities.[75] In Germany, the median stay on a stroke unit is 7 days, though the total length of stay in the hospital may be longer, depending on the speed and degree of recovery.[75]

During hospitalization (and increasingly, already during prehospital provision of stroke care, e.g. in a mobile stroke unit[3]), it is comparatively convenient to obtain blood samples from stroke patients, as laboratory tests are a part of the clinical routine. However, this convenience comes with an important caveat for coagulation and inflammation research. Namely, measurements performed in blood sampled during this period, especially in the first days after the index event, may be vulnerable to capturing elevated levels of specific factors if they are impacted by the aforementioned acute phase response. This means that FVIII measurements ascertained from blood sampled during this period likely represent a mixture of pre-stroke FVIII levels and temporary acute phase elevation. Intravenous thrombolysis administration also happens early on in this phase, which could further impact some coagulation factor measurements in the very short-term.

In-hospital biosampling makes clinical sense as a measurement time point, as blood sampling is routine in this setting, and measured biomarkers may also be used to inform future treatment strategies and clinical decision-making for both the short- and long-term. Taking blood samples at the time of hospital discharge, temporally further from ischemic stroke onset, may provide coagulation factor estimates closer to pre-stroke levels, however, one should be cautious if inferring that these levels represent pre-stroke levels, especially if they are known to be impacted by the acute phase.

In the case of the second subproject using PROSCIS-B data, for pragmatic reasons, blood was sampled during the time of hospitalization. However, since the day of discharge varies per patient, heterogeneity in blood sampling could further complicate the interpretation of the results, representing a further possible limitation to our secondary analysis of this existing dataset.

To avoid capturing transient changes in biomarker levels induced by the acute phase

reaction, blood sampling can be performed after a longer period of time has elapsed, for example, three months after stroke, or even one or more years after the index event, concurrent with long-term outcome measurements. The 3-month time period may be particularly practical, since many stroke patient cohorts include a 3-month visit to assess a series of “short term” outcome parameters. However, samples taken at 3-months post-stroke time point were, unfortunately, not available from the data sources used in my dissertation project.

Measurements taken in the longer-term may be preferred in common case-control study designs with survivorship sampling, as measuring during these late time periods leads to the lowest risk of ‘reverse causation;’ a phenomenon in which a consequence is mistaken for the actual cause.[76] We encountered such a situation in the RATIO case-control study dataset, which we used to assess associations between the individual genetic variants and disease phenotypes.[45] In the RATIO study, coagulation factor measurements were performed in blood samples taken 5 years after the index event (in cases) or their index date (for the matched control participants).[48,49] Though this approach also overcomes the limitation of taking measurements in the acute phase, it is important to note that using measurements from blood sampled at this late time point means that only participants who survived to the time point of blood sampling could be included, introducing a potential survivorship bias. Additionally, among survivors of vascular events, we cannot rule out that some preventative treatments (e.g., anticoagulation medication use after ischemic stroke) might have altered these biomarker measurements from their pre-stroke values.

Taken together, the aforementioned examples from this dissertation work illustrate how the timing of blood sampling of biomarkers is highly relevant. Specifically, measurements taken at different time points provide different information and represent different underlying constructs. In the existing literature, these limitations find little acknowledgement. The results of my work illustrate why care should be taken in the reporting and interpretation of observational, clinical studies involving coagulation factor measurements. In my dissertation, I took care to interpret and contextualize my results in light of the time of measurement, as failing to do so could obscure understanding and even lead to bias, as illustrated in the next section.

8.2.2. DAGs: visual tools to identify common biases in observational studies

Much of the work in clinical epidemiology stems from identifying biases that arise in observational studies and developing strategies to mitigate these biases in study design and analysis. In the process of writing this dissertation, I became increasingly exposed to directed acyclic graphs (DAGs). Progressively, modern epidemiologists are using DAGs across applied clinical disciplines to inform study design and analysis, although they are not (yet) commonly implemented in the field of stroke research.[77]

DAGs are visual tools based on Bayesian Networks, which were developed largely through work in the field of computer science.[78,79] DAGs are intuitive representations of the underlying so-called data generation process, encapsulating cause and effect relationships surrounding a given causal research question.[78,80,81] In this section, I will briefly introduce DAGs, which I will use in the remainder of the Discussion together with examples from my dissertation work to explore some possible limitations outside of those already mentioned in the individual publicaitons, highlighting relevent considerations for future research on coagulation factors.

A DAG consists of a set of “nodes,” each corresponding to a variable. These nodes can be connected to each other by directed arrows, which represent direct causal effects, and point from individual causes to their effects.[78,80] Using DAGs, researchers can intuitively generate visualizations of the reality that can provide a framework useful for communication between clinical subject-matter experts and epidemiologists.[77] Under a set of common assumptions, there is a direct correspondence between the DAG and conditional independence between variables.[78,80] This means, that if the DAG is correct, under certain conditions, it is possible to estimate a causal effect from observational data, which is the goal of many relevant clinical lines of scientific inquiry.[80,82]

Paths connecting any two nodes in a DAG through one or more arrows can be classified as “open” or “closed”[78,80,81]. If there is an open path in a DAG between two nodes, the corresponding variables will be associated in the corresponding dataset.[78] On the other hand, if no path exists or an existing path is “closed” (which can occur naturally, due to study design (e.g., restriction), or analytically, e.g., through

adjustment), then, the two variables will be independent.[78,80,81] Many comprehensive introductions to DAGs, including practical examples, have been published elsewhere.[78,83–85]

8.2.3. Coagulation factors in the context of ischemic stroke: confounding considerations

Earlier, I discussed how FVIII levels are altered by the occurrence of ischemic stroke as a part of the acute phase response. Intuitively, therefore, measurements of this factor sampled during the acute phase reflect a combination of pre-stroke FVIII levels as well as the acute phase elevation, a phenomenon not only true for FVIII, but also for many other inflammation biomarkers. As mentioned in the section “Time points of the measurement” (8.2.1), failing to explicitly consider the timing of coagulation factor measurements can lead to bias. I would like to explore this further in the context of one of my research questions from the second subproject (PROSCIS-B).

Below, I have created a DAG (Figure 4), as introduced in the previous section, of the data generation process behind the research question of interest, which was to estimate the effect of FVIII activity levels measured after ischemic stroke ($FVIII_1$) on the combined endpoint including recurrent stroke, myocardial infarction or death by any cause. In this simplified example for purposes of critical reflection, I have depicted *Death* as the outcome of interest, even though the following reasoning also applies to the other vascular endpoints.

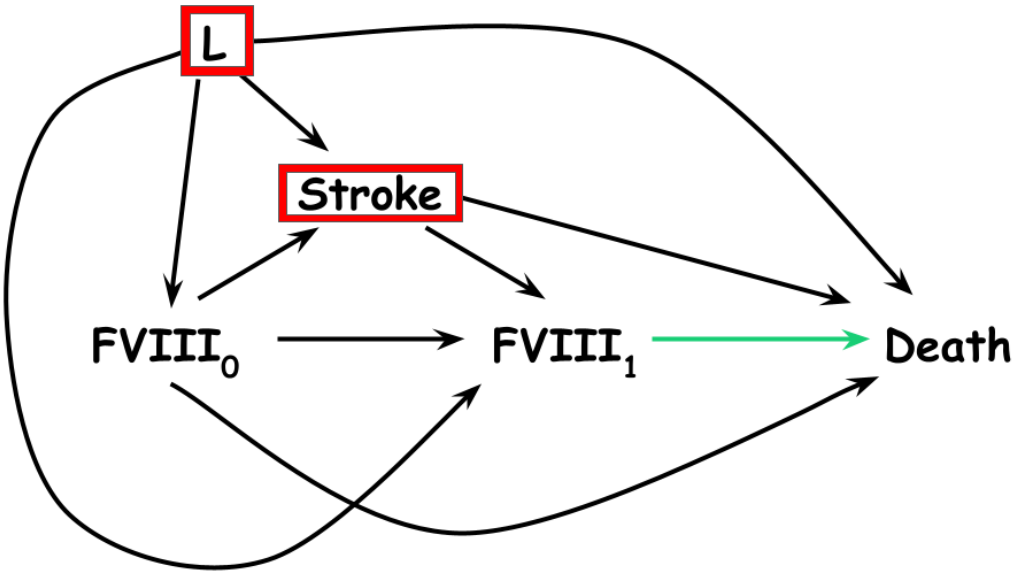

Figure 4. Directed acyclic graph for the PROSCIS-B subproject example. The green arrow indicates the research question under study. *FVIII*$_1$ denotes the measured exposure (FVIII activity level measured at a given time point after ischemic stroke); *FVIII*$_0$ denotes the unmeasured FVIII activity levels just before the index stroke event; *L* stands for the set of confounding variables we adjusted for in the analyses. The red boxes represent conditioning due to the study design (i.e. restriction to only ischemic stroke patients indicated by the box around *Stroke*) or analysis (i.e. regression adjustment for potential confounding sources indicated by the box around *L*).

In Figure 4, the causal effect of interest is indicated by the green arrow. A further node in the DAG, *FVIII*$_0$, represents pre-stroke FVIII activity levels. As indicated by the arrows in the DAG, *FVIII*$_0$ influence FVIII activity levels measured after ischemic stroke (*FVIII*$_1$). I have also depicted the assumption that *FVIII*$_0$ also influences *Death* through a series of more downstream, intermediate pathways. Since *FVIII*$_0$ is known to be a contributing cause of ischemic stroke (*Stroke*), this is also depicted by a further arrow. Among individuals who experience an ischemic stroke, FVIII elevation occurs in the acute phase, represented by Stroke → FVIII1. The DAG also shows that individuals who experience ischemic stroke are at higher risk for death (Stroke → Death). Finally, both *FVIII*$_0$ and *FVIII*$_1$ are influenced by a set of cardiovascular risk factors, here summarized as a single node *L*, which are also known to cause both ischemic stroke and death (for example, age, sex, and smoking).

The effect of *FVIII*$_1$ on the outcome was estimated in my subproject by fitting a regression model adjusted for the set of cardiovascular risk factors represented by *L*

only using data of stroke patients (*Stroke*=1), indicated in the DAG by the red boxes around nodes *L* and *Stroke*, respectively. The “blocking” of nodes reflects conditioning for these variables in either the analysis phase (regression adjustment) or already in the design phase (restriction to include only stroke patients).[78,80,81]

Blocking a node on any open path in DAG closes that path.[78] Indeed, we can obtain a valid estimate for the causal contrast of interest if there are no other open paths between exposure ($FVIII_1$) and outcome (*Death*) in the DAG other than the causal path(s) of interest (here: $FVIII_1 \rightarrow Death$). This is the reason why we adjust for common causes of exposure and outcome when we aim to estimate a causal effect in observational studies[78,80,81].

If open noncausal paths remain in a DAG, our estimation of the association between exposure and outcome will reflect a mixture of the causal effect of interest (if one is present) with these noncausal associations.[78] Confounding generally refers to the presence of open, non-causal “backdoor” paths that link the exposure to the outcome via a cause of the exposure.[78,80] In this case, through the restriction to only stroke patients and the adjustment for covariates *L,* many of the otherwise open non-causal paths are blocked (e.g., $FVIII_1 \leftarrow Stroke \rightarrow Death$, $FVIII_1 \leftarrow L \rightarrow Death$). However, even after adjustment for all the measured confounding variables (*L*) relevant in this DAG, there remains one open non-causal path between $FVIII_1$ and *Death*, which is due to the fact that pre-stroke FVIII levels ($FVIII_0$) are a common cause of post-stroke FVIII levels ($FVIII_1$) and the outcome.

Therefore, the conditional association we estimated in the PROSCIS-B data between $FVIII_1$ and the outcome likely represents a mixture of the true causal effect of interest and some residual confounding introduced by a spurious, non-causal association between $FVIII_1$ and the outcome due to the fact that $FVIII_1$ shares a common cause with the outcome ($FVIII_0$) for which we have no information and therefore cannot be blocked.

The DAG transparently showcases an important limitation of our findings for this research question, namely, an unavoidable residual confounding bias in our result due to the timing at which the blood sample was taken in the original study. This bias,

however, is not unique to our study or this particular question, and I suspect it is present in most, if not all, stroke patient cohort studies investigating biomarkers. The only way to further isolate the causal effect of interest would be to further adjust or restrict for levels of $FVIII_0$. Unfortunately, it is rare that multiple blood samples are available in existing studies, especially since large general population based-studies with repeated measures of FVIII measures would be needed to achieve this, though feasibility is difficult in practice and careful planning is required.

Moving beyond this simplified example, I would like to showcase a more complex application of DAGs to address confounding, this time using an example from the third subproject. Here, I will focus on the specific aim of quantifying the relationship between FVIII activity levels and white matter lesions in a general population sample of older persons in the CHS. I used the web application Dagitty[86] to construct the relevant DAG for this research question, displayed in Figure 5, this time including each of the relevant covariates as individual nodes. As evidenced by the intricate web of arrows, the DAG shows how many of the traditional cardiovascular risk factors influence each other as well as both the exposure FVIII activity (*FVIII*) and outcome of interest (*WMH*). Blocking the set of variables that we *a priori* selected to control for confounding (marked as white colored nodes) results in no remaining open non-causal paths between the exposure and the outcome, and leaves only the causal path of interest (*FVIII→WMH*). Assuming this DAG is correct, and that common statistical assumptions required for casual inference are fulfilled,[80] then the measured association between FVIII activity and WMH reported in our study should represent a deconfounded estimate of the causal quantity of interest.

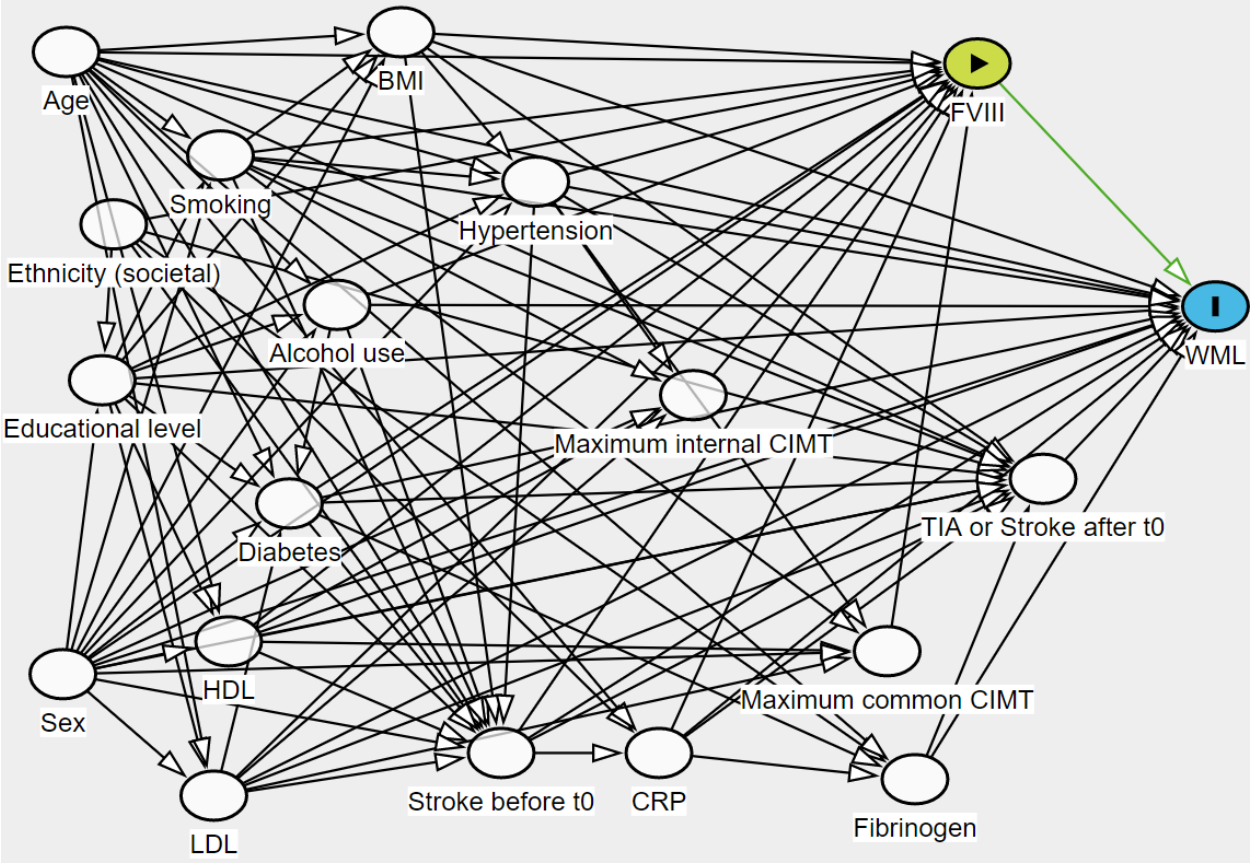

Figure 5. Directed acyclic graph for the CHS subproject example.
Constructed using the Dagitty.[86] Abbreviations: BMI, body mass index; CIMT, carotid intima-media thickness; CRP, C reactive protein; FVIII, coagulation factor FVIII; HDL, high-density lipoprotein, LDL, low-density lipoprotein, t0; "time zero" (study baseline); TIA, transient ischemic attack; WML, white matter lesions (hyperintensities).

Importantly, this DAG represents a much more transparent presentation of analytic and design choices that can readily aid the reader in assessing methodological quality of analyses of causal questions and the validity of the point estimate. Future work should consider using DAGs to inform discussions with coauthor teams pertaining to the adjustment set of confounding variables for causal questions, or even in the project planning stage, to inform study design choice and inclusion/exclusion criteria.
A further suggestion would be to publish DAGs to public repositories; this certainly would have saved me considerable effort in identifying known relationships between confounding variables, the exposures and outcomes of interest across my subprojects. As illustrated in this section, DAGs can be very useful to detect confounding and develop strategies for its control in design and analyses. DAGs find a further useful

application in detecting another type of bias that poses an equal threat to validity, however, is far less intuitive and is perhaps thus often overlooked in published work. I present this bias, called collider stratification bias, using another example from this dissertation project in the following section.

8.2.4. Collider stratification bias

Collider stratification bias is thought to be behind many paradoxes described in the epidemiologic literature, such as the “birth weight paradox” and a barrage of otherwise inexplicable results from studies of COVID-19.[87,88] This particular bias emerges upon blocking a particular type of node called a “collider”, which is represented in a DAG as a node influenced by two other nodes[80,81]. Collider stratification bias can arise either due to structural elements pertaining to study design (e.g., selection bias) or be introduced due to choices made in the analysis phase (e.g., through regression adjustment).

In cases of collider stratification bias, it is possible that an exposure appears protective when it is actually harmful. One well-known example is the so-called BMI (or obesity) paradox[89]; applying this in the context of ischemic stroke, upon selecting only individuals experiencing a first stroke into a study, high BMI can appear to be protective against mortality. This counterintuitive association can arise because first stroke is caused by high BMI as well as a combination of other, more-lethal risk factors.[89] Among the individuals who had a first stroke (i.e., those included in patient cohort studies), the ones with high BMI are less likely to have the more lethal risk factors.[89]

Though this problem is well-described in the epidemiological literature, in the field of applied clinical research in stroke and other domains, very few applications have appropriately addressed this problem.[77,88] This may be because these not particularly intuitive biases can only be readily understood and detected when the analyses are transparently designed and informed by DAGs, as described above.

In an effort to draw further attention to this often-neglected problem and provide an example relevant for the field of stroke research, I will again use a DAG to discuss how collider stratification bias may pose a further possible limitation to my PROSCIS-B subproject. Let us return to the first example from the previous section; the effect of

FVIII measured shortly after stroke (*FVIII*$_1$) on post-stroke outcomes (simplified to *Death*, as before). This research question is represented in Figure 6 by the green arrow. To improve readability, I do not show the node *L,* representing the measured confounders, in this DAG, which does not alter this example.

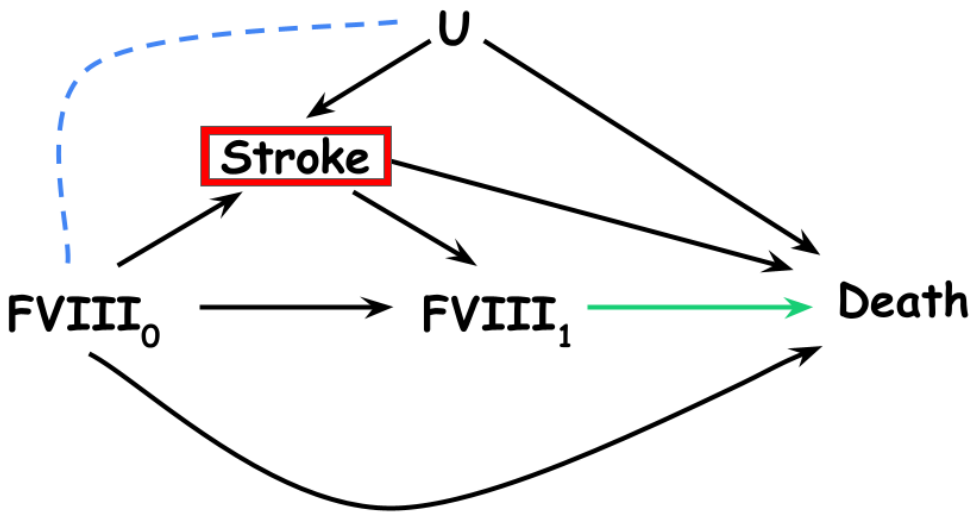

Figure 6. Directed acyclic graph illustrating possible collider stratification bias in the example evaluating FVIII levels and risk of death among patients with ischemic stroke. The green arrow indicates the research question under study. *FVIII*$_1$ denotes the measured exposure (FVIII activity level measured at a given time point after ischemic stroke); *FVIII*$_0$ denotes the unmeasured FVIII activity levels just before the index stroke event; *U* stands for unmeasured variables that affect both *Stroke* and *Death*. The red box represents conditioning due to the study design (i.e. restriction to only ischemic stroke patients indicated by the box around *Stroke*).

In this DAG (Figure 6), there is a new node *U*, which represents unmeasured variables that affect both *Stroke* and *Death*. If these variables exist, then through the restriction of the study sample to stroke patients (represented by the red box around *Stroke* in Figure 6), we induce a further non-causal association between *FVIII*$_1$ and the outcome, *Death*. Indeed, if we do not condition on *Stroke*, the non-causal path *FVIII*$_1$←*FVIII*$_0$→*Stroke*←*U*→*Death* is naturally blocked due to the fact that the stroke node is a "collider" on this path. This is visually indicated in the DAG by the colliding arrowheads coming from the nodes *FVIII*$_0$ and *U* into the node *Stroke* (Figure 6).

In the presence of a collider, a path is naturally blocked, and does not contribute to confounding, which arises from open, unblocked non-causal paths. However, since this study was performed using data from a patient cohort study, *Stroke* is again blocked by design. Blocking a collider results in the induction of a spurious association between the causes of the collider node[78,80,81]. In this case, this is represented through the dashed blue line connecting *U* and $FVIII_0$ after blocking *Stroke*. This association between $FVIII_0$ and *U* results in a non-causal association between the exposure, $FVIII_1$ and the outcome. This non-causal association induced by the collider stratification bias may have introduced an additional bias in our estimation of the causal effect of interest, in addition to the confounding bias described in the previous section.

In general, I suspect this collider stratification bias is common in most biomarker research questions in stroke patient cohorts, since it is highly likely that unmeasured risk factors influencing post-stroke outcomes exist in these settings. However, very few discussions of this topic exist in the stroke literature body, and I could not identify any in the context of coagulation factors. The limitations pertaining to collider stratification bias are on top of other issues frequently encountered in stroke patient cohort studies including small sample sizes, restrictive inclusion criteria and/or interventions limiting generalizability, and extremely short follow-up periods.[43,90,91].

Since stroke patient cohorts provide an important data resource to answer pressing questions pertaining to secondary prevention, I believe further exploration of commonly encountered biases, especially through the explicit presentation of assumptions surrounding causal questions using DAGs, could be of great relevance for future research work.

9. Conclusion

In my PhD research project, I made use of existing data from already established studies with some additional laboratory assays, requiring no new patient enrollment and yet obtaining important information from existing data (with some primary data collection). We approached the research aims from three unique angles: (1) genetic determinants of coagulation factor activities, (2) coagulation factor activities and post-stroke outcomes, and (3) assessing whether the coagulation factor activity levels may be important in covert white matter hyperintensities and cognitive decline in healthy older persons. Our findings expand upon existing findings from preliminary laboratory studies and in-hospital studies of these promising biomarkers using data from large patient cohort and population-based studies.

In addition to the clinical relevance of the findings, there were also some important methodological lessons learned. For example, using DAGs to inform design and analytic strategies, already in the planning phase of a study, could be particularly useful in the context of coagulation factor research, especially since the time point at which the coagulation factors are measured can heavily influence the interpretation of results. Since DAGs are visual tools that encode assumptions surrounding causal relationships under study, regular use of DAGs could also contribute to reproducibility and more readily facilitate replication of existing results. Further research involving stroke patient cohorts should carefully consider potential residual confounding by unmeasured factors prior to the stroke event and induced collider stratification bias.

10. References

1. GBD 2016 Stroke Collaborators. Global, regional, and national burden of stroke, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* **18,** 439–458 (2019).

2. Lackland, D. T., Roccella, E. J., Deutsch, A. F., Fornage, M., George, M. G., Howard, G., Kissela, B. M., Kittner, S. J., Lichtman, J. H., Lisabeth, L. D., Schwamm, L. H., Smith, E. E., Towfighi, A., American Heart Association Stroke Council, Council on Cardiovascular and Stroke Nursing, Council on Quality of Care and Outcomes Research & Council on Functional Genomics and Translational Biology. Factors influencing the decline in stroke mortality: a statement from the American Heart Association/American Stroke Association. *Stroke* **45,** 315–353 (2014).

3. Langhorne, P., Audebert, H. J., Cadilhac, D. A., Kim, J. & Lindsay, P. Stroke systems of care in high-income countries: what is optimal? *Lancet* **396,** 1433–1442 (2020).

4. de Leeuw, F. E., de Groot, J. C., Achten, E., Oudkerk, M., Ramos, L. M., Heijboer, R., Hofman, A., Jolles, J., van Gijn, J. & Breteler, M. M. Prevalence of cerebral white matter lesions in elderly people: a population based magnetic resonance imaging study. The Rotterdam Scan Study. *J. Neurol. Neurosurg. Psychiatry* **70,** 9–14 (2001).

5. Debette, S. & Markus, H. S. The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *BMJ* **341,** c3666 (2010).

6. Mohan, K. M., Wolfe, C. D. A., Rudd, A. G., Heuschmann, P. U., Kolominsky-Rabas, P. L. & Grieve, A. P. Risk and cumulative risk of stroke recurrence: a systematic review and meta-analysis. *Stroke* **42,** 1489–1494 (2011).

7. Amarenco, P., Lavallée, P. C., Monteiro Tavares, L., Labreuche, J., Albers, G. W., Abboud, H., Anticoli, S., Audebert, H., Bornstein, N. M., Caplan, L. R., Correia, M., Donnan, G. A., Ferro, J. M., Gongora-Rivera, F., Heide, W., Hennerici, M. G., Kelly, P. J., Král, M., Lin, H.-F., Molina, C., Park, J. M., Purroy, F., Rothwell, P. M., Segura, T., Školoudík, D., Steg, P. G., Touboul, P.-J., Uchiyama, S., Vicaut, É., Wang, Y., Wong, L. K. S. & TIAregistry.org Investigators. Five-Year Risk of Stroke after TIA or Minor Ischemic Stroke. *N. Engl. J. Med.* **378,** 2182–2190 (2018).

8. Boekholdt, S. M. & Kramer, M. H. H. Arterial thrombosis and the role of thrombophilia. *Semin. Thromb. Hemost.* **33,** 588–596 (2007).

9. Chan, M. Y., Andreotti, F. & Becker, R. C. Hypercoagulable states in cardiovascular disease. *Circulation* **118,** 2286–2297 (2008).

10. Maino, A., Rosendaal, F. R., Algra, A., Peyvandi, F. & Siegerink, B. Hypercoagulability Is a Stronger Risk Factor for Ischaemic Stroke than for Myocardial Infarction: A Systematic Review. *PLoS One* **10,** e0133523 (2015).

11. Astrup, T. The Haemostatic Balance)). *Thrombosis and Haemostasis* **02,** 347–357 (1958).

12. Gaffney, P. J., Edgell, T. A. & Whitton, C. M. The haemostatic balance -- Astrup revisited. *Haemostasis* **29,** 58–71 (1999).

13. Davie, E. W. & Ratnoff, O. D. Waterfall Sequence For Intrinsic Blood Clotting. *Science* **145,** 1310–1312 (1964).

14. Macfarlane, R. G. An Enzyme Cascade In The Blood Clotting Mechanism, And Its Function As A Biochemical Amplifier. *Nature* **202,** 498–499 (1964).

15. Virchow, R. *Phlogose und Thrombose im Gefäßsytem. Gesammelte Abhandlungen zur Wissenschaftlichen Medizin*. (Staatsdrückerei Frankfurt, 1856).

16. Silverberg, M. & Diehl, S. V. The activation of the contact system of human plasma by polysaccharide sulfates. *Ann. N. Y. Acad. Sci.* **516,** 268–279 (1987).

17. Long, A. T., Kenne, E., Jung, R., Fuchs, T. A. & Renné, T. Contact system revisited: an interface between inflammation, coagulation, and innate immunity. *Journal of Thrombosis and Haemostasis* **14,** 427–437 (2016).

18. Renné, T., Schmaier, A. H., Nickel, K. F., Blombäck, M. & Maas, C. In vivo roles of factor XII. *Blood* **120,** 4296–4303 (2012).

19. Schmaier, A. H. & Mccrae, K. R. The plasma kallikrein-kinin system: its evolution from contact activation. *Journal of Thrombosis and Haemostasis* **5,** 2323–2329 (2007).

20. Wu, Y. Contact pathway of coagulation and inflammation. *Thromb. J.* **13,** 17 (2015).

21. Henderson, L. M., Figueroa, C. D., Müller-Esterl, W. & Bhoola, K. D. Assembly of contact-phase factors on the surface of the human neutrophil membrane. *Blood* **84,** 474–482 (1994).

22. Kleinschnitz, C., Stoll, G., Bendszus, M., Schuh, K., Pauer, H.-U., Burfeind, P., Renné, C., Gailani, D., Nieswandt, B. & Renné, T. Targeting coagulation factor XII provides protection from pathological thrombosis in cerebral ischemia without interfering with hemostasis. *J. Exp. Med.* **203,** 513–518 (2006).

23. Renné, T., Pozgajová, M., Grüner, S., Schuh, K., Pauer, H.-U., Burfeind, P., Gailani, D. & Nieswandt, B. Defective thrombus formation in mice lacking coagulation factor XII. *Journal of Experimental Medicine* **202,** 271–281 (2005).

24. Salomon, O., Steinberg, D. M., Koren-Morag, N., Tanne, D. & Seligsohn, U. Reduced incidence of ischemic stroke in patients with severe factor XI deficiency. *Blood* **111,** 4113–4117 (2008).

25. Salomon, O., Steinberg, D. M., Zucker, M., Varon, D., Zivelin, A. & Seligsohn, U. Patients with severe factor XI deficiency have a reduced incidence of deep-vein

thrombosis. *Thromb. Haemost.* **105,** 269–273 (2011).

26. Kraft, P., De Meyer, S. F. & Kleinschnitz, C. Next-generation antithrombotics in ischemic stroke: preclinical perspective on 'bleeding-free antithrombosis'. *J. Cereb. Blood Flow Metab.* **32,** 1831–1840 (2012).

27. Gailani, D., Bane, C. E. & Gruber, A. Factor XI and contact activation as targets for antithrombotic therapy. *J. Thromb. Haemost.* **13,** 1383–1395 (2015).

28. Doggen, C. J. M., Rosendaal, F. R. & Meijers, J. C. M. Levels of intrinsic coagulation factors and the risk of myocardial infarction among men: Opposite and synergistic effects of factors XI and XII. *Blood* **108,** 4045–4051 (2006).

29. Govers-Riemslag, J. W. P., Smid, M., Cooper, J. A., Bauer, K. A., Rosenberg, R. D., Hack, C. E., Hamulyak, K., Spronk, H. M. H., Miller, G. J. & ten Cate, H. The plasma kallikrein-kinin system and risk of cardiovascular disease in men. *J. Thromb. Haemost.* **5,** 1896–1903 (2007).

30. Siegerink, B., Govers-Riemslag, J. W. P., Rosendaal, F. R., Ten Cate, H. & Algra, A. Intrinsic coagulation activation and the risk of arterial thrombosis in young women: results from the Risk of Arterial Thrombosis in relation to Oral contraceptives (RATIO) case-control study. *Circulation* **122,** 1854–1861 (2010).

31. Johansson, K., Jansson, J.-H., Johansson, L., Bylesjö, I., Nilsson, T. K., Eliasson, M., Söderberg, S. & Lind, M. Factor XII as a Risk Marker for Hemorrhagic Stroke: A Prospective Cohort Study. *Cerebrovasc. Dis. Extra* **7,** 84–94 (2017).

32. Zhang, H., Löwenberg, E. C., Crosby, J. R., MacLeod, A. R., Zhao, C., Gao, D., Black, C., Revenko, A. S., Meijers, J. C. M., Stroes, E. S., Levi, M. & Monia, B. P. Inhibition of the intrinsic coagulation pathway factor XI by antisense oligonucleotides: a novel antithrombotic strategy with lowered bleeding risk. *Blood* **116,** 4684–4692 (2010).

33. Tucker, E. I., Marzec, U. M., White, T. C., Hurst, S., Rugonyi, S., McCarty, O. J. T., Gailani, D., Gruber, A. & Hanson, S. R. Prevention of vascular graft occlusion and thrombus-associated thrombin generation by inhibition of factor XI. *Blood* **113,** 936–944 (2009).

34. Crosby, J. R., Marzec, U., Revenko, A. S., Zhao, C., Gao, D., Matafonov, A., Gailani, D., MacLeod, A. R., Tucker, E. I., Gruber, A., Hanson, S. R. & Monia, B. P. Antithrombotic effect of antisense factor XI oligonucleotide treatment in primates. *Arterioscler. Thromb. Vasc. Biol.* **33,** 1670–1678 (2013).

35. Wang, X., Smith, P. L., Hsu, M.-Y., Gailani, D., Schumacher, W. A., Ogletree, M. L. & Seiffert, D. A. Effects of factor XI deficiency on ferric chloride-induced vena cava thrombosis in mice. *J. Thromb. Haemost.* **4,** 1982–1988 (2006).

36. Siegerink, B., Rosendaal, F. R. & Algra, A. Antigen levels of coagulation factor XII, coagulation factor XI and prekallikrein, and the risk of myocardial infarction and ischemic stroke in young women. *J. Thromb. Haemost.* **12,** 606–613 (2014).

37. Büller, H. R., Bethune, C., Bhanot, S., Gailani, D., Monia, B. P., Raskob, G. E., Segers, A., Verhamme, P., Weitz, J. I. & FXI-ASO TKA Investigators. Factor XI antisense oligonucleotide for prevention of venous thrombosis. *N. Engl. J. Med.* **372,** 232–240 (2015).

38. Folsom, A. R., Rosamond, W. D., Shahar, E., Cooper, L. S., Aleksic, N., Nieto, F. J., Rasmussen, M. L. & Wu, K. K. Prospective study of markers of hemostatic function with risk of ischemic stroke. The Atherosclerosis Risk in Communities (ARIC) Study Investigators. *Circulation* **100,** 736–742 (1999).

39. Yap, E. S., Timp, J. F., Flinterman, L. E., van Hylckama Vlieg, A., Rosendaal, F. R., Cannegieter, S. C. & Lijfering, W. M. Elevated levels of factor VIII and subsequent risk of all-cause mortality: results from the MEGA follow-up study. *J. Thromb. Haemost.* **13,** 1833–1842 (2015).

40. Koster, T., Blann, A. D., Briët, E., Vandenbroucke, J. P. & Rosendaal, F. R. Role of clotting factor VIII in effect of von Willebrand factor on occurrence of deep-vein thrombosis. *Lancet* **345,** 152–155 (1995).

41. Kyrle, P. A., Minar, E., Hirschl, M., Bialonczyk, C., Stain, M., Schneider, B., Weltermann, A., Speiser, W., Lechner, K. & Eichinger, S. High plasma levels of factor VIII and the risk of recurrent venous thromboembolism. *N. Engl. J. Med.* **343,** 457–462 (2000).

42. Kerr, R., Stirling, D. & Ludlam, C. A. Interleukin 6 and haemostasis. *Br. J. Haematol.* **115,** 3–12 (2001).

43. Chang, T. R., Albright, K. C., Boehme, A. K., Dorsey, A., Sartor, E. A., Kruse-Jarres, R., Leissinger, C. & Martin-Schild, S. Factor VIII in the setting of acute ischemic stroke among patients with suspected hypercoagulable state. *Clin. Appl. Thromb. Hemost.* **20,** 124–128 (2014).

44. Samai, A., Monlezun, D., Shaban, A., George, A., Dowell, L., Kruse-Jarres, R., Schluter, L., El Khoury, R. & Martin-Schild, S. Von Willebrand factor drives the association between elevated factor VIII and poor outcomes in patients with ischemic stroke. *Stroke* **45,** 2789–2791 (2014).

45. Rohmann, J. L., de Haan, H. G., Algra, A., Vossen, C. Y., Rosendaal, F. R. & Siegerink, B. Genetic determinants of activity and antigen levels of contact system factors. *J. Thromb. Haemost.* **17,** 157–168 (2019).

46. Rohmann, J. L., Huo, S., Sperber, P. S., Piper, S. K., Rosendaal, F. R., Heuschmann, P. U., Endres, M., Liman, T. G. & Siegerink, B. Coagulation factor XII, XI, and VIII activity levels and secondary events after first ischemic stroke. *J. Thromb. Haemost.* **18**, 3316-3324 (2020).

47. Rohmann, J. L., Longstreth, W. T., Jr, Cushman, M., Fitzpatrick, A. L., Heckbert, S. R., Rice, K., Rosendaal, F. R., Sitlani, C. M., Psaty, B. M. & Siegerink, B. Coagulation factor VIII, white matter hyperintensities and cognitive function: Results from the Cardiovascular Health Study. *PLoS One* **15,** e0242062 (2020).

48. Tanis, B. C., van den Bosch, M. A., Kemmeren, J. M., Cats, V. M., Helmerhorst, F. M., Algra, A., van der Graaf, Y. & Rosendaal, F. R. Oral contraceptives and the risk of myocardial infarction. *N. Engl. J. Med.* **345,** 1787–1793 (2001).

49. Kemmeren Jeanet M., Tanis Bea C., van den Bosch Maurice A.A.J., Bollen Edward L.E.M., Helmerhorst Frans M., van der Graaf Yolanda, Rosendaal Frits R. & Algra Ale. Risk of Arterial Thrombosis in Relation to Oral Contraceptives (RATIO) Study. *Stroke* **33,** 1202–1208 (2002).

50. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21,** 263–265 (2005).

51. Siegerink, B., Rosendaal, F. R. & Algra, A. High-molecular-weight kininogen and the risk of a myocardial infarction and ischemic stroke in young women: the RATIO case-control study. *J. Thromb. Haemost.* **10,** 2409–2412 (2012).

52. Tanis, B., Algra, A., van der Graaf, Y., Helmerhorst, F. & Rosendaal, F. Procoagulant factors and the risk of myocardial infarction in young women. *Eur. J. Haematol.* **77,** 67–73 (2006).

53. Sinnwell, J. P., Schaid, D. J. & Sinnwell. haplo.stats: Statistical Analysis of Haplotypes with Traits and Covariates when Linkage Phase is Ambiguous. (2016). at <https://CRAN.R-project.org/package=haplo.stats>

54. Liman, T. G., Zietemann, V., Wiedmann, S., Jungehuelsing, G. J., Endres, M., Wollenweber, F. A., Wellwood, I., Dichgans, M. & Heuschmann, P. U. Prediction of vascular risk after stroke - protocol and pilot data of the Prospective Cohort with Incident Stroke (PROSCIS). *Int. J. Stroke* **8,** 484–490 (2013).

55. Fried, L. P., Borhani, N. O., Enright, P., Furberg, C. D., Gardin, J. M., Kronmal, R. A., Kuller, L. H., Manolio, T. A., Mittelmark, M. B. & Newman, A. The Cardiovascular Health Study: design and rationale. *Ann. Epidemiol.* **1,** 263–276 (1991).

56. Cushman, M., Cornell, E. S., Howard, P. R., Bovill, E. G. & Tracy, R. P. Laboratory methods and quality assurance in the Cardiovascular Health Study. *Clin. Chem.* **41,** 264–270 (1995).

57. Longstreth, W. T., Jr, Arnold, A. M., Beauchamp, N. J., Jr, Manolio, T. A., Lefkowitz, D., Jungreis, C., Hirsch, C. H., O'Leary, D. H. & Furberg, C. D. Incidence, manifestations, and predictors of worsening white matter on serial cranial magnetic resonance imaging in the elderly: the Cardiovascular Health Study. *Stroke* **36,** 56–61 (2005).

58. Longstreth, W. T., Jr, Manolio, T. A., Arnold, A., Burke, G. L., Bryan, N., Jungreis, C. A., Enright, P. L., O'Leary, D. & Fried, L. Clinical correlates of white matter findings on cranial magnetic resonance imaging of 3301 elderly people. The Cardiovascular Health Study. *Stroke* **27,** 1274–1282 (1996).

59. Manolio, T. A., Kronmal, R. A., Burke, G. L., Poirier, V., O'Leary, D. H., Gardin, J. M., Fried, L. P., Steinberg, E. P. & Bryan, R. N. Magnetic resonance abnormalities

and cardiovascular disease in older adults. The Cardiovascular Health Study. *Stroke* **25,** 318–327 (1994).

60. Teng, E. L. & Chui, H. C. The Modified Mini-Mental State (3MS) examination. *J. Clin. Psychiatry* **48,** 314–318 (1987).

61. Salthouse, T. A. The role of memory in the age decline in digit-symbol substitution performance. *J. Gerontol.* **33,** 232–238 (1978).

62. Arnold, A. M., Newman, A. B., Dermond, N., Haan, M. & Fitzpatrick, A. Using Telephone and Informant Assessments to Estimate Missing Modified Mini-Mental State Exam Scores and Rates of Cognitive Decline. *Neuroepidemiology* **33,** 55–65 (2009).

63. Williams, R. Generalized Ordered Logit/ Partial Proportional Odds Models for Ordinal Dependent Variable. *The Stata Journal* **6,** 58–82 (2006).

64. Williams, R. Understanding and interpreting generalized ordered logit models. *J. Math. Sociol.* **40,** 7–20 (2016).

65. Glymour, M. M., Weuve, J., Berkman, L. F., Kawachi, I. & Robins, J. M. When is baseline adjustment useful in analyses of change? An example with education and cognitive change. *Am. J. Epidemiol.* **162,** 267–278 (2005).

66. Huffman, J. E. Examining the current standards for genetic discovery and replication in the era of mega-biobanks. *Nat. Commun.* **9,** 5054 (2018).

67. Replicating genotype–phenotype associations. *Nature* **447,** 655–660 (2007).

68. Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533,** 452–454 (2016).

69. Sennblad, B., Basu, S., Mazur, J., Suchon, P., Martinez-Perez, A., van Hylckama Vlieg, A., Truong, V., Li, Y., Gådin, J. R., Tang, W., Grossman, V., de Haan, H. G., Handin, N., Silveira, A., Souto, J. C., Franco-Cereceda, A., Morange, P.-E., Gagnon, F., Soria, J. M., Eriksson, P., Hamsten, A., Maegdefessel, L., Rosendaal, F. R., Wild, P., Folsom, A. R., Trégouët, D.-A. & Sabater-Lleal, M. Genome-wide association study with additional genetic and post-transcriptional analyses reveals novel regulators of plasma factor XI levels. *Hum. Mol. Genet.* **26,** 637–649 (2017).

70. Siegler, J. E., Samai, A., Albright, K. C., Boehme, A. K. & Martin-Schild, S. Factoring in Factor VIII With Acute Ischemic Stroke. *Clin. Appl. Thromb. Hemost.* **21,** 597–602 (2015).

71. Tóth, N. K., Székely, E. G., Czuriga-Kovács, K. R., Sarkady, F., Nagy, O., Lánczi, L. I., Berényi, E., Fekete, K., Fekete, I., Csiba, L. & Bagoly, Z. Elevated Factor VIII and von Willebrand Factor Levels Predict Unfavorable Outcome in Stroke Patients Treated with Intravenous Thrombolysis. *Front. Neurol.* **8,** 721 (2017).

72. Ceciliani, F., Giordano, A. & Spagnolo, V. The systemic reaction during inflammation: the acute-phase proteins. *Protein Pept. Lett.* **9,** 211–223 (2002).

73. Elkind, M. S. V., Coates, K., Tai, W., Paik, M. C., Boden-Albala, B. & Sacco, R. L. Levels of acute phase proteins remain stable after ischemic stroke. *BMC Neurol.* **6,** 37 (2006).

74. Richter, D., Weber, R., Eyding, J., Bartig, D., Misselwitz, B., Grau, A., Hacke, W. & Krogias, C. Acute ischemic stroke care in Germany - further progress from 2016 to 2019. *Neurol Res Pract* **3,** 14 (2021).

75. Hillmann, S., Wiedmann, S., Rücker, V., Berger, K., Nabavi, D., Bruder, I., Koennecke, H.-C., Seidel, G., Misselwitz, B., Janssen, A., Burmeister, C., Matthis, C., Busse, O., Hermanek, P., Heuschmann, P. U. & German Stroke Register Study Group (ADSR). Stroke unit care in germany: the german stroke registers study group (ADSR). *BMC Neurol.* **17,** 49 (2017).

76. Gail, M. H., Altman, D. G., Cadarette, S. M., Collins, G., Evans, S. J., Sekula, P., Williamson, E. & Woodward, M. Design choices for observational studies of the effect of exposure on disease incidence. *BMJ Open* **9,** e031031 (2019).

77. Tennant, P. W. G., Murray, E. J., Arnold, K. F., Berrie, L., Fox, M. P., Gadd, S. C., Harrison, W. J., Keeble, C., Ranker, L. R., Textor, J., Tomova, G. D., Gilthorpe, M. S. & Ellison, G. T. H. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int. J. Epidemiol.* (2020). doi:10.1093/ije/dyaa213

78. Greenland, S., Pearl, J. & Robins, J. M. Causal diagrams for epidemiologic research. *Epidemiology* **10,** 37–48 (1999).

79. Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. (Morgan Kaufmann, 1988).

80. Hernán, M. A. & Robins, J. M. Causal inference: what if. (2020).

81. Pearl, J. & Mackenzie, D. *The Book of Why: The New Science of Cause and Effect*. (Penguin UK, 2018).

82. Hernán, M. A. The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data. *Am. J. Public Health* **108,** 616–619 (2018).

83. Hernán, M. A., Hernández-Díaz, S., Werler, M. M. & Mitchell, A. A. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am. J. Epidemiol.* **155,** 176–184 (2002).

84. Howards, P. P., Schisterman, E. F., Poole, C., Kaufman, J. S. & Weinberg, C. R. 'Toward a clearer definition of confounding' revisited with directed acyclic graphs. *Am. J. Epidemiol.* **176,** 506–511 (2012).

85. Hernán, M. A., Hernández-Díaz, S. & Robins, J. M. A structural approach to selection bias. *Epidemiology* **15,** 615–625 (2004).

86. Textor, J., van der Zander, B., Gilthorpe, M. S., Liśkiewicz, M. & Ellison, G. T. H.

Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International Journal of Epidemiology* dyw341 (2017). doi:10.1093/ije/dyw341

87. Hernández-Díaz, S., Schisterman, E. F. & Hernán, M. A. The birth weight 'paradox' uncovered? *Am. J. Epidemiol.* **164,** 1115–1120 (2006).

88. Griffith, G. J., Morris, T. T., Tudball, M. J., Herbert, A., Mancano, G., Pike, L., Sharp, G. C., Sterne, J., Palmer, T. M., Davey Smith, G., Tilling, K., Zuccolo, L., Davies, N. M. & Hemani, G. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat. Commun.* **11,** 5749 (2020).

89. Banack, H. R. & Kaufman, J. S. The 'Obesity Paradox' Explained. *Epidemiology* **24,** 461 (2013).

90. de Montmollin, E., Terzi, N., Dupuis, C., Garrouste-Orgeas, M., da Silva, D., Darmon, M., Laurent, V., Thiéry, G., Oziel, J., Marcotte, G., Gainnier, M., Siami, S., Sztrymf, B., Adrie, C., Reignier, J., Ruckly, S., Sonneville, R., Timsit, J.-F. & OUTCOMEREA Study Group. One-year survival in acute stroke patients requiring mechanical ventilation: a multicenter cohort study. *Ann. Intensive Care* **10,** 53 (2020).

91. Haeusler, K. G., Grittner, U., Fiebach, J. B., Endres, M., Krause, T. & Nolte, C. H. HEart and BRain interfaces in Acute ischemic Stroke (HEBRAS)--rationale and design of a prospective oberservational cohort study. *BMC Neurol.* **15,** 213 (2015).

11. Statutory Declaration

I, Jessica L. Rohmann, by personally signing this document in lieu of an oath, hereby affirm that I prepared the submitted dissertation on the topic *"The role of hypercoagulability in neurovascular disorders: Insights from clinical epidemiology"* (*"Die Rolle der Hyperkoagulabilität bei neurovaskulären Erkrankungen: Einblicke aus der klinischen Epidemiologie"*), independently and without the support of third parties, and that I used no other sources and aids than those stated.

All parts which are based on the publications or presentations of other authors, either in letter or in spirit, are specified as such in accordance with the citing guidelines. The sections on methodology (in particular regarding practical work, laboratory regulations, statistical processing) and results (in particular regarding figures, charts and tables) are exclusively my responsibility.

Furthermore, I declare that I have correctly marked all of the data, the analyses, and the conclusions generated from data obtained in collaboration with other persons, and that I have correctly marked my own contribution and the contributions of other persons (cf. declaration of contribution). I have correctly marked all texts or parts of texts that were generated in collaboration with other persons.

My contributions to any publications to this dissertation correspond to those stated in the below joint declaration made together with the supervisor. All publications created within the scope of the dissertation comply with the guidelines of the ICMJE (International Committee of Medical Journal Editors; www.icmje.org) on authorship. In addition, I declare that I shall comply with the regulations of Charité – Universitätsmedizin Berlin on ensuring good scientific practice.

I declare that I have not yet submitted this dissertation in identical or similar form to another Faculty. The significance of this statutory declaration and the consequences of a false statutory declaration under criminal law (Sections 156, 161 of the German Criminal Code) are known to me.

Date: _____ Signature: _____

12. Declaration of your own contribution to the publications

Jessica L. Rohmann contributed the following to the below listed publications:

Publication 1: **Rohmann JL**, de Haan HG, Algra A, Vossen CY, Rosendaal FR, Siegerink B. Genetic determinants of activity and antigen levels of contact system factors. J Thromb Haemost. 2019 Jan.

Journal Impact Factor (2016): 5.287

Contribution: Jessica L. Rohmann, together with coauthor HG de Haan performed all genetic and statistical analyses for this study. Jessica L. Rohmann independently created the participant flowchart (Fig 1). She assisted HG de Haan with the selection of the tag SNPs using HapMap reference data in Haploview. Jessica L. Rohmann performed the descriptive statistical analysis to summarize the participants' characteristics and determined major/minor allele frequencies within the study population by herself. She also independently ran all the age-adjusted and conditional regression models for the quantitative trait associations with the coagulation factor levels as outcomes, as well as the logistic regression models with the disease outcomes. The exploratory analyses investigating pleiotropy versus potential mediation, described on page 161 of the paper, were performed by Jessica L. Rohmann in consultation with HG de Haan. Jessica L. Rohmann created the haplotypes and performed the haplotype analyses with the haplo.stats package in R together with coauthor HG de Haan. Jessica L. Rohmann independently created the Tables 1, 2, 3, 4, 5, S1b, S2, S3, S4, S5, S6 and S7, in which these aforementioned analyses were reported. She further assisted HG de Haan with the creation and design of the regional association plots (Figures S1 and S2). Together with the project supervisor, Bob Siegerink, and another coauthor, she interpreted and contextualized the results. Jessica L. Rohmann wrote the first draft of the manuscript by herself. She drafted the response to peer reviewers and revised the manuscript for final publication, incorporating additional reviewer and co-author feedback.

Publication 2: **Rohmann JL**, Huo S, Sperber PS, Piper SK, Rosendaal FR, Heuschmann PU, Endres M, Liman TG, Siegerink B. Coagulation factor XII, XI, and VIII

activity levels and secondary events after first ischemic stroke. J Thromb Haemost. 2020 Dec.

Journal Impact Factor (2018): 4.662

Contribution: Jessica L. Rohmann conceived this secondary data analysis project together with her supervisor Bob Siegerink. Together, they developed the study design and statistical analysis plan. Jessica L. Rohmann sorted and prepared all frozen blood plasma samples for analysis. She traveled to the Leiden University Medical Center as a part of a visiting student exchange to oversee and collect the requisite data from the coagulation factor activity measurements, as a part of a collaboration with the Dept. of Thrombosis and Hemostasis. Jessica L. Rohmann wrote the Stata code for all statistical analyses presented in the paper, which was then checked by her supervisor, Bob Siegerink. Together with the coauthor team, she discussed and interpreted the findings in light of existing literature evidence. She made the two figures and both tables in the publication by herself, created the supplemental material, and wrote the first draft of the manuscript independently. She incorporated feedback from all coauthors and submitted the final version of the manuscript to the journal. She also managed all correspondence with the journal, and coordinated the revision and resubmission process.

Publication 3: **Rohmann JL**, Longstreth WT Jr, Cushman M, Fitzpatrick AL, Heckbert SR, Rice K, Rosendaal FR, Sitlani CM, Psaty BM, Siegerink B. Coagulation factor VIII, white matter hyperintensities and cognitive function: Results from the Cardiovascular Health Study. PLoS One. 2020 Nov.

Journal Impact Factor (2018): 2.776

Contribution: Jessica L. Rohmann drafted a complete project proposal for this third subproject, including a statistical analysis plan by herself, which was then reviewed by her supervisor, Bob Siegerink. Jessica L. Rohmann established the collaboration with the Cardiovascular Health Study sponsor (co-author Bruce Psaty) and the other CHS coauthors. She submitted this proposal to the CHS's "Publications and Presentations" committee, for which she received full approval. She selected the requisite variables needed for her analysis and set up a formal data transfer agreement with the Charité to obtain the requisite data. In parallel, together with Bob Siegerink, she drafted an ethics

review application for the secondary data analysis project. Jessica L. Rohmann independently produced a first draft of the statistical analysis code for all analyses presented in the published manuscript using the program Stata. She later incorporated the feedback from code review by her supervisor Bob Siegerink. She then ran all the analyses by herself. After discussing the interpretation of the findings with the international coauthor team, Jessica L. Rohmann independently drafted the first version of the scientific manuscript. She also created all 5 tables and the 2 figures in the manuscript by herself. She further served as the corresponding author for the manuscript submission and coordinated all communication with the journal as well as her coauthors throughout the revision process.

_____

Signature, date and stamp of first supervising university professor / lecturer

_____

Signature of doctoral candidate

13. List of selected publications

13.1. Publication 1: Genetic determinants of activity and antigen levels of contact system factors

Rohmann JL, de Haan HG, Algra A, Vossen CY, Rosendaal FR, Siegerink B. Genetic determinants of activity and antigen levels of contact system factors. J Thromb Haemost. 2019 Jan;17(1):157–68. Available from: http://dx.doi.org/10.1111/jth.14307

The extract from ISI Web of Knowledge Journal Summary List and the original published version of the article follow on next 13 pages. I have obtained permission to reproduce this work here from the publisher "John Wiley and Sons".

Journal Data Filtered By: **Selected JCR Year: 2016** Selected Editions: SCIE,SSCI
Selected Categories: **"HEMATOLOGY"** Selected Category Scheme: WoS
Gesamtanzahl: 70 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	CIRCULATION RESEARCH	49,784	13.965	0.079890
2	BLOOD	161,962	13.164	0.313600
3	LEUKEMIA	23,538	11.702	0.059800
4	HAEMATOLOGICA	15,075	7.702	0.040460
5	Lancet Haematology	571	7.123	0.002680
6	ARTERIOSCLEROSIS THROMBOSIS AND VASCULAR BIOLOGY	32,950	6.607	0.051690
7	Journal of Hematology & Oncology	2,879	6.350	0.007920
8	BLOOD REVIEWS	2,380	6.342	0.005310
9	TRANSFUSION MEDICINE REVIEWS	1,254	5.745	0.002760
10	BRITISH JOURNAL OF HAEMATOLOGY	23,280	5.670	0.041040
11	THROMBOSIS AND HAEMOSTASIS	17,662	5.627	0.029740
12	STEM CELLS	20,822	5.599	0.038100
13	JOURNAL OF THROMBOSIS AND HAEMOSTASIS	18,059	5.287	0.041260
14	AMERICAN JOURNAL OF HEMATOLOGY	8,776	5.275	0.021330
15	JOURNAL OF CEREBRAL BLOOD FLOW AND METABOLISM	16,998	5.081	0.029520
16	CRITICAL REVIEWS IN ONCOLOGY HEMATOLOGY	6,296	4.971	0.011240
17	BIOLOGY OF BLOOD AND MARROW TRANSPLANTATION	9,904	4.704	0.025270
18	SEMINARS IN HEMATOLOGY	2,157	4.042	0.003430
19	JOURNAL OF LEUKOCYTE BIOLOGY	17,441	4.018	0.023810
20	BONE MARROW TRANSPLANTATION	11,896	3.874	0.021220
21	SEMINARS IN THROMBOSIS AND HEMOSTASIS	4,054	3.629	0.007400
22	HAEMOPHILIA	6,137	3.569	0.012260
23	STEM CELLS AND DEVELOPMENT	7,446	3.562	0.018710
24	TRANSFUSION	12,469	3.386	0.021790
25	HEMATOLOGY-ONCOLOGY CLINICS OF NORTH AMERICA	2,120	3.226	0.004840
26	CYTOTHERAPY	4,952	3.203	0.008800

Journal of Thrombosis and Haemostasis, **17**: 157–168 DOI: 10.1111/jth.14307

ORIGINAL ARTICLE

Genetic determinants of activity and antigen levels of contact system factors

J. L. ROHMANN,*† H. G. DE HAAN,‡ A. ALGRA,‡§¶ C. Y. VOSSEN,‡ F. R. ROSENDAAL‡ and B. SIEGERINK*‡

*Center for Stroke Research Berlin, Charité – Universitätsmedizin Berlin; †Institute of Public Health, Charité – Universitätsmedizin Berlin, Berlin, Germany; ‡Department of Clinical Epidemiology, Leiden University Medical Center, Leiden; §Department of Neurology and Neurosurgery, Brain Center Rudolph Magnus, University Medical Center Utrecht and Utrecht University; and ¶Julius Centre for Health Sciences and Primary Care, University Medical Center Utrecht and Utrecht University, Utrecht, the Netherlands

To cite this article: Rohmann JL, de Haan HG, Algra A, Vossen CY, Rosendaal FR, Siegerink B. Genetic determinants of activity and antigen levels of contact system factors. *J Thromb Haemost* 2019; **17**: 157–68.

Essentials

- Genetic variation may provide valuable insight into the role of the contact system in thrombosis.
- Explored associations of genetic variants with activity, antigen, and disease in RATIO study.
- Two novel loci were identified: *KLKB1* rs4253243 for prekallikrein; *KNG1* rs5029980 for HMWK levels.
- Contact system variants and haplotypes were not associated with myocardial infarction or stroke.

Summary. *Background:* The complex, interdependent contact activation system has been implicated in thrombotic disease, although few genetic determinants of levels of proteins from this system are known. *Objectives:* Our primary aim was to study the influence of common *F11*, *F12*, *KLKB1*, and *KNG1* variants on factor (F) XI activity and FXI, FXII, prekallikrein (PK) and high-molecular-weight kininogen (HMWK) antigen levels, as well as the risk of myocardial infarction and ischemic stroke. *Patients/methods:* We analyzed samples from all 630 healthy participants, 182 ischemic stroke patients and 216 myocardial infarction patients in the RATIO case–control study of women aged < 50 years. Forty-three tagging single nucleotide variants (SNVs) were genotyped to represent common genetic variation in the contact system genes. Antigen and activity levels were measured with sandwich-ELISA-based and one-stage clotting assays. We performed single variant, age-adjusted, linear regression analyses per trait and disease phenotype, assuming additive inheritance and determined conditionally independent associations. Haplotypes based on the lead SNV and all conditionally independent SNVs were tested for association with traits and disease. *Results:* We identified two novel associations of *KLKB1* SNV rs4253243 with PK antigen ($\beta_{conditional} = -12.38$; 95% CI, −20.07 to −4.69) and *KNG1* SNV rs5029980 with HMWK antigen ($\beta_{conditional} = 5.86$; 95% CI, 2.40–9.32) and replicated previously reported associations in a single study. Further analyses probed whether the observed associations were indicative of linkage, pleiotropic effects or mediation. No individual SNVs or haplotypes were associated with the disease outcomes. *Conclusion:* This study adds to current knowledge of how genetic variation influences contact system protein levels and clarifies interdependencies.

Keywords: factor XI; factor XII; genetic variation; kininogen, high-molecular-weight; prekallikrein.

Introduction

The contact activation system, upstream in the intrinsic coagulation pathway, plays an important but until recently underappreciated role in pathological thrombus formation [1–4]. This system's numerous interdependencies make it complex: upon binding to a negatively charged surface, coagulation factor XII (FXII) is activated, which in turn activates factor XI (FXI), triggering the downstream procoagulant intrinsic coagulation cascade [5,6]. Simultaneously, FXIIa sets the proinflammatory kinin-kallikrein system into motion by activating prekallikrein (PK) to kallikrein, which in turn cleaves high-molecular-weight kininogen (HMWK) into bradykinin [7]. Non-enzymatic HMWK molecules serve as a

Correspondence: Bob Siegerink, Center for Stroke Research Berlin, Charité – Universitätsmedizin Berlin, Charitеplatz 1, 10117 Berlin, Germany
Tel.: +49 30 450 560621
E-mail: bob.siegerink@charite.de

Received: 2 April 2018
Manuscript handled by: J. Douketis
Final decision: J. Douketis, 17 September 2018

cofactor in the activation of both FXI and PK, for example by anchoring FXI and PK to neutrophil membranes [6,8].

High plasma levels of contact system factors, especially FXI, contribute to a hypercoagulable state and have been associated with increased risk of venous [9–12] and arterial thrombosis [13–16]. Several genetic variants of these factors have also been implicated in thrombotic diseases, but their means of exerting influence often remain unclear [17–22]. A considerable amount of variation in both activity and antigen levels of these coagulation factors in healthy individuals can be attributed to genetic variation, with heritability estimates of 67% for FXII levels and 45% for FXI [23]. Effects of single nucleotide variants (SNVs) in *F12* (chromosome(chr) 5) and *F11* genes (chr 4) on circulating levels of FXII and FXI activity and antigen, respectively, have previously been explored [24–26]. Less is known about the influence of genetic variation in *KLKB1* (kallikrein B1, chr 4) on PK levels and in *KNG1* (chr 3) on HMWK levels, apart from recent protein genome-wide association study (GWAS) findings [27]. Furthermore, some across-trait associations within the contact system are known. Recent studies have shown associations of *KNG1* SNVs with FXI levels [25,26], as well as *KLKB1* and *KNG1* SNVs with PK and HMWK levels [27,28]. Although generally less well studied, these may provide valuable insights into regulatory mechanisms.

In this study, we explore associations between genetic variation in *F12*, *F11*, *KLKB1* and *KNG1* and FXII, FXI, PK and HMWK levels via tagging SNVs as well as haplotype analyses. We additionally explore whether this genetic variation, especially in haplotype combinations, is associated with ischemic stroke (IS) and myocardial infarction (MI) in young female Dutch RATIO study participants.

Materials and methods

Study design and participants

The Risk of Arterial Thrombosis In relation to Oral contraceptives (RATIO) multicenter, population-based, case–control study has been previously described in detail [29–31]. Briefly, the first phase of the study assessed risk factors for MI and IS in young women aged 18–50 years. Cases were recruited from 16 participating hospitals in the Netherlands between 1990 and 1995. Control women with no history of arterial thrombosis were recruited using random digit dialing and frequency matched on age, calendar year of the event and area of residence. Blood samples and buccal swabs were collected in a second phase (1998–2002) and DNA was harvested and stored for later analyses. All RATIO study participants provided written informed consent and the ethics committees of the participating hospitals approved the study.

SNV selection and genotyping

For each of the genes, using HapMap reference data from the CEU population (build GRCh37), tag SNVs were selected using Haploview software (version 4.2) [32] based on standard settings and a minor allele frequency (MAF) of $> 5\%$ and $r^2 > 0.80$. Gene boundaries were not extended. Complete tagger results are reported in Table S1A. In total, 43 SNVs, two in *F12*, 11 in *F11*, 10 in *KLKB1* and 20 in *KNG1*, were selected to represent common genetic variation in these genes.

Genotyping was performed by technicians who were blinded to the case/control status of the participants using KASPar SNP genotyping assays according to the manufacturer's instructions (KBioscience, Hoddesdon, UK). Because of failures in assay design, for three SNVs, TaqMan genotyping assays were used (Thermo Fisher Scientific, Waltham, MA, USA). Assay design failed again for *KNG1* SNV rs5030072 and thus this SNV was not analyzed. The average call rate per SNV in control subjects with DNA available ($n = 755$) was 94.3%, and *KNG1* SNV rs266723 had the lowest call rate (90.5%) of all SNVs assayed (Table S1B).

Contact system protein measurements

Blood samples were obtained from 638 control subjects 5 years after the index date, on average. Antigen levels of FXI, PK, HMWK and FXII (FXI:ag, PK:ag, HWMK:ag and FXII:ag) were measured using sandwich ELISA-based assays with polyclonal antibodies in duplicate and expressed as percentages of normal pooled plasma, as previously described [15,33]. Inter-assay coefficients of variation were 9.3%, 4.9%, 9.2% and 12.0%, respectively. FXI activity (FXI:C) was additionally measured in duplicate in MI patients and control subjects using a one-stage clotting assay and FXI-deficient plasma, with a coefficient of variation $< 10\%$, as detailed elsewhere [34].

Statistical analyses

Study population characteristics were summarized using mean and median values (with standard deviations or ranges) or absolute numbers and percentages, as appropriate. Antigen and activity measurements were normally distributed and did not require transformation. Departures from the Hardy–Weinberg equilibrium were tested using χ^2 goodness-of-fit tests with a continuity correction or exact tests if expected counts were below five (Table S1B).

To investigate associations between SNVs and the five traits, we performed single variant analyses in control subjects only using age-adjusted linear regression models. An additive model of inheritance was assumed in all quantitative trait loci (QTL) analyses to investigate the change in respective contact system factor level per copy

of the minor allele (β) to maximize power. To account for multiple testing, we used a Bonferroni-corrected overall significance level of 0.00116 based on the total number of variants tested per trait ($n = 43$).

When multiple variants within one gene were associated with the same trait, we performed conditional regression analyses to identify associations conditionally independent of the lead SNV. SNVs with persisting associations were mutually adjusted and only remaining contributing signals were carried forward to the haplotype analyses with the lead SNV. To assess associations between SNVs and other traits, we first adjusted for the levels of the encoding gene's trait. IBM SPSS Statistics v23 was used for all variant association analyses.

We constructed haplotypes for each gene, including the lead SNV and any secondary signals. We used the haplo.em function of the haplo.stats R package to estimate haplotypes per gene for each individual, accounting for the unphased nature of our data, and the haplo.glm function to run age-adjusted regressions for all haplotypes with a frequency > 1% in control subjects [35]. For each gene, the haplotype containing major allele copies was used as a reference and effect size (β), standard error (SE) and *P*-value estimates were calculated for the other inherited combinations.

All variants associated with at least one of the traits following conditional analysis as well as the previously constructed haplotypes were additionally tested for association with risk of MI and IS using logistic regression models, assuming an additive mode of inheritance. We adjusted these models for the matching variables including age, area of residence and index year.

Results

In total, 767 controls, 218 MI patients and 190 IS patients (including 50 additional IS patients) entered the second phase of the RATIO study. As shown in Fig. 1, 755 control subjects, 216 MI patients and 182 IS patients with at least one SNV call were included in the variant-disease association analyses. Only control subjects with DNA and blood plasma samples were included in the protein QTL analyses ($n = 630$). Control subjects' baseline characteristics are displayed in Table 1 (MI and IS patients' characteristics in Table S2).

Quantitative trait associations

F11 variation Five of the 11 *F11* SNVs showed associations with FXI activity below the Bonferroni-corrected threshold (Table 2; remainder in Table S3). Lead SNV rs4253399 was associated with an increase in FXI:C of 9.24 units (95% confidence interval [CI], 6.69–11.79; $P = 3.32 \times 10^{-12}$; MAF = 38.9%) per minor allele (G) copy, measured as percentage of activated normal pooled plasma, explaining 7.8% of the total variance in FXI:C. SNV rs4253399 is in relatively high linkage disequilibrium (LD) ($r^2 = 0.77$) with rs2289252 in 1000Genomes data of Europeans, which has been associated with FXI:ag [18], FXI:C and activated partial thromboplastin time (APTT) [36].

Table 3 shows association results conditional on the lead SNV rs4253399. Of the three remaining associations with FXI:C, the strongest was with rs1593 ($\beta_{conditional} = -5.65$; 95% CI, −9.46 to −1.83; $P_{conditional} = 0.004$; MAF = 12.2%), previously associated with APTT [37]. Furthermore, rs4253421, tagged by rs1593, was associated with a decrease in FXI antigen/activity levels in a pooled GWAS [26]. Additional adjustment for rs1593 diminished remaining associations.

Four of the *F11* SNVs significantly associated with FXI activity as well as rs4253430 were associated with FXI antigen (Table 2). The lead SNV rs2036914 (β = −11.38; 95% CI, −14.18 to −8.58; $P = 7.87 \times 10^{-15}$; MAF = 46.3%) explained 9.1% of variance in FXI:ag. SNV rs2036914 has been shown to be associated with FXI:ag [18], FXI:C and prolonged APTT [36], and tagged rs4241824, which has been associated with FXI:ag [12, 25]. After conditioning on the lead SNV, two associations with FXI:ag remained: rs1593 and rs4253399, the same two SNVs from the FXI:C trait analyses. Both signals remained after additional mutual adjustment (rs1593: β = −7.47; 95% CI, −11.99 to −2.96; $P = 0.001$; rs4253399: β = 6.11; 95% CI, 2.02–10.20; $P = 0.003$).

All *F11* SNVs were additionally tested for associations with the other three traits: PK:ag, HMWK:ag and FXII:ag (Table S4); only rs4253406 showed suggestive evidence of an association with PK:ag.

Overall, the effect sizes observed for associations between *F11* SNVs and FXI:ag were larger in magnitude than those for FXI:C. Exploratory analyses shown in Table 4 indicate the majority of the associations with FXI:C were greatly attenuated upon adjustment for antigen levels. Although the association between rs4253399 and FXI:C diminished upon adjustment for FXI:ag, a conditional association not explained by FXI:ag ($\beta_{adjusted} = 3.07$; 95% CI, 0.95–5.20; $P_{adjusted} = 0.005$) remained. Reciprocal adjustment of the association with FXI:ag for FXI:C resulted in a modest effect size reduction.

KLKB1 variation Six of the 10 *KLKB1* SNVs tested showed below-threshold associations with PK:ag, with effect size magnitudes ranging from 6.55 to 16.93 (Table 2; above-threshold associations in Table S3). Lead variant rs2304595 (β = 10.33; 95% CI, 6.65–14.02; $P = 5.58 \times 10^{-8}$; MAF = 43.2%) explained 5.1% of the total variance in PK:ag and tagged rs1511802, previously reported to be associated with PK levels [27]. SNV rs4253243 had the largest effect size of all *KLKB1* SNVs tested; after conditioning on rs2304595, it was the only remaining association with PK:ag ($\beta_{conditional} = -12.38$;

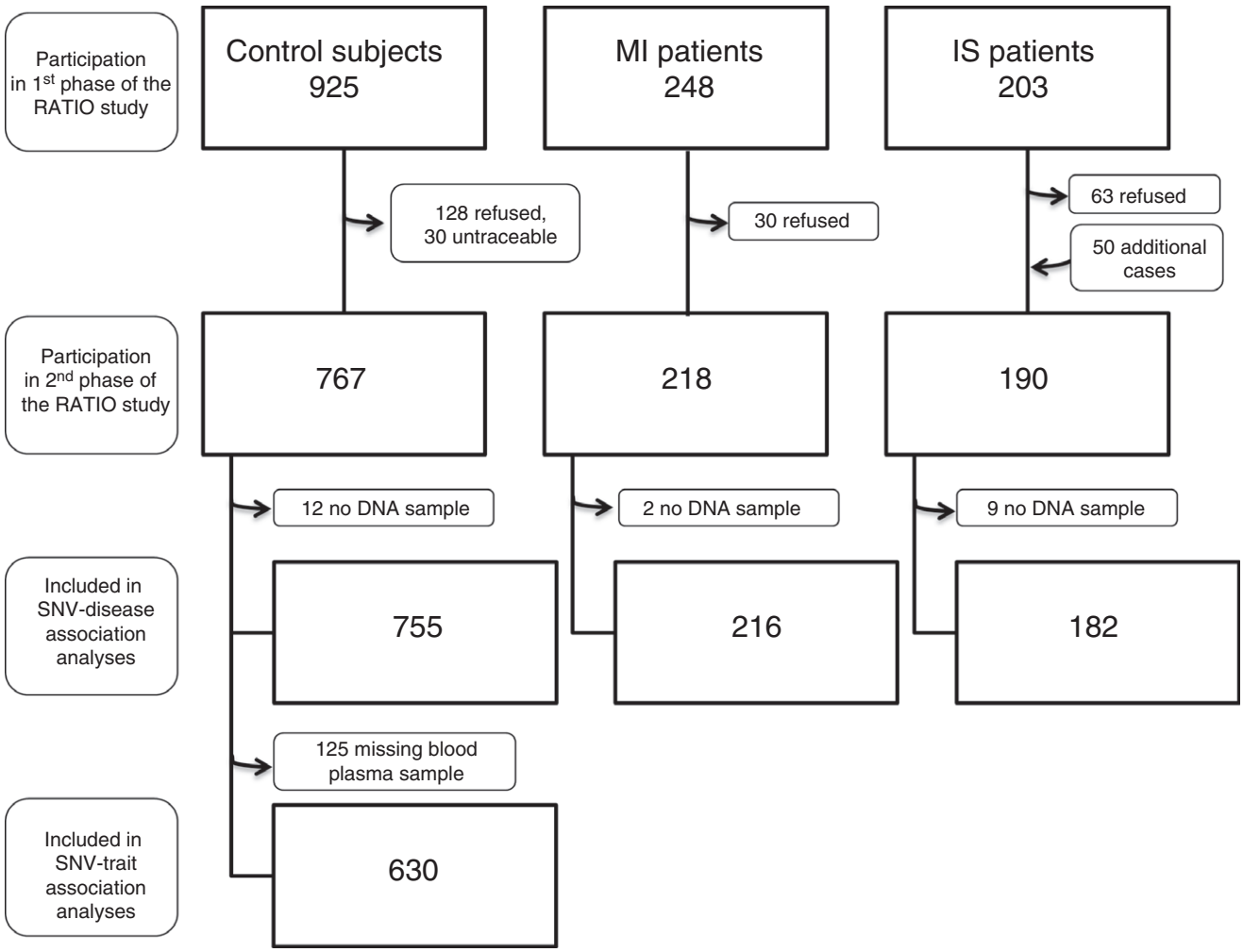

Fig. 1. Flow chart of study participant inclusion. MI, myocardial infarction; IS, ischemic stroke; SNV, single nucleotide variant.

95% CI, −20.07 to −4.69; $P_{conditional} = 0.0017$) and is a novel finding (Table 3).

Three *KLKB1* SNVs, rs1511801, rs2304595 and rs3087505, showed significant associations with both FXI:C and FXI:ag in single variant analyses (Table S4). After adjustment for PK:ag or lead F11 SNV rs2036914, the associations between *KLKB1* SNVs rs1511801 and rs2304595 and both FXI:C and FXI:ag were reduced (Table 4), indicating the observed effects could be explained by linkage between *KLKB1* and *F11* SNVs (e.g. pairwise LD of rs1511801 and rs2036914; $r^2 = 0.41$; $D' = 0.64$). The associations between rs3087505 and the two FXI traits were not affected by the adjustment for PK:ag; however, adjustment for *F11* SNV rs2036914 attenuated these associations considerably, whereas adjustment for rs1593, an *F11* SNV in moderate LD with rs3087505 ($r^2 = 0.71$, $D' = 1.00$), diminished them entirely.

KNG1 variation Eleven of the 20 *KNG1* SNVs had significant single variant associations with HMWK:ag (Table 2, non-significant results in Table S3). Lead SNV rs5030062 was strongly associated with HMWK:ag ($\beta = 10.23$; 95% CI, 7.87–12.59; $P = 1.63 \times 10^{-16}$; MAF = 38.3%) and explained 10.9% of the variance in HMWK:ag. This confirms a reported association of a tagged SNV rs3856930 with HMWK levels [27]. After adjustment for the lead SNV, eight associations remained (Table 3), the strongest of which was with rs2304456 ($\beta_{conditional} = -7.96$; 95% CI, −11.60 to −4.32; $P_{conditional} = 2.1 \times 10^{-5}$; MAF = 11.4%), recently described in the literature [27]. Additionally, a novel association between rs5029980 and HMWK:ag remained after adjustment for both rs5030062 and rs2304456 ($\beta_{indepedent} = 5.86$; 95% CI, 2.40–9.32; $P_{indepedent} = 0.001$). All other associations were attenuated.

Two *KNG1* SNVs (rs5030062 and rs5030060) showed across-trait associations with PK:ag, and four *KNG1* SNVs (rs5030062, rs5030060, rs1469859 and rs166479) showed associations with FXI:C (Table S4). The latter three SNVs are in moderate LD with rs5030062, with r^2 values of 0.65 ($D' = 1.00$), 0.53 ($D' = 0.86$) and 0.29 ($D' = 0.75$), and represent the same association signal. In

Table 1 Characteristics of RATIO control subjects with available DNA samples

Characteristics	Control subjects (n = 755)	
Age in years, median (range)	39.8 (34.9)	
Caucasian ethnicity, n (%)	712 (94)	
History of,* n (%)		
Hypertension	47 (6)	
Diabetes	10 (1)	
Hypercholesteremia	22 (3)	
Oral contraceptive use	267 (35)	
Smoking	316 (42)	
Plasma sample available, n (%)	630 (83)	
	Median (range)	**Mean (SD)**
Activity level†		
FXI:C	106 (144)	107.7 (22.9)
Antigen levels‡		
FXI:ag	112 (146)	115.4 (26.4)
PK:ag	131 (196)	130.0 (31.6)
HMWK:ag	115 (136)	117.4 (21.0)
FXII:ag	123 (279)	126.2 (40.1)

Range was calculated as maximum minus minimum value. SD indicates standard deviation. *In the year prior to index date of control subject, self-reported. PK, prekallikrein; HMWK, high-molecular-weight kininogen. †Measured as a percentage of activated normal pooled plasma. ‡Measured as a percentage of normal pooled plasma.

addition to FXI:C, rs5030062 was associated with FXI:ag after conditioning on FXI:C. This SNV as well as a tagged missense variant rs710446 have been associated with FXI levels in both pooled activity/antigen GWAS and activity-only analyses [25–27,38].

We investigated whether the associations between *KNG1* SNVs and the other contact system traits are a result of pleiotropy, when a single gene exerts an effect on multiple traits, or can be explained by a potentially mediating effect via HMWK. These exploratory analyses revealed that the association between rs5030062 and FXI:C remained after adjustment for HMWK:ag and FXI:ag, unlike the association of rs5030062 with FXI:ag, which was greatly reduced following adjustment for either HMWK:ag or FXI:C (Table 4). The replicated association between rs5030062 and PK:ag [27] disappeared following adjustment for HMWK:ag and was greatly reduced upon adjustment for FXI levels. The association between rs5030062 and HMWK:ag was only slightly reduced following adjustment for FXI:C and FXI:ag.

F12 variation Both rs1801020 (lead SNV) and rs17876032 have been associated with FXII:ag in previous studies [24,39], and rs1801020 with APTT [40]. Each showed strong associations with FXII:ag: $\beta = -41.54$ (95% CI, −45.67 to −37.41; $P = 6.85 \times 10^{-67}$; MAF = 25.3%) and $\beta = -28.09$ (95% CI, −32.52 to −23.66; $P = 1.42 \times 10^{-31}$; MAF = 34.9%) (Table 2). SNV rs1801020 explained 39.5% of the variance in FXII:ag, the largest of the five traits analyzed. No conditional effect of rs17876032 in non-carriers of rs1801020 was observed following adjustment (Table 3). Neither *F12* variant showed associations with the other traits (Table S4).

To address the potential issue of confounding as a result of population admixture, we additionally adjusted the analyses of the lead variants for ethnicity in a sensitivity analysis and observed no meaningful differences in the results (S5).

Haplotype analyses

Using the haplo.em function, haplotypes with frequencies of at least 1% in control subjects were constructed from various combinations of the lead and conditionally independent SNVs for each gene.

F11 The lead SNVs from both FXI traits (rs4253399 and rs2036914) and rs1593, a third SNV associated with both traits after conditioning for the other two, were included in the constructed *F11* haplotypes (Table 5). Haplotype 4, containing at least one copy of the rs4253399 major allele (T), the rs2036914 minor allele (T) and the rs1593 minor allele (T), was associated with a substantial decrease in FXI:ag ($\beta = -12.9$; SE = 2.8; $P < 0.0001$) compared with the reference haplotype containing all major alleles. The haplotypes were then tested for association with FXI:C; a moderate association was observed for Haplotype 4 ($\beta = -7.7$; SE = 2.5; $P = 0.002$; Table S6).

KLKB1 Our results revealed that the minor alleles of the two *KLKB1* SNVs with conditional effects did not occur in an inherited combination in our study population (Table S6), which is likely to be attributable to low linkage ($r^2 = 0.046$; $D' = 1.00$).

KNG1 Five haplotypes were constructed from the three conditionally independent *KNG1* SNVs. Compared with the reference haplotype, Haplotype 5, having at least one copy of minor alleles of rs5029980 (C) and rs5030062 (C), which occurred in 4.4% of control subjects, was associated with a 25.2 unit increase in HMWK:ag in control subjects (SE = 3.4; $P < 0.0001$).

F12 No haplotypes were generated for *F12* because no secondary associations remained after adjustment for the lead *F12* SNV (rs1801020).

Associations with IS and MI

None of the 26 SNVs, each significantly associated with levels of the respective protein for which the gene

Table 2 Statistically significant single variant associations in control subjects

Trait: FXI activity*						
F11 SNV	Chr. position	A1/A2	MAF (%)	β	95% CI	*P*
rs4253399	4:186266940	T/G	38.9%	9.24	6.69 to 11.79	3.32×10^{-12}
rs2036914	4:186271327	C/T	46.3%	−8.45	−10.93 to −5.98	4.64×10^{-11}
rs1593	4:186274397	A/T	12.2%	−8.51	−12.25 to −4.76	1.00×10^{-5}
rs4253417	4:186277851	T/C	41.7%	5.42	2.79 to 8.06	6.00×10^{-5}
rs4253418	4:186278343	G/A	4.9%	−9.99	−15.75 to −4.23	7.06×10^{-4}
Trait: FXI antigen†						
F11 SNV	Chr. position	A1/A2	MAF (%)	β	95% CI	*P*
rs2036914	4:186271327	C/T	46.3%	−11.38	−14.18 to −8.58	7.87×10^{-15}
rs4253399	4:186266940	T/G	38.9%	11.32	8.42 to 14.23	8.39×10^{-14}
rs1593	4:186274397	A/T	12.2%	−11.88	−16.13 to −7.63	6.07×10^{-8}
rs4253417	4:186277851	T/C	41.7%	6.25	3.25 to 9.25	4.90×10^{-5}
rs4253430	4:186288910	G/C	35.7%	−6.17	−9.17 to −3.17	6.00×10^{-5}
Trait: PK antigen†						
KLKB1 SNV	Chr. position	A1/A2	MAF (%)	β	95% CI	*P*
rs2304595	4:186251126	G/A	43.2%	10.33	6.65 to 14.02	5.58×10^{-8}
rs1511801	4:186229556	T/A	45.5%	−8.41	−12.00 to −4.82	5.00×10^{-6}
rs4253243	4:186232357	T/C	6.9%	−16.93	−24.19 to −9.67	6.00×10^{-6}
rs4253327	4:186257459	T/A	28.9%	−9.29	−13.43 to −5.14	1.30×10^{-5}
rs4253326	4:186257445	T/C	18.4%	−9.14	−13.86 to −4.42	1.60×10^{-4}
rs925453	4:186258056	C/T	31.0%	−6.40	−10.19 to −2.61	9.73×10^{-4}
Trait: HMWK antigen†						
KNG1 SNV	Chr. position	A1/A2	MAF (%)	β	95% CI	*P*
rs5030062	3:186736391	A/C	38.3%	10.23	7.87 to 12.59	1.63×10^{-16}
rs5030039	3:186730370	T/C	25.6%	−8.30	−10.88 to −5.71	5.57×10^{-10}
rs166479	3:186725461	C/T	45.3%	−7.74	−10.16 to −5.32	6.61×10^{-10}
rs5030060	3:186735128	C/T	30.0%	8.16	5.51 to 10.82	2.95×10^{-9}
rs1621816	3:186721384	T/C	28.9%	−7.87	−10.50 to −5.24	7.17×10^{-9}
rs2304456	3:186727263	T/G	11.4%	−10.81	−14.51 to −7.12	1.43×10^{-8}
rs5029999	3:186722770	C/T	20.5%	−8.03	−10.97 to −5.10	1.12×10^{-7}
rs1469859	3:186722454	G/A	32.1%	6.23	3.62 to 8.84	3.00×10^{-6}
rs266723	3:186729258	A/C	48.2%	−5.15	−7.61 to −2.69	4.60×10^{-5}
rs5029980	3:186720155	T/C	12.9%	6.58	2.95 to 10.22	4.10×10^{-4}
rs1648722	3:186731200	C/T	39.3%	−4.18	−6.66 to −1.69	0.00102
Trait: FXII antigen†						
F12 SNV	Chr. position	A1/A2	MAF (%)	β	95% CI	*P*
rs1801020	5:177409531	C/T	25.3%	−41.54	−45.67 to −37.41	6.85×10^{-67}
rs17876032	5:177403626	A/G	34.9%	−28.09	−32.52 to −23.66	1.42×10^{-31}

β coefficients representing the change in contact system trait level per copy of the minor allele were calculated from linear regression using an additive model adjusted for age. A significance level of 0.00116 was used based on a Bonferroni correction for 43 tests. Only significant associations between variants and the levels for which the gene is coding are shown here. Variants are ordered by level of statistical significance. A1 indicates major allele; A2, minor allele; Chr. position, chromosomal position (build GRCh37); CI, confidence interval; PK, prekallikrein; HMWK, high-molecular-weight kininogen; MAF, minor allele frequency in control subjects; SNV, single nucleotide variant. *Measured as a percentage of activated normal pooled plasma. †Measured as a percentage of normal pooled plasma.

encodes, showed single variant associations with either IS or MI (Table S7). The previously described haplotypes were also tested for associations with IS and MI in cases and controls adjusted for matching variables (Table S8). A weak relationship was observed between carriers of the *F11* Haplotype 4 and IS (OR_{IS} = 1.31; 95% CI, 0.80–2.16) or MI (OR_{MI} = 1.24; 95% CI, 0.77–2.01). *KNG1* Haplotype 5 did not show an association with IS or MI

Table 3 Conditional analyses of significant variants adjusted for lead SNV

FXI activity,* adjusted for rs4253399

F11 SNV	Chr. position	Dist. (kb)	LD (r^2)	β	95% CI	*P*
rs1593	4:186274397	7.46	0.10	−5.65	−9.46 to −1.83	0.004
rs4253418	4:186278343	11.40	0.06	−8.02	−13.62 to −2.41	0.005
rs2036914	4:186271327	4.39	0.60	−3.56	−7.11 to −0.02	0.049
rs4253417	4:186277851	10.91	0.74	0.59	−2.43 to 3.60	0.70

FXI antigen,† adjusted for rs2036914

F11 SNV	Chr. position	Dist. (kb)	LD (r^2)	β	95% CI	*P*
rs1593	4:186274397	3.07	0.17	−7.21	−11.70 to −2.73	0.002
rs4253399	4:186266940	4.39	0.60	6.20	2.10 to 10.34	0.003
rs4253417	4:186277851	6.52	0.46	1.77	−1.39 to 4.93	0.27
rs4253430	4:186288910	17.58	0.40	1.12	−2.46 to 4.70	0.54

PK antigen,† adjusted for rs2304595

KLKB1 SNV	Chr. position	Dist. (kb)	LD (r^2)	β	95% CI	*P*
rs4253243	**4:186232357**	**18.77**	**0.06**	**−12.38**	**−20.07 to −4.69**	**0.002**
rs4253326	4:186257445	6.32	0.15	−4.72	−9.97 to 0.56	0.08
rs4253327	4:186257459	6.33	0.31	−4.34	−9.38 to 0.71	0.09
rs1511801	4:186229556	21.57	0.64	−0.81	−6.78 to 5.16	0.79
rs925453	4:186258056	6.93	0.27	−0.58	−5.34 to 4.19	0.81

HMWK antigen,† adjusted for rs5030062

KNG1 SNV	Chr. position	Dist. (kb)	LD (r^2)	β	95% CI	*P*
rs2304456	3:186727263	9.13	0.06	−7.96	−11.60 to −4.32	2.1×10^{-5}
rs5029980	**3:186720155**	**16.24**	**0.00**	**7.02**	**3.59 to 10.44**	$\mathbf{6.4 \times 10^{-5}}$
rs5030039	3:186730370	6.02	0.19	−5.63	−8.32 to −2.95	4.4×10^{-4}
rs1621816	3:186721384	15.01	0.11	−4.36	−7.09 to −1.63	0.002
rs5029999	3:186722770	13.62	0.07	−4.67	−7.64 to −1.69	0.002
rs1469859	3:186722454	13.94	0.50	−4.31	−7.96 to −0.66	0.02
rs166479	3:186725461	10.93	0.29	−3.04	−5.86 to −0.21	0.04
rs5030060	3:186735128	1.26	0.69	−3.13	−7.72 to 1.47	0.18
rs1648722	3:186731200	5.19	0.19	−0.74	−3.28 to 1.80	0.57
rs266723	3:186729258	7.13	0.27	−0.70	−3.39 to 1.98	0.61

FXII antigen,† adjusted for rs1801020

F12 SNV	Chr. position	Dist. (kb)	LD (r^2)	β	95% CI	*P*
rs17876032	5:177403626	5.91	0.57	−0.84	−6.61 to 4.93	0.78

β coefficients representing the change in contact system trait level per copy of the minor allele were calculated from linear regression using an additive model adjusted for age and lead SNV for each trait. Variants are ordered by significance. Novel associations are shown in bold. Chr. position, chromosomal position (build GRCh37); CI, confidence interval; PK, prekallikrein; HMWK, high-molecular-weight kininogen; SNV, single nucleotide variant. Dist. indicates the distance in kilobase pairs between the lead SNV and the tested SNV; LD indicates linkage disequilibrium with corresponding lead SNV based on 1000 Genomes Project. *Measured as percentage of activated normal pooled plasma. †Measured as percentage of normal pooled plasma.

(OR_{IS} = 1.08, 95% CI, 0.53–2.23; OR_{MI} = 1.00, 95% CI, 0.46–2.21).

Discussion

In this study, we assessed relationships between tagging SNVs representing common genetic variation in the four contact system genes and protein levels of each factor in a single homogenous population. We identified *KLKB1* SNV rs4253243 and *KNG1* SNV rs5029980 as novel loci for PK:ag and HMWK:ag, respectively, and confirmed across-trait associations between *KNG1* variants and FXI levels, which can only in part be explained by HMWK:ag. We determined that many previously described across-trait associations can largely be explained by coinheritance of variants and potentially mediating effects via trait levels.

Table 4 Results from exploratory analyses

F11	Single variant analyses			Adjusted for FXI antigen*			Adjusted for FXI activity†		
Trait	β	95% CI	*P*	β	95% CI	*P*	β	95% CI	*P*
Variant rs2036914									
FXI:C	−8.45	−10.93 to −5.98	4.6×10^{-11}	−2.23	−4.31 to −0.15	0.04	–	–	–
FXI:ag	−11.38	−14.18 to −8.58	7.9×10^{-15}	–	–	–	−5.52	−7.80 to −3.24	3.0×10^{-6}
Variant rs4253399									
FXI:C	9.24	6.69 to 11.79	3.3×10^{-12}	3.07	0.95 to 5.20	0.005	–	–	–
FXI:ag	11.32	8.42 to 14.23	8.4×10^{-14}	–	–	–	4.80	2.42 to 7.17	8.2×10^{-5}
Variant rs1593									
FXI:C	−8.51	−12.25 to −4.76	1.0×10^{-5}	−1.89	−4.90 to 1.12	0.22	–	–	–
FXI:ag	−11.88	−16.13 to −7.63	6.1×10^{-8}	–	–	–	−5.74	−9.09 to −2.40	7.9×10^{-4}

KLKB1	Single variant analyses			Adjusted for PK antigen*			Adjusted for *F11* SNV rs2036914			Adjusted for *F11* SNV rs1593		
Trait	β	95% CI	*P*	β	95% CI	*P*	β	95% CI	*P*	β	95% CI	*P*
Variant rs1511801												
FXI:C	−5.34	−7.87 to −2.80	4.0×10^{-5}	−3.46	−5.96 to −0.96	0.007	−0.40	−3.46 to 2.66	0.80	–	–	–
FXI:ag	−6.20	−9.10 to −3.30	3.0×10^{-5}	−4.28	−7.11 to −1.45	0.003	0.95	−2.50 to 4.39	0.59	–	–	–
PK:ag	−8.41	−12.0 to −4.82	5.0×10^{-6}	–	–	–	−8.02	−12.47 to −3.57	4.3×10^{-4}	–	–	–
Variant rs2304595												
FXI:C	6.20	3.60 to 8.79	3.0×10^{-6}	4.17	1.56 to 6.78	0.002	2.13	−1.00 to 5.26	0.18	–	–	–
FXI:ag	6.69	3.73 to 9.65	1.1×10^{-5}	4.20	1.28 to 7.11	0.005	−0.01	−3.53 to 3.52	1.0	–	–	–
PK:ag	10.33	6.65 to 14.02	5.6×10^{-8}	–	–	–	10.53	5.97 to 15.09	7.0×10^{-6}	–	–	–
Variant rs3087505												
FXI:C	−9.12	−13.35 to −4.89	2.6×10^{-5}	−8.94	−13.03 to −4.86	2.0×10^{-5}	−5.81	−10.25 to −1.37	0.01	−2.58	−9.78 to 4.62	0.48
FXI:ag	−10.28	−15.13 to −5.44	3.5×10^{-5}	−9.70	−14.31 to −5.08	4.2×10^{-5}	−5.06	−10.10 to −0.02	0.05	1.69	−6.67 to 0.04	0.69
PK:ag	−2.67	−8.78 to 3.44	0.39	–	–	–	−0.71	−7.29 to 5.86	0.83	1.00	−9.65 to 11.66	0.85

KNG1	Single variant analyses			Adjusted for HMWK antigen*			Adjusted for FXI activity†			Adjusted for FXI antigen*		
Trait	β	95% CI	*P*	β	95% CI	*P*	β	95% CI	*P*	β	95% CI	*P*
Variant rs5030062												
FXI:C	9.88	7.29 to 12.46	2.3×10^{-13}	6.30	3.67 to 8.93	3.0×10^{-6}	–	–	–	6.25	4.22 to 8.28	2.8×10^{-9}
FXI:ag	6.80	3.75 to 9.84	1.4×10^{-5}	2.72	−0.36 to 5.80	0.08	−0.75	−3.19 to 1.70	0.55	–	–	–
PK:ag	7.08	3.29 to 10.86	2.6×10^{-4}	−0.70	−4.23 to 2.84	0.70	3.24	−0.59 to 7.07	0.10	4.60	0.92 to 8.28	0.01
HMWK:ag	10.23	7.87 to 12.59	1.6×10^{-16}	–	–	–	7.26	4.92 to 9.60	2.1×10^{-9}	8.60	6.32 to 10.89	5.4×10^{-13}

β coefficients were calculated from linear regression using an additive model adjusted for age and the additional adjustment factor indicated in each column. CI, confidence interval; PK, prekallikrein; SNV, single nucleotide variant; HMWK, high-molecular-weight kininogen. *Measured as a percentage of normal pooled plasma. †Measured as a percentage of activated normal pooled plasma.

Table 5 *F11* and *KNG1* haplotypes and associations with antigen levels of FXI and HMWK

FXI antigen*							
F11 SNVs: (A1/A2)	rs4253399 (T/G)	rs2036914 (C/T)	rs1593 (A/T)	Frequency, %	β	SE	*P*
Haplotype 1	T	C	A	15.2	Ref	–	–
Haplotype 2	T	T	A	33.7	−6.7	2.1	0.002
Haplotype 3	G	C	A	38.2	5.1	2.1	0.02
Haplotype 4	T	T	T	12.0	−12.9	2.8	< 0.0001
HMWK antigen*							
KNG1 SNVs: (A1/A2)	rs5029980 (T/C)	rs2304456 (T/G)	rs5030062 (A/C)	Frequency, %	β	SE	*P*
Haplotype 1	T	T	A	43.2	Ref	–	–
Haplotype 2	T	T	C	31.9	7.1	1.4	< 0.0001
Haplotype 3	T	G	A	10.6	−7.7	2.0	0.0001
Haplotype 4	C	T	A	9.0	−0.5	2.4	0.84
Haplotype 5	C	T	C	4.4	25.2	3.4	< 0.0001

Haplotype frequencies are based on results from regression analyses. The reference haplotype (1) contains only major alleles. Other haplotypes contain at least one copy of the specified minor allele(s). Effect size estimates (β) represent change in levels relative to the respective reference category. A1 indicates major allele; A2, minor allele; HMWK, high-molecular-weight kininogen; SE, standard error. Ref denotes the reference haplotype. *Percentage of normal pooled plasma.

Additionally, we probed the influence of these variants in inherited combinations as well as their potential roles in IS and MI.

Our study design provided the unique opportunity to assess whether variants were associated with FXI:C, FXI:ag or both, and whether these associations persisted after mutual adjustment. This distinction is rarely possible in smaller-scope studies or larger GWAS analyses, because activity and antigen levels are often studied in isolation or are pooled.

In the present study, three of 11 tested variants were found to have persistent associations with FXI:C or FXI:ag. SNV rs1593 tagged rs4253421, an intronic SNV classified as a genic enhancer [41], which was associated with pooled FXI levels in a recent GWAS [26]. The observed association between rs1593 and FXI:C disappeared after adjustment for antigen levels. This suggests the effect of this SNV (or a tagged SNV) influences FXI:C via its effect on FXI:ag (i.e. it has a quantitative rather than qualitative effect on the protein). Similar results were observed for *F11* SNV rs2036914, which is an expression quantitative trait locus (eQTL) for *F11* in lung tissue [42]. Results of our explanatory analyses clarified that the observed associations of rs2036914 and rs1593 with FXI activity can be completely explained by their associations with FXI:ag. We found that rs4253399 has a direct association with FXI:C not explained by FXI:ag.

In *KLKB1*, we found a strong association between lead SNV rs2304595 and PK:ag. SNV rs2304595 is an eQTL for *F11* and *CYP4V2* in several tissues but is not associated with *KLKB1* expression [42]. This association may be a result of the proximity of the *KLKB1* and *F11* genes, especially because an association with FXI levels was observed; however, to better understand this relationship, a formal mediation analysis is warranted, for which this study lacks the statistical power. Of the 10 variants tested, one additional novel association with rs4253243 remained associated with PK:ag after adjustment. SNV rs4253243 is an eQTL for *KLKB1* in lung, heart (atrial appendage) and nerve (tibia) tissue and as well for *F11* in brain tissue [42].

In *KNG1*, we confirmed two recently identified associations of rs5030062 and rs2304456 with HMWK:ag and found an additional novel association with rs5029980.

We did not observe much common variation in *F12* that influenced FXI:ag levels; only two SNVs, rs17876032 and rs1801020, captured the common variation in this gene. *F12* is the smallest of the four contact system genes analyzed, and this may explain the limited common variation. Interestingly, upon interrogation, only rs1801020 remained associated with FXII:ag. SNV rs1801020, a frequently occurring C->T polymorphism in the *F12* promoter region, was the top SNV in a previous FXII study and has been reported as an eQTL for *F12* in liver tissue [24,42].

Multiple associations were observed between *F11* and *KLKB1* SNVs and FXI:C, FXI:ag and PK:ag phenotypes. Our exploratory analyses demonstrated these observations can largely be explained by the close proximity of these genes (about seven kilobase pairs; see regional association plots in Figures S1–S3). We also observed associations between *KNG1* SNV rs5030062 and FXI:C and FXI:ag, confirming previous findings that *KNG1* loci influence FXI levels in plasma [25–27]. Our study adds detail to these observations; namely, that the effect of rs5030062 on FXI:ag can be explained by FXI:C. These findings support a previously proposed mechanism of *KNG1*-driven genetic regulation of FXI levels, in which the C allele

of missense rs710446 (tagged by rs5030062) has been postulated to impact the functionality of HMWK by altering binding, which, in turn, could impact FXI:C, because FXI exists in complex with HMWK in circulation [26]. Both rs710446 and rs5030062 have histone enhancer markers in liver tissue [41]. A functional study would be required to explore the exact mechanism.

In the present study, *KNG1* SNV rs5030062 was also found to be associated with PK:ag, as was previously described [27]. However, in our analyses, this association disappeared following adjustment for HMWK:ag and was greatly reduced after adjustment for FXI levels, suggesting this association with PK:ag may be mediated by the association between rs5030062 and HMWK:ag and/or FXI:C.

We constructed haplotypes for *F11* and *KNG1* to assess joint effects between SNVs in moderate LD. The *F11* haplotype occurring in 12.0% of control subjects, containing a T allele of rs2036914, T allele of rs2036914 and T allele of rs1593, had the largest joint effect on FXI:ag ($\beta = -12.9$; SE = 2.5; $P = 0.002$) and showed a slight association with IS (OR = 1.31; 95% CI, 0.80–2.16) and MI (OR = 1.24; 95% CI, 0.77–2.01). The *KNG1* haplotype containing the minor allele (C) of rs5029980, major allele (T) of rs2304456 and minor allele (C) of rs5030062 (Haplotype 5) had a substantially higher impact on HMWK:ag levels ($\beta = 25.2$, SE = 3.4, $P < 0.0001$) compared with the reference than would be expected based on the sum of the effects of the two haplotypes containing only one minor allele each (Haplotype 2: T-T-C, $\beta = 7.1$; Haplotype 4: C-T-A, $\beta = -0.5$). Variant rs5029980 is a synonymous SNV, and although it does not change the amino acid sequence, it may still affect protein expression, conformation and even function [43]. SNV rs5029980 has both promoter histone markers and enhancer markers in liver tissue [41]. Both rs2304456 and rs710446, a SNV tagged by rs5030062, are known missense variants and may alter protein function. This *KNG1* haplotype did not show an association with IS or MI ($OR_{IS} = 1.08$, 95% CI 0.53–2.23; $OR_{MI} = 1.00$, 95% CI 0.46–2.21), which may be because of low power, as this haplotype was carried by only 4.4% of control subjects.

Limitations

Even though our study was well powered to assess associations between single SNVs and multiple contact system traits, this study had limited power to test for associations of single SNVs with IS and MI risk, which is why we focused on haplotype combinations. We observed weak associations between *F11* and *KNG1* haplotypes containing relatively common variants and IS and MI.

Genotyping assays failed twice for *KNG1* SNV rs530072 and it was not included in our analyses. The failed SNV is in moderate LD with one of our lead SNVs (rs5030062, $r^2 = 0.79$ in 1000 Genomes data), and was dropped in the conditional analysis of a previous fine-mapping study on FXI activity [25], so it is unlikely we would have observed a strong effect on FXI levels by this SNV.

Remaining potential confounding factors in this study include population stratification and variants in linkage disequilibrium; as most women included in RATIO study were of European descent, population substructure was unlikely to be present, but the results can only be generalized to this population. Allelic frequencies and genomic distribution of polymorphisms were in Hardy–Weinberg equilibrium for all except five variants. This could be a result of chance, genotyping errors, or population admixture, although the latter is unlikely because four of five variants still tested below the 0.05 threshold after excluding non-Caucasians in a sensitivity analysis.

Since we chose SNVs that tag common variation in each gene, it is important to appreciate they may not be causal variants, rather just *linked* to causal variants. As in all case–control studies, survival bias could be a potential issue in the disease association analyses, and only control subjects were used for all other analyses. Furthermore, it has been demonstrated that adjusting genetic associations for correlated traits can bias effect estimates [44]; however, because we did not observe any directional changes in the effect estimates in the conditional analyses, this is unlikely to have impacted our overall results.

Conclusion

Our study adds to current knowledge of how genetic variation, also in haplotype combinations, influences levels of contact system proteins and clarifies interdependencies via exploratory analyses. Our findings explain the differential influence of genetic variation on FXI activity and antigen levels and provide additional support for proposed hypotheses regarding mechanisms of action for downstream regulation of FXI levels via *KNG1* loci. The newly observed associations between *KLKB1* SNV rs4253243 and PK:ag, as well as *KNG1* SNV rs5029980 and HMWK:ag, still require replication in other studies, and detailed research is needed to better understand which variants are causal and whether the haplotype associations with IS and MI can be confirmed in larger studies.

Addendum

The original design of the RATIO study was conceptualized by A. Algra and F. R. Rosendaal. This particular genetic sub-study was planned by C. Y. Vossen and B. Siegerink. H. G. de Haan and J. L. Rohmann performed the genetic and statistical analyses and generated all tables and figures. J. L. Rohmann, H. G. de Haan and B. Siegerink interpreted and contextualized results. J. L. Rohmann drafted the manuscript and it was critically revised by H. G. de Haan, B. Siegerink, A. Algra, C. Y.

Vossen and F. R. Rosendaal. All authors take responsibility for the manuscript content.

Acknowledgements

We thank P. Noordijk, L. Mahic and A. Hoenderdos for performing laboratory assays as well as I. de Jonge for data management. This work and the original RATIO study were supported by grants from the Dutch Heart Foundation (Grant 2001.069), the Leducq Foundation for the development of Transatlantic Networks of Excellence in Cardiovascular Research, Paris, France (Grant 04 CVD 02), and the Virchow Scholarship from the Center for Stroke Research Berlin (H. G. de Haan). The sponsors of this study are public or non-profit organizations that support science in general. They had no role in gathering, analyzing or interpreting the data. J. L. Rohmann is a PhD candidate at the Charité – Universitätsmedizin Berlin, and this work is submitted in partial fulfillment of her PhD requirements.

Disclosure of Conflict of Interests

F. R. Rosendaal is listed on patents of several prothrombotic variants, including factor XI. J. L. Rohmann, H. G. de Haan, A. Algra, C. Y. Vossen and B. Siegerink state that they have no conflict of interest.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Table S1A. Tagger results.
Table S1B. Minor allele frequencies, call rates and test of Hardy–Weinberg equilibrium.
Table S2. Characteristics of RATIO myocardial infarction and ischemic stroke patients with DNA samples.
Table S3. Non-significant *within-trait* single variant associations in control subjects.
Table S4. *Across-trait* single variant associations in control subjects with FXII, PK and HMWK antigen levels
Table S5. Sensitivity analysis: lead *within-trait* single variant associations in control subjects additionally adjusted for ethnicity.
Table S6. *F11* haplotype associations with FXI activity levels.
Table S7. Single variant associations with risk of myocardial infarction and ischemic stroke by gene.
Table S8. *F11* and *KNG1* haplotypes: associations with ischemic stroke and myocardial infarction.
Fig. S1. Regional association plot: *F11* and *KLKB1* SNVs with FXI activity.
Fig. S2. Regional association plot: *F11* and *KLKB1* SNVs with FXI antigen.
Fig. S3. Regional association plot: *F11* and *KLKB1* SNVs with PK antigen.

References

1 Gailani D, Renne T. Intrinsic pathway of coagulation and arterial thrombosis. *Arterioscler Thromb Vasc Biol* 2007; **27**: 2507–13.
2 Borissoff JI, Heeneman S, Kilinc E, Kassak P, Van Oerle R, Winckers K, Govers-Riemslag JW, Hamulyak K, Hackeng TM, Daemen MJ, ten Cate H, Spronk HM. Early atherosclerosis exhibits an enhanced procoagulant state. *Circulation* 2010; **122**: 821–30.
3 Zhang H, Lowenberg EC, Crosby JR, MacLeod AR, Zhao C, Gao D, Black C, Revenko AS, Meijers JC, Stroes ES, Levi M, Monia BP. Inhibition of the intrinsic coagulation pathway factor XI by antisense oligonucleotides: a novel antithrombotic strategy with lowered bleeding risk. *Blood* 2010; **116**: 4684–92.
4 Buller HR, Bethune C, Bhanot S, Gailani D, Monia BP, Raskob GE, Segers A, Verhamme P, Weitz JI. Factor XI antisense oligonucleotide for prevention of venous thrombosis. *N Engl J Med* 2015; **372**: 232–40.
5 Silverberg M, Diehl SV. The activation of the contact system of human plasma by polysaccharide sulfates. *Ann N Y Acad Sci* 1987; **516**: 268–79.
6 Long AT, Kenne E, Jung R, Fuchs TA, Renne T. Contact system revisited: an interface between inflammation, coagulation, and innate immunity. *J Thromb Haemost* 2016; **14**: 427–37.
7 Schmaier AH, McCrae KR. The plasma kallikrein–kinin system: its evolution from contact activation. *J Thromb Haemost* 2007; **5**: 2323–9.
8 Henderson LM, Figueroa CD, Muller-Esterl W, Bhoola KD. Assembly of contact-phase factors on the surface of the human neutrophil membrane. *Blood* 1994; **84**: 474–82.
9 Meijers JC, Tekelenburg WL, Bouma BN, Bertina RM, Rosendaal FR. High levels of coagulation factor XI as a risk factor for venous thrombosis. *N Engl J Med* 2000; **342**: 696–701.
10 Cushman M, O'Meara ES, Folsom AR, Heckbert SR. Coagulation factors IX through XIII and the risk of future venous thrombosis: the Longitudinal Investigation of Thromboembolism Etiology. *Blood* 2009; **114**: 2878–83.
11 Salomon O, Steinberg DM, Zucker M, Varon D, Zivelin A, Seligsohn U. Patients with severe factor XI deficiency have a reduced incidence of deep-vein thrombosis. *Thromb Haemost* 2011; **105**: 269–73.
12 Folsom AR, Tang W, Roetker NS, Heckbert SR, Cushman M, Pankow JS. Prospective study of circulating factor XI and incident venous thromboembolism: the Longitudinal Investigation of Thromboembolism Etiology (LITE). *Am J Hematol* 2015; **90**: 1047–51.
13 Doggen CJ, Rosendaal FR, Meijers JC. Levels of intrinsic coagulation factors and the risk of myocardial infarction among men: opposite and synergistic effects of factors XI and XII. *Blood* 2006; **108**: 4045–51.
14 Suri MF, Yamagishi K, Aleksic N, Hannan PJ, Folsom AR. Novel hemostatic factor levels and risk of ischemic stroke: the Atherosclerosis Risk in Communities (ARIC) Study. *Cerebrovasc Dis* 2010; **29**: 497–502.
15 Siegerink B, Rosendaal FR, Algra A. Antigen levels of coagulation factor XII, coagulation factor XI and prekallikrein, and the risk of myocardial infarction and ischemic stroke in young women. *J Thromb Haemost* 2014; **12**: 606–13.
16 Maino A, Rosendaal FR, Algra A, Peyvandi F, Siegerink B. Hypercoagulability is a stronger risk factor for ischaemic stroke

than for myocardial infarction: a systematic review. *PLoS ONE* 2015; **10**: e0133523.
17 Hanson E, Nilsson S, Jood K, Norrving B, Engstrom G, Blomstrand C, Lindgren A, Melander O, Jern C. Genetic variants of coagulation factor XI show association with ischemic stroke up to 70 years of age. *PLoS ONE* 2013; **8**: e75286.
18 Li Y, Bezemer ID, Rowland CM, Tong CH, Arellano AR, Catanese JJ, Devlin JJ, Reitsma PH, Bare LA, Rosendaal FR. Genetic variants associated with deep vein thrombosis: the F11 locus. *J Thromb Haemost* 2009; **7**: 1802–8.
19 Tang W, Teichert M, Chasman DI, Heit JA, Morange PE, Li G, Pankratz N, Leebeek FW, Pare G, de Andrade M, Tzourio C, Psaty BM, Basu S, Ruiter R, Rose L, Armasu SM, Lumley T, Heckbert SR, Uitterlinden AG, Lathrop M, *et al*. A genome-wide association study for venous thromboembolism: the extended cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium. *Genet Epidemiol* 2013; **37**: 512–21.
20 Bezemer ID, Bare LA, Doggen CJ, Arellano AR, Tong C, Rowland CM, Catanese J, Young BA, Reitsma PH, Devlin JJ, Rosendaal FR. Gene variants associated with deep vein thrombosis. *JAMA* 2008; **299**: 1306–14.
21 Morange PE, Oudot-Mellakh T, Cohen W, Germain M, Saut N, Antoni G, Alessi MC, Bertrand M, Dupuy AM, Letenneur L, Lathrop M, Lopez LM, Lambert JC, Emmerich J, Amouyel P, Tregouet DA. KNG1 Ile581Thr and susceptibility to venous thrombosis. *Blood* 2011; **117**: 3692–4.
22 Reiner AP, Lange LA, Smith NL, Zakai NA, Cushman M, Folsom AR. Common hemostasis and inflammation gene variants and venous thrombosis in older adults from the Cardiovascular Health Study. *J Thromb Haemost* 2009; **7**: 1499–505.
23 Souto JC, Almasy L, Borrell M, Garí M, Martínez E, Mateo J, Stone WH, Blangero J, Fontcuberta J. Genetic determinants of hemostasis phenotypes in Spanish families. *Circulation* 2000; **101**: 1546–51.
24 Calafell F, Almasy L, Sabater-Lleal M, Buil A, Mordillo C, Ramirez-Soriano A, Sikora M, Souto JC, Blangero J, Fontcuberta J, Soria JM. Sequence variation and genetic evolution at the human F12 locus: mapping quantitative trait nucleotides that influence FXII plasma levels. *Hum Mol Genet* 2010; **19**: 517–25.
25 Sabater-Lleal M, Martinez-Perez A, Buil A, Folkersen L, Souto JC, Bruzelius M, Borrell M, Odeberg J, Silveira A, Eriksson P, Almasy L, Hamsten A, Soria JM. A genome-wide association study identifies KNG1 as a genetic determinant of plasma factor XI level and activated partial thromboplastin time. *Arterioscler Thromb Vasc Biol* 2012; **32**: 2008–16.
26 Sennblad B, Basu S, Mazur J, Suchon P, Martinez-Perez A, van Hylckama Vlieg A, Truong V, Li Y, Gadin JR, Tang W, Grossman V, de Haan HG, Handin N, Silveira A, Souto JC, Franco-Cereceda A, Morange PE, Gagnon F, Soria JM, Eriksson P, *et al*. Genome-wide association study with additional genetic and post-transcriptional analyses reveals novel regulators of plasma factor XI levels. *Hum Mol Genet* 2017; **26**: 637–49.
27 Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, Sarwath H, Thareja G, Wahl A, DeLisle RK, Gold L, Pezer M, Lauc G, El-Din Selim MA, Mook-Kanamori DO, Al-Dous EK, Mohamoud YA, Malek J, Strauch K, Grallert H, *et al*. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun* 2017; **8**: 14357.
28 Solomon T, Smith EN, Matsui H, Braekkan SK, Wilsgaard T, Njølstad I, Mathiesen EB, Hansen J-B, Frazer KA; Consortium I. Associations between common and rare exonic genetic variants and serum levels of 20 cardiovascular-related proteins: the Tromsø Study. *Circ Cardiovas Genet* 2016;**9**:375–83.
29 Kemmeren JM, Tanis BC, van den Bosch MAAJ, Bollen ELEM, Helmerhorst FM, van der Graaf Y, Rosendaal FR, Algra A. Risk of Arterial Thrombosis in Relation to Oral Contraceptives (RATIO) Study: oral contraceptives and the risk of ischemic. *Stroke* 2002; **33**: 1202–8.
30 Pruissen DM, Slooter AJ, Rosendaal FR, van der Graaf Y, Algra A. Coagulation factor XIII gene variation, oral contraceptives, and risk of ischemic stroke. *Blood* 2008; **111**: 1282–6.
31 Tanis BC, van den Bosch MAAJ, Kemmeren JM, Cats VM, Helmerhorst FM, Algra A, van der Graaf Y, Rosendaal FR. Oral contraceptives and the risk of myocardial infarction. *N Engl J Med* 2001; **345**: 1787–93.
32 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–5.
33 Siegerink B, Rosendaal FR, Algra A. High-molecular-weight kininogen and the risk of a myocardial infarction and ischemic stroke in young women: the RATIO case-control study. *J Thromb Haemost* 2012; **10**: 2409–12.
34 Tanis B, Algra A, van der Graaf Y, Helmerhorst F, Rosendaal F. Procoagulant factors and the risk of myocardial infarction in young women. *Eur J Haematol* 2006; **77**: 67–73.
35 Sinnwell J, Schaid D. haplo.stats: Statistical Analysis of Haplotypes with Traits and Covariates when Linkage Phase is Ambiguous. R package version 1.7.7. https://CRAN.R-project.org/package = haplo.stats. 2016 [Available from: https://CRAN.R-project.org/package = haplo.stats.
36 Lunghi B, Cini M, Legnani C, Bernardi F, Marchetti G. The F11 rs2289252 polymorphism is associated with FXI activity levels and APTT ratio in women with thrombosis. *Thromb Res* 2012; **130**: 563–4.
37 Tang W, Schwienbacher C, Lopez LM, Ben-Shlomo Y, Oudot-Mellakh T, Johnson AD, Samani NJ, Basu S, Gogele M, Davies G, Lowe GD, Tregouet DA, Tan A, Pankow JS, Tenesa A, Levy D, Volpato CB, Rumley A, Gow AJ, Minelli C, *et al*. Genetic associations for activated partial thromboplastin time and prothrombin time, their gene expression profiles, and risk of coronary artery disease. *Am J Hum Genet* 2012; **91**: 152–62.
38 Martin-Fernandez L, Gavidia-Bovadilla G, Corrales I, Brunel H, Ramirez L, Lopez S, Souto JC, Vidal F, Soria JM. Next generation sequencing to dissect the genetic architecture of KNG1 and F11 loci using factor XI levels as an intermediate phenotype of thrombosis. *PLoS ONE* 2017; **12**: e0176301.
39 Sabater-Lleal M, Chillón M, Mordillo C, Martínez Á, Gil E, Mateo J, Blangero J, Almasy L, Fontcuberta J, Soria JM. Combined cis-regulator elements as important mechanism affecting FXII plasma levels. *Thromb Res* 2010; **125**: e55–60.
40 Houlihan LM, Davies G, Tenesa A, Harris SE, Luciano M, Gow AJ, McGhee KA, Liewald DC, Porteous DJ, Starr JM, Lowe GD, Visscher PM, Deary IJ. Common variants of large effect in F12, KNG1, and HRG are associated with activated partial thromboplastin time. *Am J Hum Genet* 2010; **86**: 626–31.
41 Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* 2016; **44**: D877–81.
42 Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015; **348**: 648–60.
43 Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet* 2011; **12**: 683–91.
44 Aschard H, Vilhjalmsson BJ, Joshi AD, Price AL, Kraft P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am J Hum Genet* 2015; **96**: 329–39.

13.2. Publication 2: *Coagulation factor XII, XI, and VIII activity levels and secondary events after first ischemic stroke*

Rohmann JL, Huo S, Sperber PS, Piper SK, Rosendaal FR, Heuschmann PU, Endres M, Liman TG, Siegerink B. Coagulation factor XII, XI, and VIII activity levels and secondary events after first ischemic stroke. J Thromb Haemost. 2020 Dec; 18(12):3316-3324. Available from: http://dx.doi.org/10.1111/jth.15092

The extract from ISI Web of Knowledge Journal Summary List and the original published version of the article follow on next 10 pages.

Journal Data Filtered By: **Selected JCR Year: 2018** Selected Editions: SCIE,SSCI
Selected Categories: **"HEMATOLOGY"** Selected Category Scheme: WoS
Gesamtanzahl: 73 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	BLOOD	161,827	16.562	0.240720
2	CIRCULATION RESEARCH	52,988	15.862	0.072290
3	Lancet Haematology	1,934	11.990	0.010520
4	LEUKEMIA	24,555	9.944	0.054750
5	Journal of Hematology & Oncology	5,366	8.731	0.013620
6	Blood Cancer Journal	2,247	7.895	0.009060
7	HAEMATOLOGICA	16,255	7.570	0.037660
8	ARTERIOSCLEROSIS THROMBOSIS AND VASCULAR BIOLOGY	33,223	6.618	0.036000
9	AMERICAN JOURNAL OF HEMATOLOGY	10,375	6.137	0.022930
10	BLOOD REVIEWS	2,889	6.125	0.005980
11	JOURNAL OF CEREBRAL BLOOD FLOW AND METABOLISM	19,766	6.040	0.028050
12	STEM CELLS	21,467	5.614	0.030220
13	BRITISH JOURNAL OF HAEMATOLOGY	23,963	5.206	0.037720
14	CRITICAL REVIEWS IN ONCOLOGY HEMATOLOGY	7,401	5.012	0.012890
15	THROMBOSIS AND HAEMOSTASIS	16,590	4.733	0.022810
16	BONE MARROW TRANSPLANTATION	12,031	4.674	0.020710
17	JOURNAL OF THROMBOSIS AND HAEMOSTASIS	18,886	4.662	0.028230
18	CYTOTHERAPY	5,969	4.297	0.009690
19	JOURNAL OF LEUKOCYTE BIOLOGY	16,921	4.012	0.019570
20	SEMINARS IN HEMATOLOGY	2,157	3.738	0.003950
21	TRANSFUSION MEDICINE REVIEWS	1,434	3.610	0.002890

Received: 1 March 2020 | Accepted: 24 August 2020

DOI: 10.1111/jth.15092

ORIGINAL ARTICLE

Coagulation factor XII, XI, and VIII activity levels and secondary events after first ischemic stroke

Jessica L. Rohmann[1,2] | Shufan Huo[1,3,4] | Pia S. Sperber[1,4] | Sophie K. Piper[5,6] | Frits R. Rosendaal[7] | Peter U. Heuschmann[8,9] | Matthias Endres[1,3,4,10,11] | Thomas G. Liman[1,3,4] | Bob Siegerink[1,7]

[1]Center for Stroke Research Berlin, Charité – Universitätsmedizin Berlin, Berlin, Germany

[2]Institute of Public Health, Charité – Universitätsmedizin Berlin, Berlin, Germany

[3]Klinik für Neurologie, Charité – Universitätsmedizin Berlin, Berlin, Germany

[4]Deutsches Zentrum für Herz-Kreislauf-Forschung DZHK, Berlin, Germany

[5]Insitute for Biometry and Clinical Epidemiology, Charité – Universitätsmedizin Berlin, Berlin, Germany

[6]Berlin Institute of Health (BIH), Berlin, Germany

[7]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands

[8]Institute of Clinical Epidemiology and Biometry, University of Würzburg, Würzburg, Germany

[9]Clinical Trial Center Würzburg, University Hospital Würzburg, Würzburg, Germany

[10]German Center for Neurodegenerative Disease DZNE, Charité – Universitätsmedizin Berlin, Berlin, Germany

[11]Excellence Cluster Neurocure, Charité – Universitätsmedizin Berlin, Berlin, Germany

Correspondence
Jessica L. Rohmann, Center for Stroke Research, Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10117, Berlin, Germany.
Email: jessica.rohmann@charite.de

Funding Information
The PROSCIS-B study received funding from the Federal Ministry of Education and Research via the grant Center for Stroke Research Berlin (01 EO 0801).

Abstract

Background: Though risk for recurrent vascular events is high following ischemic stroke, little knowledge about risk factors for secondary events post-stroke exists.

Objectives: Coagulation factors XII, XI, and VIII (FXII, FXI, and FVIII) have been implicated in first thrombotic events, and our aim was to estimate their effects on vascular outcomes within 3 years after first stroke.

Patients/Methods: In the Prospective Cohort with Incident Stroke Berlin (PROSCIS-B) study, we followed participants aged 18 and older for 3 years after first mild to moderate ischemic stroke event or until occurrence of recurrent stroke, myocardial infarction, or all-cause mortality. We compared high coagulation factor activity levels to normal and low levels and also analyzed activities as continuous variables. We used Cox proportional hazards models adjusted for age, sex, and cardiovascular risk factors to estimate hazard ratios (HRs) for the combined endpoint.

Results: In total, 94 events occurred in 576 included participants, resulting in an absolute rate of 6.6 events per 100 person-years. After confounding adjustment, high FVIII activity showed the strongest relationship with the combined endpoint (HR = 2.05, 95% confidence interval [CI] 1.28–3.29). High FXI activity was also associated with a higher hazard (HR = 1.80, 95% CI 1.09–2.98), though high FXII activity was not (HR = 0.86, 95% CI 0.49–1.51). Continuous analyses yielded similar results.

Conclusions: In our study of mild to moderate ischemic stroke patients, high activity levels of FXI and FVIII but not FXII were associated with worse vascular outcomes in the 3-year period after first ischemic stroke.

KEYWORDS

coagulation, factor VIII, factor XI, factor XII, ischemic stroke

Manuscript handled by: Matthew T. Rondina

Final decision: Matthew T. Rondina, 24 August 2020

1 | INTRODUCTION

Globally, stroke remains a leading cause of disability and mortality.[1,2] Following first-ever ischemic stroke, the risk of secondary events is high.[3-5] Although many risk factors for first ischemic stroke have been identified, comparatively little is known about factors that contribute to secondary post-stroke events. In fact, a recent systematic review of biomarkers of hemostasis found no conclusive evidence of a single marker ready for use in practice, largely due to the limited number of existing studies.[6]

The coagulation factors XII, XI, and VIII (FXII, FXI, and FVIII) are promising candidates for further investigation in this context. As the initiating factor in the contact activation system, FXII was quickly implicated in first-ever vascular events in animal studies; however, conflicting evidence exists regarding the potential role of FXII in ischemic stroke events in humans.[7-10] The role of FXI in hypercoagulability has been more consistently demonstrated, and high levels of FXI have been linked to thrombotic events, especially ischemic stroke.[11-13] Given its role in thrombin activation and thrombus formation, it is unsurprising that high levels of FVIII have also been implicated in vascular events.[14,15] FVIII elevation is observed during the acute phase of stroke as part of the inflammatory response;[16] however, a dose-dependent relationship between FVIII and thrombosis, independent of this acute phase response, has also been described.[14,15,17,18] A recent study found that acute ischemic stroke patients with elevated FVIII experienced a higher frequency of recurrent thrombotic events while in hospital.[19] It remains unknown whether this increased risk due to FVIII elevation also persists in the longer term for future incident thrombotic events.

In the present study, we aimed to estimate the effects of FXII, FXI, and FVIII activity levels on risk for secondary vascular events among ischemic stroke patients.

Essentials

- Factors XII, XI, and VIII are linked with first vascular events; role in secondary events unclear.
- We followed adult stroke patients for 3 years or until stroke, myocardial infarction, or death.
- We report confounding-adjusted estimates for effect of factor activities on the combined endpoint.
- High FXI and FVIII but not FXII activities were associated with worse post-stroke vascular outcomes.

2 | MATERIALS AND METHODS

2.1 | Study population

We used data from the Prospective Cohort with Incident Stroke Berlin (PROSCIS-B; clinicaltrials.gov registration number: NCT01363856). This longitudinal, hospital-based, observational cohort study has been described in detail elsewhere.[20] Participants (or legal representatives) provided written informed consent for study participation. The study protocol was approved by the internal review board of the Charité – Universitätsmedizin Berlin (EA1/218/09) and was conducted in accordance with ethical principles described in the Declaration of Helsinki.

In brief, between January 2010 and February 2013, patients aged 18 or older presenting at one of the three tertiary stroke units at the Charité – Universitätsmedizin in Berlin with first-ever stroke (defined by World Health Organization criteria[21]), including ischemic stroke, primary hemorrhage, or sinus venous thrombosis, were recruited. Participants underwent a baseline visit within 1 week of the initial event, including a detailed interview, clinical examination, and the collection of blood samples stored for later analysis. Participants were contacted annually over a period of 3 years via telephone interview to document vital status, any incident cardiovascular events, and to assess functional outcome. Those who were not reachable by phone were mailed surveys.

As shown in Figure 1, in this study, we excluded non-ischemic stroke patients and patients with severe strokes (defined as having a National Institute of Health Stroke Scale [NIHSS] assessment of >15). Overall, activity measurements for at least one coagulation factor were available for 576 PROSCIS-B participants, who were subsequently included in these analyses.

2.2 | Participant characteristics

At baseline, age, sex, and cardiovascular risk factors were assessed. In the baseline clinical assessment, body mass index (BMI, in kg/m^2), high density lipoprotein (HDL, in mg/dL), and low-density lipoprotein (LDL, in mg/dL) cholesterol were measured. Participants were asked to provide information about lifestyle-related risk factors including smoking (never, former, or current); whether they consumed alcohol regularly; and whether they had a history of diabetes mellitus, hypertension, and acute coronary syndrome (myocardial infarction or angina pectoris). The stroke units provided information on whether the patient received recombinant tissue-type plasminogen activator (rt-PA) treatment, the suspected stroke etiology according to the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) classification,[22] and the severity of the stroke based on the NIHSS (mild: 0–4, moderate: 5–15, severe: >15).[23]

2.3 | Exposure assessment: coagulation factors

Citrate-buffered blood samples were obtained from PROSCIS-B participants after an overnight fast within 1 week of the initial stroke event and aliquots were stored at −80°C degrees until thawed once for the laboratory assays. Between the initial stroke event and the time of blood sampling, a median of 4 days elapsed (interquartile range [IQR] limits: 3–5). Coagulation factor activity

levels (:C) were measured using a one-stage clotting assay and are reported as percentages of activated normal pooled plasma (standard activity units). Some of the samples had too little plasma; in these cases, FXI:C followed by FXII:C measurements were prioritized as less is known about these factors in the context of secondary vascular risk compared to FVIII:C. Coagulation factor measurements were performed blinded to participant characteristics and outcome status.

2.4 | Outcome: combined endpoint

We generated the combined endpoint outcome as the composite of relevant secondary event occurrence; first of either recurrent stroke, myocardial infarction, or death attributable to any cause during follow-up. During follow-up, participants were requested to provide information about the occurrence of any of these events since the last time of contact. We confirmed these self-reported outcomes using the Charité – Universitätsmedizin hospital discharge records or, when not available, using information obtained from the treating hospital or general practitioner.

We performed additional screening of the Charité hospital records to identify any events of interest not self-reported by participants during follow-up. Information about death from any cause was supplemented using city registration office's records. For one participant, the exact date of death could not be determined and was assigned as the halfway point between last contact and the date on the returned postal questionnaire.

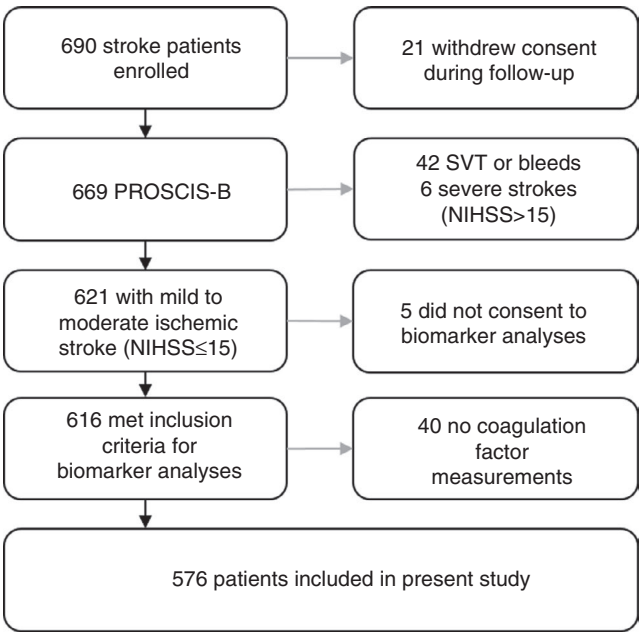

FIGURE 1 Participant inclusion flowchart. Abbreviations: NIHSS, National Institutes of Health Stroke Scale; SVT, sinus venous thrombosis. Of the 40 participants with no recorded coagulation factor activity measurements, 15 had samples that failed in the laboratory and the remainder had no stored citrate available for analysis

2.5 | Endpoint committee

All cardiovascular endpoints were confirmed using medical records from the treating hospital or physician and validated by an independent endpoint committee consisting of two senior vascular neurologists. We used only these committee-confirmed endpoints in our analyses.

2.6 | Statistical analysis

We summarized the baseline characteristics for the full PROSCIS-B cohort using medians and IQR limits for continuous variables and frequencies and percentages for nominal variables.

In the primary analysis, we categorized the FXII:C, FXI:C, and FVIII:C levels into quartile groups and compared participants with the highest fourth (>75th percentile) to the remainder (reference) for each factor. In an additional analysis, for each factor, we analyzed the activity measurements as continuous exposure variables divided by the standard deviation of all measurements of that factor to allow for better comparisons of the estimated effect sizes between the three factors. All reported analyses are complete case analyses.

In the time-to-event analyses, we calculated person-time from the date of the initial ischemic stroke to the date of occurrence of the combined endpoint during follow-up (first occurrence of either recurrent stroke, myocardial infarction, or death by any cause), loss to follow-up, or the study end, whichever came first. We used Kaplan-Meier curves to estimate event-free survivorship and the log-rank test to measure overall crude differences in survivorship curves between groups after visual inspection of fulfillment of the proportional hazards assumption. Dropouts were censored on the date of last contact.

We used Cox proportional hazards models to estimate the hazard ratios (HR) and 95% confidence intervals (CI) for the combined endpoint outcome adjusted for potential confounding factors. We used multiple models for confounding control: Model 1 was adjusted only for age and sex. In Model 2, we additionally adjusted for cardiovascular factors determined to contribute to confounding based on *a priori* knowledge. In addition to age and sex, Model 2 included the continuous variables BMI, HDL, and LDL cholesterol levels; the categorical variable smoking status (never, ever, current); and the following dichotomous variables: regular alcohol consumption, hypertension, diabetes mellitus, and acute coronary syndrome.

As a sensitivity analysis, in a third model (Model 3), we adjusted for all Model 2 covariates as well as rt-PA treatment status and stroke severity (NIHSS). Because these are consequences of the stroke, they could be intermediates in the causal path of interest and may not contribute to confounding directly. However, as they may also be proxies for relevant pre-stroke confounders, we decided to explore how the effect estimates change with their inclusion in Model 3.

We performed all analyses using STATA IC version 14.2 (Stata Corp.). The syntax is available on request from the corresponding author.

3 | RESULTS

Participant characteristics at baseline are displayed in Table 1. PROSCIS participants with mild to moderate ischemic stroke (N = 621) were predominantly male (61%) and had a median age of 69 years (IQR limits: 58–76). Arterial hypertension was observed in 65% and diabetes mellitus in 22% of participants. Twenty percent of participants received rt-PA treatment. Median activity levels for FXII:C, FXI:C, and FVIII:C were 108 (IQR limits: 91–127), 113 (99–130), and 140 (115–166), respectively. At least one of the three coagulation factor measurements was available for 576 participants (93%) who were included in the present analyses. All three factor activity measurements were available for 553 participants. The distribution of high versus low/normal activity levels across the TOAST stroke subtypes is displayed in Table S1 in the Supporting Information. All other variables relevant for this study measured at baseline had <5% missing values aside from the LDL and HDL cholesterol levels, for which 34 participants had missing values.

After a pursuit follow-up time of 3.0 years resulting in 1419.5 contributed person-years, 94 combined endpoint events occurred. Of these, 41 were recurrent ischemic strokes, 5 were myocardial infarctions, and 48 were deaths. The overall crude observed incidence rate for included participants was 6.6 events per 100 person-years. The absolute cumulative risk for the combined outcome during follow-up among the 576 included participants was 16.3%.

We generated Kaplan-Meier curves to compare participants with coagulation factor levels in the highest fourth (>p75) with the remainder for the three factors of interest (Figure 2A-C). In the crude comparison, no significant difference between >p75 and ≤p75 groups of FXII:C was observed in the log-rank test ($P = .48$). However, clear differences were visible for both the FXI:C and FVIII:C curve comparisons. Visually, participants with high levels had consistently higher cumulative probabilities of the combined endpoint compared to the reference group (FXI:C: $P = .06$; FVIII:C: $P = .0001$).

The multivariable adjusted HRs are shown in Table 2. In the fully adjusted model (Model 2), high FXII:C levels (>p75) were not associated with the combined endpoint (HR = 0.86, 95% CI 0.49–1.51). Having high FXI:C levels was associated with a higher hazard for the combined endpoint: (HR = 1.80, 95% CI 1.09–2.98), as was having high FVIII:C levels (HR = 2.05, 95% CI 1.28–3.29), compared to low/normal levels. In the secondary analyses treating the coagulation factor levels as continuous variables, we obtained similar results (Table 2). One standard deviation of FXII:C, FXI:C, and FVIII:C levels corresponded to 29.3, 28.8, and 45.4 units, respectively.

In a sensitivity analysis (Model 3), we further adjusted Model 2 for NIHSS and rt-PA treatment. This additional adjustment did not substantially change the results (Table 2).

TABLE 1 Baseline characteristics of PROSCIS-B participants with mild to moderate ischemic stroke

PROSCIS-B participants with mild to moderate ischemic stroke (N=621)[a]		
Age in years, median (IQRL)	69	(58–76)
Female sex, N (%)	242	(39%)
BMI in kg/m^2, median (IQRL)	27	(24–30)
HDL cholesterol in mg/dL, median (IQRL)	49	(40–60)
LDL cholesterol in mg/dL, median (IQRL)	117	(96–147)
Hypertension, N (%)	406	(65%)
Acute coronary syndrome, N (%)	99	(16%)
Diabetes mellitus, N (%)	137	(22%)
Habitual alcohol consumption, N (%)	217	(35%)
Smoking, N (%)		
Former	201	(33%)
Current	171	(28%)
TOAST subtype, N (%)		
Large-artery atherosclerosis	167	(27%)
Cardioembolism	145	(23%)
Small vessel occlusion	96	(15%)
Other determined etiology	22	(4%)
Undetermined etiology[b]	191	(31%)
NIHSS, N (%)		
Mild, 0–4	470	(76%)
Moderate, 5–15	151	(24%)
Thrombolysis treatment (rt-PA), N (%)	125	(20%)
Coagulation measurements available for at least one factor[c], N (%)	576	(93%)
FXII:C[d], median (IQRL)	108	(91–127)
FXI:C, median (IQRL)	133	(99–130)
FVIII:C, median (IQRL)	140	(115–166)

Abbreviations: BMI, body mass index; FVIII:C, coagulation factor VIII activity; FXI:C, coagulation factor XI activity; FXII:C, coagulation factor XII activity; HDL, high-density lipoprotein; IQRL, interquartile range limits; LDL, low-density lipoprotein; NIHSS, National Institutes of Health Stroke Scale; PROSCIS-B, Prospective Cohort with Incident Stroke Berlin; rt-PA, recombinant tissue-type plasminogen activator; TOAST, stroke etiology according to Trial of Org 10172 in Acute Stroke Treatment.

[a]Owing to missing data, percentages may not total 100%. 5 participants did not consent to biomarker measurements and were excluded. All variables have <5% missing values except for the coagulation factor and cholesterol measurements. In total, 34 study participants were missing LDL and HDL cholesterol measurements, and 40 participants were missing all three coagulation factor activity measurements.

[b]Includes cryptogenic stroke despite complete work-up, stroke of two or more etiologies, and stroke with incomplete work-up.

[c]Laboratory measurements available for at least one of the coagulation factors of interest; 553 had all three coagulation factor activity measurements.

[d]Activities were measured as percentages of activated normal pooled plasma.

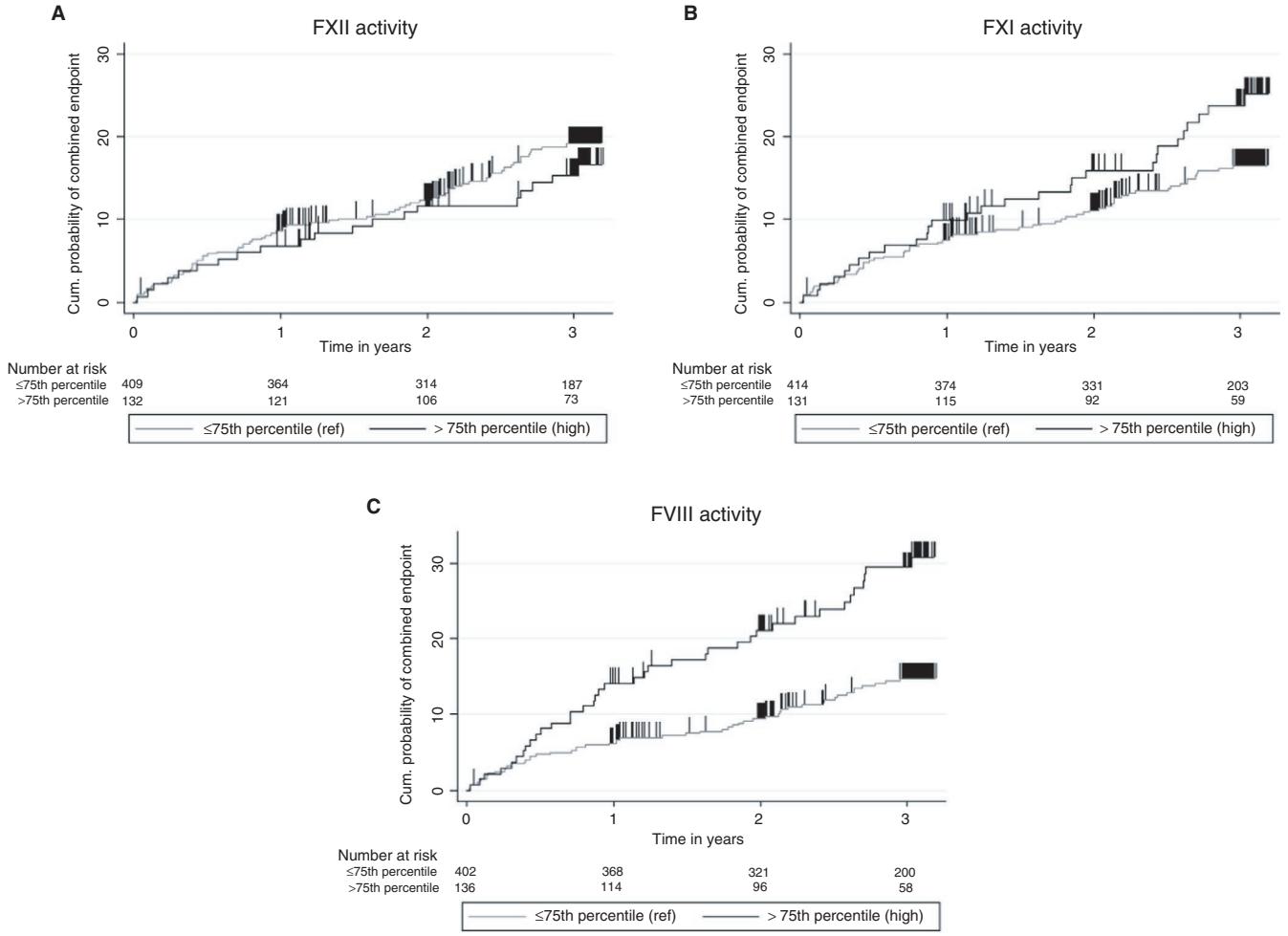

FIGURE 2 Cumulative probabilities of the combined vascular endpoint. Kaplan-Meier estimates among stroke patient participants with high activity levels (>75th percentile) of each indicated coagulation factor (Panel A) FXII, (Panel B) FXI, and (Panel C) FVIII, compared to those with lower activity levels of each factor (≤75th percentile, reference). Activity levels were measured as percentages of activated normal pooled plasma. Abbreviations: Cum., cumulative; FVIII, coagulation factor VIII; FXI, coagulation factor XI; FXII, coagulation factor XII; ref, reference.

4 | DISCUSSION

In this prospective patient cohort study of individuals with mild to moderate ischemic stroke, having high levels of FXI or FVIII activity was associated with a higher hazard for the combined vascular endpoint within 3 years compared with individuals having low/normal levels after confounding adjustment. Such a relationship was not observed for having high FXII activity. Our findings expand on the recent literature and fill some important gaps with longitudinal results.

Specifically in the context of post-stroke outcomes, the search for meaningful etiologic or prognostic hemostatic biomarkers has not been straightforward. A recent systematic review of hemostatic biomarkers in ischemic stroke revealed a large heterogeneity in existing studies and did not find enough evidence to provide clear recommendations for a prognostic marker to be used in practice; however, some biomarkers, including FXI and FVIII, seemed promising in some of the included studies.[6]

In general, the existing evidence on coagulation factor activity and post-stroke outcomes is quite limited. Our findings pertaining to FXI activity and long-term outcomes add to previous findings from a cross-sectional study of ischemic stroke and transient ischemic attack patients aged 70 or younger, which concluded that the presence of circulating activated FXI during the acute phase of cerebral ischemia (defined as having a detectable response to inhibitory monoclonal antibodies in plasma harvested within 72 hours of symptom onset) at hospital admission was associated with a higher NIHSS score, higher Modified Rankin Scale (mRS), and lower Barthel index at discharge.[24]

Another study found that acute ischemic stroke patients with elevated FVIII (>1.50 IU/mL) at hospital admission experienced a higher frequency of recurrent thrombotic events (defined as new ischemic stroke, progressive stroke, myocardial infarction, deep vein thrombosis, or pulmonary embolism) while in hospital.[19] Our findings add to these results by showing that this relationship also appears to persist in the longer term for future incident events. Regarding ischemic stroke patients who underwent thrombolysis, a study found that having elevated FVIII activity levels (defined as >168%, the upper limit of the reference level), both immediately and 24 hours

after thrombolysis, led to a higher risk for poor functional outcome (mRS ≥ 3) at 90 days.[25]

There have been some important recent developments regarding hemostatic factors in the context of first stroke that may also provide relevant insights in the context of secondary prevention. For instance, FXI appears to be a stronger risk factor for first ischemic stroke than myocardial infarction[26] and may therefore be a particularly attractive target. Numerous laboratory and animal studies have further demonstrated a prothrombotic role of FXII and FXI in thrombosis but non-integral role of these factors in normal hemostasis, making them promising targets for future anticoagulant drug development with potentially lower bleeding risk.[7,27,28] In addition to the successful trial conducted among knee arthroplasty patients, in which FXI antisense oligonucleotides prevented venous thrombotic events without increasing bleeding risk,[29] a recent genetic study investigating variants known to alter FXI levels and increase relative activated partial thromboplastin time found that genetic disposition to lower FXI levels was associated with lower odds for ischemic stroke without increasing risk for major bleeding.[30] This decrease was equivalent to the FXI level reduction that can be achieved through pharmacological modulation.[30]

Furthermore, a 2018 study that used two-sample Mendelian randomization found that genetically determined FXI levels had a causal effect on the risk of any ischemic stroke but not myocardial infarction or intracerebral hemorrhage, with the strongest effect observed amongst the cardioembolism subgroup.[31] Shortly after, a 2019 study that also used Mendelian randomization techniques in a larger meta-analysis, integrating phenomic, genomic, and proteomic databases, assessed the role of 653 proteins as potential mediators for ischemic stroke subtypes and relevant side effects.[32] In this study, genetically determined FXI levels were identified as one of five causal mediators of ischemic stroke, with the cardioembolic subtype appearing to drive this effect.[32] In both studies, no adverse side effects appeared to be linked to the genetic influences on variation in FXI levels, providing further justification for clinical trials on FXI-related interventions in the context of ischemic stroke.[32] Another study published in 2020 used two-sample Mendelian randomization to assess the causal relationships among numerous coagulation factor and other hematological traits on ischemic stroke and its subtypes using data from the MEGASTROKE Consortium.[33] Specifically, genetically higher levels of FXI activity and FVIII antigen were each found to be associated with increased ischemic stroke risk as well as specific risk for the cardioembolic subtype, but not with small-vessel stroke risk.[33] Interestingly, reduced FVIII activity was associated with ischemic stroke risk, and specifically the cardioembolic and large artery atherosclerosis subtypes (the latter only among the European population).[33] Our study was not powered to investigate stratum-specific effects across the stroke subtypes, but this area warrants future research also in the context of secondary event prevention.

In light of our findings that stroke patients with high FXI:C had a higher risk for secondary events after first stroke, the population of stroke patients with high FXI levels may particularly benefit from such targeted interventions. FXI:C may be a promising biomarker for identifying individuals who are most likely to benefit from such interventions.

TABLE 2 Hazard ratios from Cox proportional hazards regression models for FXII:C, FXI:C, and FVIII:C

FXII:C[a]	n	Combined EP events	HR1[b]	95% CI	HR2[c]	95% CI	HR3[d]	95% CI
≤p75	428	71	1	ref	1	ref	1	ref
>p75 (high)	137	20	1.00	(0.60–1.67)	0.86	(0.49–1.51)	0.91	(0.51–1.62)
Per SD[e]	565	91	0.94	(0.76–1.16)	0.90	(0.71–1.13)	0.92	(0.73–1.17)
FXI:C	**n**	**Combined EP events**	**HR1**	**95% CI**	**HR2**	**95% CI**	**HR3**	**95% CI**
≤p75	433	63	1	ref	1	ref	1	ref
>p75 (high)	137	29	1.81	(1.15–2.84)	1.80	(1.09–2.98)	1.84	(1.11–3.07)
Per SD	570	92	1.26	(1.04–1.53)	1.26	(1.00–1.58)	1.23	(0.98–1.55)
FVIII:C	**n**	**Combined EP events**	**HR1**	**95% CI**	**HR2**	**95% CI**	**HR3**	**95% CI**
≤p75	420	54	1	ref	1	ref	1	ref
>p75 (high)	140	38	2.10	(1.38–3.19)	2.05	(1.28–3.29)	2.18	(1.35–3.52)
Per SD	560	92	1.33	(1.11–1.59)	1.37	(1.10–1.71)	1.37	(1.10–1.72)

Abbreviations: CI, confidence interval; EP, endpoint; FVIII:C, coagulation factor VIII activity; FXI:C, coagulation factor XI activity; FXII:C, coagulation factor XII activity; HR, hazard ratio; p75, 75th percentile; ref, reference; SD, standard deviation.

[a]Activities were measured as percentages of activated normal pooled plasma.

[b]Model 1: Adjusted for age and sex.

[c]Model 2: Adjusted for Model 1 variables plus BMI, HDL, LDL, smoking status, regular alcohol consumption, hypertension, diabetes mellitus, and acute coronary syndrome (see Methods for detailed variable definitions).

[d]Model 3 (sensitivity analysis): Adjusted for Model 2 variables plus thrombolysis treatment (rt-PA) status and NIHSS (National Institutes of Health Stroke Scale).

[e]One standard deviation of FXII:C, FXI:C, and FVIII:C levels were 29.3, 28.8, and 45.4 units, respectively.

4.1 | Strengths and limitations

Though the relationships between coagulation factors FXII, FXI, and FVIII and primary thrombotic event risk, for both venous and arterial events, have been well studied, to the best of our knowledge, this is the first study to investigate their role in long-term vascular event risk after first stroke. Furthermore, a limited number of studies have included multiple coagulation factors in a single study, and existing studies often lack follow-up beyond hospital discharge. Coagulation activity was assayed using state-of-the-art machinery on fresh-frozen, once-thawed plasma samples as opposed to antigen level measurements. Our longitudinal design with 94 observed outcome events also afforded us the opportunity to contribute our analyses of long-term outcome risk to the literature, which is currently limited to cross-sectional and very-short-term designs.

We were able to screen for additional, unreported clinical endpoints of interest in the Charité University Hospital medical records to supplement the information provided by the patients; however, unreported clinical endpoints presenting at other clinics in Berlin or elsewhere may have been missed. When possible, we made an effort to validate any patient-reported events by requesting forwarding of medical records from other hospitals and clinics. We further confirmed the vital status for all participants at the end of the study via the local citizen's registration office in Berlin; however, due to legal restrictions, we could not obtain specific cause information from the death certificates.

Some limitations should also be considered when interpreting our results. First, self-reported patient characteristics, such as the lifestyle-related factors and the presence of chronic diseases at baseline, may be prone to recall bias. A set of standard operating procedures and training was provided for the study nurses in an effort to improve consistency in the measurements made at study enrollment. Although we believe that we have included the most important potential sources of confounding in the adjusted models, we cannot rule out that some residual confounding may be present due to unmeasured factors.

Second, we emphasize that the coagulation factor activity levels were measured in blood samples that were taken after the index stroke event, and the initial stroke event itself may activate the coagulation system.[34] Though FXII and FXI are not known to change dramatically during the acute phase of stroke, this phenomenon has been well documented for FVIII. This means that our findings for FVIII are likely a mixing of elevated FVIII as part of the acute phase and high pre-stroke FVIII levels (increase of thrombotic event risk independent of the acute phase[18]). It is also possible that these activity levels changed within the first week post-stroke, during which the blood was drawn. About one fifth of the study participants received rt-PA treatment, and we cannot rule out that some may have been on anticoagulation therapy at the time of the initial stroke event. Time-standardized measurements should be considered in future confirmatory studies, and sequential measurements could provide additional relevant insights into the changes that occur shortly after stroke.

Our reported results apply to a cohort comprised of first-ever mild to moderate ischemic stroke patients. The six patients enrolled with a baseline NIHSS >15 in the PROSCIS-B study were excluded to limit the heterogeneity of the cohort. Readers should take care not to extrapolate our conclusions to severe patients (NIHSS >15) or patients with severe comorbidities or complications (such as sepsis) who may have been less likely to participate in our study.

We do not expect that the censoring of individuals lost to follow-up was differential with respect to exposure status. However, it is feasible, despite our efforts to confirm unreported endpoints, that those who were lost to follow-up may have been more likely to experience one of the combined vascular endpoints compared to the participants actively remaining in the study. Lastly, due to a limited number of endpoint events per included independent variable in Models 2 and 3, we acknowledge that the measured associations of interest in these models may be imprecise and could even be biased in the direction of more extreme values. However, especially for observational studies with causal questions like ours in which full confounding control is crucial, simulation results indicate that use of less rigid events per variable criteria is often justifiable for Cox regression analysis in terms of important model performance measures and in the range of 5 to 9 events per variable (our Models 2 and 3), such problems are uncommon.[35]

5 | CONCLUSIONS

Our study of mild to moderate ischemic stroke patients indicates that high levels of FXI:C or FVIII:C measured within 1 week of the index event may contribute to unfavorable vascular outcomes after stroke in the longer term (3 years). We did not observe a clear relationship with FXII:C. Further research in this area should focus on obtaining time-standardized and repeated measures of coagulation factor activities after stroke. In the context of secondary prevention, we demonstrated that individuals with high levels of FXI:C after stroke have an increased risk for secondary events. This knowledge may be beneficial for potential future treatment strategies involving drugs targeting FXI.

ACKNOWLEDGMENTS

We thank P. Noordijk, L. Mahic, and A. Hoenderdos from the Leiden University Medical Center for performing laboratory assays and J. Thümmler for PROSCIS-B data management. We are indebted to the participants of the PROSCIS-B for their commitment. JLR is a PhD candidate at the Charité – Universitätsmedizin Berlin, and this work is submitted in partial fulfillment of her PhD requirements. Open access funding enabled and organized by ProjektDEAL.

CONFLICTS OF INTEREST

Jessica L. Rohmann, Sophie K. Piper, Thomas G. Liman, and Bob Siegerink report no disclosures related to this work. Shufan Huo reports funding from the Sonnenfeld Foundation from January 2017 until March 2018. Pia S. Sperber reports funding from FAZIT-STIFTUNG between March 2018 and March 2020. Frits R.

Rosendaal is listed on patents of several prothrombotic variants, including factor XI. Peter U. Heuschmann reports research grants from the German Ministry of Research and Education, German Research Foundation, European Union, Charité, Berlin Chamber of Physicians, German Parkinson Society, University Hospital Würzburg, Robert-Koch-Institute, German Heart Foundation, Federal Joint Committee (G-BA) within the Innovationsfond, Charité – Universitätsmedizin Berlin (within MonDAFIS; supported by an unrestricted research grant to the Charité from Bayer), University Göttingen (within FIND-AF-randomized; supported by an unrestricted research grant to the University Göttingen from Boehringer-Ingelheim), and University Hospital Heidelberg (within RASUNOA-prime; supported by an unrestricted research grant to the University Hospital Heidelberg from Bayer, BMS, Boehringer-Ingelheim, Daiichi Sankyo), outside of the submitted work. Matthias Endres reports grant support from Bayer, the German Research Foundation (DFG), the German Federal Ministry of Education and Research (BMBF), the German Center for Neurodegenerative Diseases (DZNE), the German Centre for Cardiovascular Research (DZHK), the European Union, Corona Foundation, and Fondation Leducq; fees paid to the Charité from Bayer, Boehringer Ingelheim, Bristol-Myers Squibb/Pfizer, Daiichi Sankyo, Amgen, GlaxoSmithKline, Sanofi, Covidien, and Novartis, all outside of the submitted work.

AUTHOR CONTRIBUTIONS

J. L. Rohmann contributed to study conception and design, coagulation factor assays, analysis and interpretation of the data, manuscript drafting, and critical revision. S. Huo and P. S. Sperber contributed to data collection, interpretation, and critical revision of the manuscript. S. K. Piper contributed to project data management and critical revision of the manuscript. F. R. Rosendaal facilitated the laboratory assays and their interpretation and contributed to critical revision of the manuscript. P. U. Heuschmann and M. Endres were involved in original study conception and design as well as critical revision of the manuscript. T. G. Liman also contributed to study conception and design, was responsible for data collection, and contributed to critical revision of the manuscript. B. Siegerink contributed to study concept and design, analysis and interpretation of data, critical revision of the manuscript, and provided project supervision.

ORCID

Jessica L. Rohmann https://orcid.org/0000-0003-2420-5716
Bob Siegerink https://orcid.org/0000-0002-8454-9142

TWITTER

Jessica L. Rohmann @JLRohmann
Bob Siegerink @bobsiegerink

REFERENCES

1. Institute for Health Metrics and Evaluation (IHME). *GBD Compare Data Visualization*. Seattle, WA: IHME, University of Washington; 2018.
2. Thrift AG, Thayabaranathan T, Howard G, et al. Global stroke statistics. *Int J Stroke*. 2017;12:13-32.
3. Mohan KM, Wolfe CDA, Rudd AG, Heuschmann PU, Kolominsky-Rabas PL, Grieve AP. Risk and cumulative risk of stroke recurrence: a systematic review and meta-analysis. *Stroke*. 2011;42:1489-1494.
4. Dennis MS, Burn JP, Sandercock PA, Bamford JM, Wade DT, Warlow CP. Long-term survival after first-ever stroke: the Oxfordshire community stroke project. *Stroke*. 1993;24:796-800.
5. Sacco RL, Wolf PA, Kannel WB, McNamara PM. Survival and recurrence following stroke. The Framingham study. *Stroke*. 1982;13:290-295.
6. Donkel SJ, Benaddi B, Dippel DWJ, Ten Cate H, de Maat MPM. Prognostic hemostasis biomarkers in acute ischemic stroke. *Arterioscler Thromb Vasc Biol*. 2019;39:360-372.
7. Gailani D, Bane CE, Gruber A. Factor XI and contact activation as targets for antithrombotic therapy. *J Thromb Haemost*. 2015;13:1383-1395.
8. Kraft P, De Meyer SF, Kleinschnitz C. Next-generation antithrombotics in ischemic stroke: preclinical perspective on "bleeding-free antithrombosis". *J Cereb Blood Flow Metab*. 2012;32:1831-1840.
9. Siegerink B, Govers-Riemslag JWP, Rosendaal FR, Ten Cate H, Algra A. Intrinsic coagulation activation and the risk of arterial thrombosis in young women: results from the risk of arterial thrombosis in relation to oral contraceptives (RATIO) case-control study. *Circulation*. 2010;122:1854-1861.
10. Govers-Riemslag JWP, Smid M, Cooper JA, et al. The plasma kallikrein-kinin system and risk of cardiovascular disease in men. *J Thromb Haemost*. 2007;5:1896-1903.
11. Yang DT, Flanders MM, Kim H, Rodgers GM. Elevated factor XI activity levels are associated with an increased odds ratio for cerebrovascular events. *Am J Clin Pathol*. 2006;126:411-415.
12. Suri MFK, Yamagishi K, Aleksic N, Hannan PJ, Folsom AR. Novel hemostatic factor levels and risk of ischemic stroke: the atherosclerosis risk in communities (ARIC) study. *Cerebrovasc Dis*. 2010;29:497-502.
13. Siegerink B, Rosendaal FR, Algra A. Antigen levels of coagulation factor XII, coagulation factor XI and prekallikrein, and the risk of myocardial infarction and ischemic stroke in young women. *J Thromb Haemost*. 2014;12:606-613.
14. Borissoff JI, Spronk HMH, ten Cate H. The hemostatic system as a modulator of atherosclerosis. *N Engl J Med*. 2011;364:1746-1760.
15. Siegler JE, Samai A, Albright KC, Boehme AK, Martin-Schild S. Factoring in factor VIII with acute ischemic stroke. *Clin Appl Thromb Hemost*. 2015;21:597-602.
16. Chang TR, Albright KC, Boehme AK, et al. Factor VIII in the setting of acute ischemic stroke among patients with suspected hypercoagulable state. *Clin Appl Thromb Hemost*. 2014;20:124-128.
17. Koster T, Blann AD, Briët E, Vandenbroucke JP, Rosendaal FR. Role of clotting factor VIII in effect of von Willebrand factor on occurrence of deep-vein thrombosis. *Lancet*. 1995;345:152-155.
18. Kamphuisen PW, Eikenboom JC, Bertina RM. Elevated factor VIII levels and the risk of thrombosis. *Arterioscler Thromb Vasc Biol*. 2001;21:731-738.
19. Gouse BM, Boehme AK, Monlezun DJ, et al. New thrombotic events in ischemic stroke patients with elevated factor VIII. *Thrombosis*. 2014;2014:302861.
20. Liman TG, Zietemann V, Wiedmann S, et al. Prediction of vascular risk after stroke - protocol and pilot data of the Prospective Cohort with Incident Stroke (PROSCIS). *Int J Stroke*. 2013;8:484-490.
21. Hatano S. Experience from a multicentre stroke register: a preliminary report. *Bull World Health Organ*. 1976;54:541-553.
22. Adams HP Jr, Bendixen BH, Kappelle LJ, 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.

23. Lyden P. Using the national institutes of health stroke scale: a cautionary tale. *Stroke*. 2017;48:513-519.
24. Undas A, Slowik A, Gissel M, Mann KG, Butenas S. Active tissue factor and activated factor XI in patients with acute ischemic cerebrovascular events. *Eur J Clin Invest*. 2012;42:123-129.
25. Tóth NK, Székely EG, Czuriga-Kovács KR, et al. Elevated factor VIII and von Willebrand factor levels predict unfavorable outcome in stroke patients treated with intravenous thrombolysis. *Front Neurol*. 2017;8:721.
26. Maino A, Rosendaal FR, Algra A, Peyvandi F, Siegerink B. Hypercoagulability is a stronger risk factor for ischaemic stroke than for myocardial infarction: a systematic review. *PLoS One*. 2015;10:e0133523.
27. Weitz JI, Fredenburgh JC. Factors XI and XII as targets for new anticoagulants. *Front Med*. 2017;4:19
28. van Montfoort ML, Meijers JCM. Anticoagulation beyond direct thrombin and factor Xa inhibitors: indications for targeting the intrinsic pathway? *Thromb Haemost*. 2013;110:223-232.
29. Büller HR, Bethune C, Bhanot S, et al. Factor XI antisense oligonucleotide for prevention of venous thrombosis. *N Engl J Med*. 2015;372:232-240.
30. Georgi B, Mielke J, Chaffin M, et al. Leveraging human genetics to estimate clinical risk reductions achievable by inhibiting factor XI. *Stroke*. 2019;50:3004-3012.
31. Gill D, Georgakis MK, Laffan M, et al. Genetically determined FXI (Factor XI) levels and risk of stroke. *Stroke*. 2018;49:2761-2763.
32. Chong M, Sjaarda J, Pigeyre M, et al. Novel drug targets for ischemic stroke identified through mendelian randomization analysis of the blood proteome. *Circulation*. 2019;140:819-830.
33. Harshfield EL, Sims MC, Traylor M, Ouwehand WH, Markus HS. The role of haematological traits in risk of ischaemic stroke and its subtypes. *Brain*. 2020;143:210-221.
34. Brouns R, De Deyn PP. The complexity of neurobiological processes in acute ischemic stroke. *Clin Neurol Neurosurg*. 2009;111:483-495.
35. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol*. 2007;165:710-718.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Rohmann JL, Huo S, Sperber PS, et al. Coagulation factor XII, XI, and VIII activity levels and secondary events after first ischemic stroke. *J. Thromb. Haemost*. 2020;18:3316–3324. https://doi.org/10.1111/jth.15092

13.3. Publication 3: *Coagulation factor VIII, white matter hyperintensities and cognitive function: Results from the Cardiovascular Health Study*

Rohmann JL, Longstreth WT Jr, Cushman M, Fitzpatrick AL, Heckbert SR, Rice K, Rosendaal FR, Sitlani CM, Psaty BM, Siegerink B. Coagulation factor VIII, white matter hyperintensities and cognitive function: Results from the Cardiovascular Health Study. PLoS One. 2020 Nov 16;15(11):e0242062. Available from: http://dx.doi.org/10.1371/journal.pone.0242062

The extract from ISI Web of Knowledge Journal Summary List and the original published version of the article follow on the next 17 pages.

Journal Data Filtered By: **Selected JCR Year: 2018** Selected Editions: SCIE,SSCI Selected Categories: **"MULTIDISCIPLINARY SCIENCES"** Selected Category Scheme: WoS

Gesamtanzahl: 69 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	NATURE	745,692	43.070	1.285010
2	SCIENCE	680,994	41.037	1.070190
3	National Science Review	1,842	13.222	0.006500
4	Science Advances	21,901	12.804	0.110010
5	Nature Communications	243,793	11.878	1.103290
6	Nature Human Behaviour	1,230	10.575	0.006550
7	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA	661,118	9.580	1.022190
8	Science Bulletin	3,569	6.277	0.009840
9	Scientific Data	3,240	5.929	0.015610
10	Frontiers in Bioengineering and Biotechnology	1,994	5.122	0.006540
11	Journal of Advanced Research	2,691	5.045	0.004780
12	Research Synthesis Methods	1,932	5.043	0.005420
13	GigaScience	2,674	4.688	0.012510
14	Annals of the New York Academy of Sciences	46,385	4.295	0.025840
15	Scientific Reports	302,086	4.011	1.061540
16	Journal of the Royal Society Interface	12,933	3.224	0.029190
17	NPJ Microgravity	203	3.111	0.000670
18	PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A-MATHEMATICAL PHYSICAL AND ENGINEERING SCIENCES	19,227	3.093	0.028200

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
19	FRACTALS-COMPLEX GEOMETRY PATTERNS AND SCALING IN NATURE AND SOCIETY	1,429	2.971	0.001120
20	Journal of Radiation Research and Applied Sciences	860	2.963	0.001860
21	MIT Technology Review	929	2.893	0.001910
22	JOURNAL OF KING SAUD UNIVERSITY SCIENCE	1,120	2.835	0.001670
23	PROCEEDINGS OF THE ROYAL SOCIETY A-MATHEMATICAL PHYSICAL AND ENGINEERING SCIENCES	18,683	2.818	0.018940
24	PLoS One	650,727	2.776	1.706770
25	COMPLEXITY	2,753	2.591	0.003890
26	Royal Society Open Science	4,118	2.515	0.017150
27	PeerJ	11,911	2.353	0.045900
28	SCIENCE AND ENGINEERING ETHICS	1,719	2.275	0.003450
29	INTERNATIONAL JOURNAL OF BIFURCATION AND CHAOS	7,008	2.145	0.007390
30	Symmetry-Basel	2,097	2.143	0.002590
31	SCIENTIFIC AMERICAN	6,609	1.946	0.003540
32	Science of Nature	508	1.839	0.002000
33	PROCEEDINGS OF THE JAPAN ACADEMY SERIES B-PHYSICAL AND BIOLOGICAL SCIENCES	1,532	1.833	0.001960
34	Journal of Taibah University for Science	779	1.640	0.001240
35	Frontiers in Life Science	241	1.622	0.000500
36	ARABIAN JOURNAL FOR SCIENCE AND ENGINEERING	3,838	1.518	0.005840
37	SCIENCE PROGRESS	521	1.500	0.000400

Selected JCR Year: 2018; Selected Categories: "MULTIDISCIPLINARY SCIENCES"

PLOS ONE

RESEARCH ARTICLE

Coagulation factor VIII, white matter hyperintensities and cognitive function: Results from the Cardiovascular Health Study

Jessica L. Rohmann[1,2]*, W. T. Longstreth, Jr.[3,4], Mary Cushman[5], Annette L. Fitzpatrick[4,6,7], Susan R. Heckbert[4,8], Kenneth Rice[9], Frits R. Rosendaal[10], Colleen M. Sitlani[8], Bruce M. Psaty[4,8,11,12,13], Bob Siegerink[1,10]

1 Charité –Universitätsmedizin Berlin, Center for Stroke Research Berlin, Berlin, Germany, **2** Charité – Universitätsmedizin Berlin, Institute of Public Health, Berlin, Germany, **3** Department of Neurology, University of Washington, Seattle, WA, United States of America, **4** Department of Epidemiology, University of Washington, Seattle, WA, United States of America, **5** Department of Medicine, Larner College of Medicine at the University of Vermont, Burlington, VT, United States of America, **6** Department of Family Medicine, University of Washington, Seattle, WA, United States of America, **7** Department of Global Health, University of Washington, Seattle, WA, United States of America, **8** Cardiovascular Health Research Unit, University of Washington, Seattle, WA, United States of America, **9** Department of Biostatistics, University of Washington, Seattle, WA, United States of America, **10** Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands, **11** Department of Medicine, University of Washington, Seattle, WA, United States of America, **12** Department of Health Services, University of Washington, Seattle, WA, United States of America, **13** Kaiser Permanente Washington Health Research Institute, Seattle, WA, United States of America

* jessica.rohmann@charite.de

OPEN ACCESS

Citation: Rohmann JL, Longstreth WT, Jr., Cushman M, Fitzpatrick AL, Heckbert SR, Rice K, et al. (2020) Coagulation factor VIII, white matter hyperintensities and cognitive function: Results from the Cardiovascular Health Study. PLoS ONE 15(11): e0242062. https://doi.org/10.1371/journal.pone.0242062

Editor: Xi Chen, McLean Hospital, UNITED STATES

Received: November 12, 2019

Accepted: October 26, 2020

Published: November 16, 2020

Data Availability Statement: The CHS data are available for any researcher upon reasonable request, as described on the CHS website: https://chs-nhlbi.org. CHS data can be accessed through the NHLBI BioLINCC repository (https://biolincc.nhlbi.nih.gov/home/).

Funding: This research was supported by contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083,

Abstract

Objective

To investigate the relationship between high FVIII clotting activity (FVIII:C), MRI-defined white matter hyperintensities (WMH) and cognitive function over time.

Methods

Data from the population-based Cardiovascular Health Study (n = 5,888, aged ≥65) were used. FVIII:C was measured in blood samples taken at baseline. WMH burden was assessed on two cranial MRI scans taken roughly 5 years apart. Cognitive function was assessed annually using the Modified Mini-Mental State Examination (3MSE) and Digit Symbol Substitution Test (DSST). We used ordinal logistic regression models adjusted for demographic and cardiovascular factors in cross-sectional and longitudinal WMH analyses, and adjusted linear regression and linear mixed models in the analyses of cognitive function.

Results

After adjustment for confounding, higher levels of FVIII:C were not strongly associated with the burden of WMH on the initial MRI scan (OR>p75 = 1.20, 95% CI 0.99–1.45; N = 2,735) nor with WMH burden worsening over time (OR>p75 = 1.18, 95% CI 0.87–1.59; N = 1,527). High FVIII:C showed no strong association with cognitive scores cross-sectionally

N01HC85086, and grants U01HL080295 and U01HL130114 from the National Heart, Lung, and Blood Institute (NHLBI; https://www.nhlbi.nih.gov/), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS; https://www.ninds.nih.gov/). Additional support was provided by R01AG023629 from the National Institute on Aging (NIA; https://www.nia.nih.gov). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This work was also supported by the German Federal Ministry of Education and Research (BMBF; https://www.bmbf.de/en/index.html, BMBF 01 EO 08 01). We acknowledge support from the German Research Foundation (DFG) and the Open Access Publication Fund of Charité – Universitätsmedizin Berlin. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: JLR, MC, ALF, SRH, KR, FRR, CMS and BS have no disclosures to report. WTL reports no disclosures except those related to funding from the NIH. BMP serves on the DSMB of a clinical trial funded by Zoll LifeCor and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

(3MSE>p75 β = -0.06, 95%CI -0.45 to 0.32, N = 4,005; DSST>p75 β = -0.69, 95%CI -1.52 to 0.13, N = 3,954) or over time (3MSE>p75 β = -0.07,95% CI -0.58 to 0.44, N = 2,764; DSST>p75 β = -0.22, 95% CI -0.97 to 0.53, N = 2,306) after confounding adjustment.

Interpretation

The results from this cohort study of older adult participants indicate no strong relationships between higher FVIII:C levels and WMH burden or cognitive function in cross-sectional and longitudinal analyses.

Introduction

White matter hyperintensities (WMH) on MRI are common among older adults. According to a Dutch community-based study, the prevalence of WMH in healthy volunteers aged between 60 and 90 years was estimated to be 95%, and both prevalence and severity were found to increase with age [1]. In addition to being associated with traditional cardiovascular risk factors, changes in brain morphology, including WMH development, have been implicated in cognitive decline, incident cognitive impairment, and the development of dementia [2–6].

Coagulation factor FVIII (FVIII) levels generally increase with age [2]. Acute elevation of FVIII is known to occur during the acute phase of stroke as part of the inflammatory response [3]. A dose-dependent relationship between FVIII levels and the occurrence of thrombotic events (including overt ischemic stroke) has also been observed [4–8]. Further studies have linked FVIII, a potential therapeutic target, with dementia risk [9,10] as well as risk for cognitive impairment in a study of men [11]. Previous research in the US-based Cardiovascular Health Study (CHS) found an association between higher FVIII and incident cardiovascular disease, stroke, and death in the older general study population [12], but these studies did not investigate covert WMH or cognitive decline.

It remains unclear whether FVIII may contribute to the severity of covert WMH burden and worsening over time. It is also unknown whether high FVIII levels contribute to cognitive decline in this population, a process that may, in turn, be mediated by WMH. The present study aims to probe these relationships to better understand the role of FVIII, if any, in the pathways leading to cognitive decline in both cross-sectional and longitudinal settings, using data from a large, population-based cohort.

Methods

Participants and design

In this cohort study, we used data from the Cardiovascular Health Study (CHS) in all analyses. The design of the full longitudinal, population-based CHS, which aimed to assess risk factors for cardiovascular disease is described in detail elsewhere [13]. Briefly, the original CHS cohort included adults aged 65 and older recruited from four United States communities using Medicare eligibility lists. The first cohort of participants (N = 5,201) were enrolled in 1989 and 1990. A second cohort oversampling African-Americans (n = 687) was enrolled in 1992 and 1993; however, since FVIII levels were not measured in this group, the second cohort could not be included in our analyses. As illustrated in Fig 1, participants were prospectively followed for 9 years after the baseline visit and completed yearly clinic visits or phone interviews. An

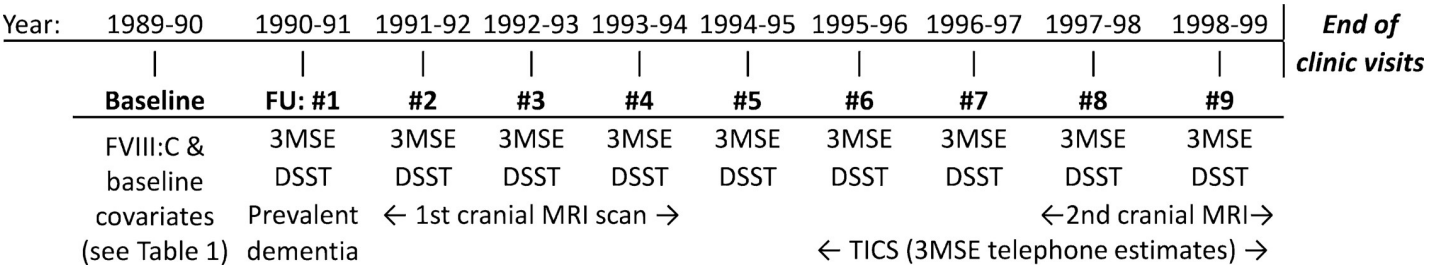

Fig 1. Timeline of CHS data collection. Measurements collected at baseline and at each follow-up are listed per year. Abbreviations: FU, follow-up; FVIII:C, coagulation factor VIII activity; 3MSE, Modified Mini-Mental State Examination; DSST, Digit Symbol Substitution Test; MRI, magnetic resonance imaging, TICS, Telephone Interview for Cognitive Status.

https://doi.org/10.1371/journal.pone.0242062.g001

overview of all relevant variables used in our study and a timeline of their acquisition in the CHS is shown in Fig 1.

Baseline assessment

The exposure variable of interest, FVIII clotting activity (FVIII:C), was measured in blood samples of 5,112 participants taken at baseline. Clotting activity was assayed using the Coag-a-mate X2 instrument with WHO standards, and activity units are expressed as percentages of normal pooled plasma [14]. The mean coefficients of variation of two control pools were 9% and 10%.

Other relevant variables measured at baseline included self-reported age, sex, and ethnicity. Participants were also asked to provide information about their highest level of completed education and information regarding smoking status and alcohol consumption. Smoking status was categorized as never, former, or current. Alcohol consumption was classified as never, occasionally, or frequently, which was defined as consuming more than 7 drinks per week, on average.

We considered other measurements taken at the baseline clinic visits as markers for cardiovascular risk. Participants' measured weight (kg) and squared height (m^2) were used to compute body mass index (BMI). Individuals presenting with $\geq$140 mmHg (systolic) or $\geq$90 mmHg (diastolic) seated blood pressure at baseline were classified as 'hypertensive' in addition to those who both reported a history of hypertension and were taking antihypertensive medication at baseline. High density lipoprotein cholesterol (HDL, mg/dL) and adjusted low-density lipoprotein cholesterol, (LDL, mg/dL) were measured from blood samples and recorded as continuous variables. Furthermore, participants' fasting plasma glucose levels were used to determine diabetes status according to current American Diabetes Association guidelines as 'normal'($<$100mg/dL), 'impaired fasting glucose' (100-125mg/dL), or 'diabetic' ($\geq$126 or reported taking insulin or oral hypoglycemics). Maximum common carotid intima-media thickness (CIMT) and maximum internal CIMT were defined as the mean of maximum wall thickness measurements made during all scans using ultrasonography used as indicators of carotid atherosclerosis [15]. Further relevant blood biomarker measurements included C reactive protein (CRP, mg/L) and fibrinogen (mg/dL). In addition to information on whether the participant had a history of stroke or transient ischemic attack (TIA) at baseline, incident stroke or incident TIA events occurring during CHS follow-up were also recorded [16].

Ascertainment of outcomes

Cranial MRI scans were performed in two waves over several years. The initial scan occurred during the 2nd through 4th follow-up visits (1991–1994), and the follow-up scan during the

8th and 9th follow-up visits (1997–1999), allowing for the assessment of changes in brain morphology in CHS participants over an average time interval of 5 years [17,18]. Areas of hyperintensities in the periventricular and subcortical regions as observed on standardized sagittal axial-spin density/T2-weighted cranial MRI images were used to quantify WMH burden, one of our outcomes of interest [19]. The WMH burden apparent on each scan was evaluated by experienced neuroradiologists at a central CHS Reading Center using a 10-point white matter grade (WMG) scale, ranging from 0 (no lesions) to 9 (most lesions) based on a library of templates [18]. In additional, the initial and follow-up MRI scans were read side-by-side to determine any worsening of the WMG between the two scans [17].

Cognitive ability was measured using multiple assessment tools. The first, the 100-point Modified Mini-Mental State Examination (3MSE) [20], was introduced during the first follow-up wave (1990–91) and administered annually thereafter during in-person clinic visits. Missing 3MSE scores could be estimated from the Telephone Interview for Cognitive Status (TICS) scores, when these data were available. TICS were first introduced during the sixth follow-up (1995–96), and these estimates have previously been confirmed to be reliable substitutions for 3MSE scores [21].

To provide an indication of the robustness of our results, we included an additional assessment of cognitive function as a second outcome: the Digit Symbol Substitution Test (DSST). This timed, 90-second test measures both attention and processing speed. The DSST overcomes the known challenge of the ceiling effect of the 3MSE [22]. The DSST was administered at baseline (1989–90) and annually thereafter during in-person clinic visits. For consistency between the two outcome measures, we considered only 3MSE scores (or the imputed TICS estimates thereof) and DSST scores measured during the 1st through 9th follow-up visits as continuous outcomes in our longitudinal analyses.

Inclusion/Exclusion criteria

In addition to the general inclusion and exclusion criteria of the CHS [13], we additionally excluded from our analyses all participants with missing FVIII:C measurements or with a history of one or more of the following at baseline: clinical stroke, TIA, prevalent dementia, and/or low cognitive function. Low cognitive function was defined as scoring less than 78 on the first 3MSE measurement, an established cutoff used in other CHS cognition studies [23], or in the lowest 10% of all scores on the first included DSST. Prevalent dementia and cognitive function (as measured by 3MSE) were first assessed in 1990–91.

Statistical analyses FVIII:C categorization

For our primary analyses, participants were categorized according to their FVIII activity levels (FVIII:C). High (>75th percentile) and low (≤25th percentile) FVIII:C levels were compared to the middle reference interval between the 25th and 75th percentiles, based on our *a priori* analysis plan. In secondary analyses, FVIII:C were analyzed continuously as normalized variables by dividing FVIII:C by the standard deviation (SD) of all FVIII:C measurements. To explore possible dose response, we used quintiles to group FVIII:C using the middle fifth as a reference. As a sensitivity analysis to assess robustness of our findings, we performed an additional categorization of FVIII:C. High (>90th percentile) and low (≤10th percentile) FVIII:C levels were compared to participants with FVIII:C in the 10th– 90th percentile interval (reference).

White matter hyperintensity burden at initial cranial MRI scan: Cross-sectional analyses

Cross-sectional analyses were conducted to determine whether FVIII:C levels were associated with WMH burden on the initial MRI scan (see Fig 1) in all participants having initial scan

results. For this analysis, the outcome variable, WMH burden, was grouped by grade into three groups (0–1, 2–3, 4–9). Differences in mean FVIII:C between WMG groups were assessed using an ANOVA.

We then used ordinal logistic regression to probe the relationship between FVIII:C and WMH burden using the gologit2 command in Stata, which can relax the proportional odds assumption as needed for specific explanatory variables (this is known as fitting a partial proportional odds model) [24,25]. In addition to an unadjusted model (model 1), we provide two additional models conditioning for potential confounding variables: model 2: adjusted for *a priori* selected demographic and socioeconomic factors including age, sex, ethnicity, and education level; and model 3: additionally adjusting for *a priori* selected lifestyle and cardiovascular risk factors: smoking, alcohol use, BMI, hypertension, diabetes, HDL and LDL cholesterol, fibrinogen, CRP, maximum common and internal CIMT, and the occurrence of a TIA or stroke event during the follow-up period prior to the initial MRI scan, as defined previously. Logarithmic transformations were used for variables with right-skewed distributions including both CIMT measurements and CRP. Since incident TIA or stroke events could be mediators for the exposure-outcome relationship, we conducted an additional sensitivity analysis in which this variable was omitted. We report ordinal odds ratios with corresponding 95% confidence intervals (95% CI) for each model.

Worsening of white matter grade between MRI scans: Longitudinal analyses

We report odds ratios with corresponding 95% CI from ordinal logistic regression models to quantify the worsening in WMG between the two MRI scans taken on average 5 years apart. Thus, only participants who underwent both MRI scans were included in these analyses. Change between scans was *a priori* categorized as no worsening in WMG, a worsening of 1 grade, or a worsening of 2 or more grades. The analyses were additionally adjusted for the elapsed time between participants' initial and follow-up MRI scans but not for baseline WML grade, as WML grade changes were likely to precede our baseline assessment and such adjustment is thus likely to introduce bias [26].

FVIII:C and cognitive function: Cross-sectional and longitudinal analyses

We used linear regression models to obtain estimates of the effect of baseline FVIII:C (β and 95% CIs) on 3MSE scores measured at the first follow-up in three adjusted models as described above. We then used linear mixed models with random intercepts for each individual to investigate the relationship between FVIII:C and cognitive function over time, as measured by serial 3MSE scores. These models handle missing data and the dependent nature of the repeated measurements within individuals. As a secondary analysis, we repeated both cross-sectional and longitudinal analyses using DSST scores instead of 3MSE. Effect estimates and 95% CIs are reported for the 3 models, as well as the model variance between individuals.

Standard protocol approvals, registrations and patient consents

Informed written consent was obtained from all CHS participants at entry into the study and at periodic intervals during follow-up. Institutional review boards at each CHS center approved the study. The IRB at the Charité - Universitätsmedizin Berlin also approved this secondary data analysis, and data was transferred in a fully anonymized form for analysis.

Results

Description of study population

After applying our inclusion and exclusion criteria as shown in the participant flowchart (Fig 2), our study population consisted of 4,295 participants whose characteristics are summarized in Table 1. The study participants had a mean age of 72.3 years, 59.3% were female, and 96.0% were white. In total, 45.5% of included participants were hypertensive and 13.9% were diabetic.

FVIII:C and white matter grade

Of the 2,735 participants who underwent the first MRI scan, mean FVIII:C differed among WMG groups (p = 0.001). Participants with the highest burden of WMH on this scan (grades 4–9) had higher mean FVIII:C (121.2) than participants with WMG of 2–3 (120.5) or WMG 0–1 (115.3) (see Table 2). Ordinal analyses using three groups of WMG (0–1, 2–3, and 4–9) revealed a weak association between high FVIII:C (>p75)–compared to normal levels–and WMG (model 3 OR = 1.20, 95% CI 0.99–1.45) in participants with initial MRI scan results (Table 2). Furthermore, each increase in FVIII:C by one standard deviation (36 units) was not associated with substantially worse WMG at the initial MRI scan after full adjustment (model 3 OR = 1.08, 95% CI 0.99–1.17).

We report results from longitudinal analyses in Table 3. Considering worsening of WMG between the two scans, after full adjustment for potential confounding factors, the ordinal logistic regression analysis revealed no meaningful association between high FVIII:C (>p75) and the degree of worsening over the time period between scans (model 3 OR = 1.18, 95% CI 0.87–1.59).

Similar findings were observed in secondary analyses considering per SD increase in FVIII:C as the exposure (model 3 OR = 1.07, 95% CI 0.94–1.22).

FVIII:C and cognitive function

Table 4 shows results from cross-sectional analyses using 3MSE (n = 4,005) and DSST (n = 3,954) scores from study year 1991–92 (first follow-up) as outcomes. Though a small association between high FVIII:C (>p75) and lower 3MSE scores was observed in the crude linear regression model (β_1 = -0.55; 95% CI -0.95 to -0.15), this association disappeared after full confounding adjustment (β_3 = -0.06, 95% CI -0.45 to 0.32). The unadjusted association with DSST score (β_1 = -2.05, 95% CI -2.94 to -1.16), was also greatly reduced after full adjustment (β_3 = -0.69, 95% CI -1.52 to 0.13). For the DSST scores, we observed a slight dose response in the crude quintile assessment. Per SD of FVIII:C, we also found no association between high FVIII:C and lower 3MSE or DSST scores in the fully adjusted models (Table 4).

Linear mixed-effects regression with random intercepts was then used to investigate whether FVIII:C measured at baseline were associated with cognitive ability over time (Table 5). The numbers of participants with available measurements in the final year of follow-up were N = 2,764 (3MSE or TICS estimate) and N = 2,306 (DSST). As was observed in the cross-sectional setting, having high (>p75) FVIII:C was not associated with 3MSE over the course of study follow-up in the fully adjusted models (β_3 = -0.07, 95% CI -0.58 to 0.44; β_3 per SD of FVIII:C = 0.15, 95% CI -0.06 to 0.37). Furthermore, no clinically relevant association was observed between high FVIII:C and change in DSST score over time, after adjustment (β_3 = -0.22, 95% CI -0.97 to 0.53; β_3 per SD of FVIII:C = -0.11, 95% CI -0.43 to 0.22).

As a sensitivity analysis, we repeated the above analyses with p10/p90 cutoffs, which did not substantially alter the results. Furthermore, omitting the variable 'occurrence of a TIA or

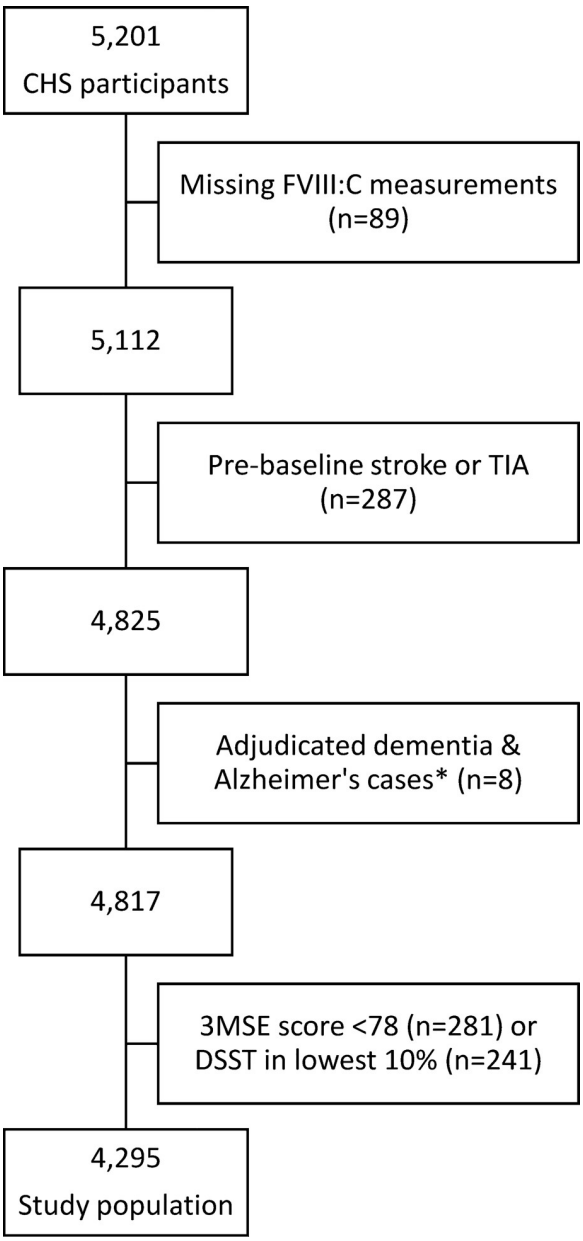

Fig 2. Study inclusion flow chart. Dementia was first adjudicated in 1990–91. DSST and 3MSE scores from that study year were also used to determine participant inclusion for this study. Abbreviations: FVIII:C, coagulation factor VIII activity; FU, follow-up; 3MSE, Modified Mini-Mental State Examination; DSST, Digit Symbol Substitution Test; MRI, magnetic resonance imaging; TICS, Telephone Interview for Cognitive Status; TIA, transient ischemic attack.

https://doi.org/10.1371/journal.pone.0242062.g002

stroke event during the follow-up' from the third model (as it may be both an intermediate and confounder) resulted in nearly identical point estimates.

Discussion

In the present study, we estimated the effects of FVIII:C on WMH burden and cognitive function both cross-sectionally and over time in a population-based sample of older persons in the CHS. Although participants with high WMG on the first MRI scan had higher mean FVIII:C

Table 1. Baseline characteristics of study population.

Characteristic (n = 4,295)	
Mean age (SD), years	72.3 (± 5.3)
Female sex, N (%)	2,549 (59.3%)
White, N (%)	4,124 (96.0%)
Education, N (%)[a]	
< Grade 12	986 (23.0%)
Completed high school/GED	1,296 (30.2%)
Vocational or some college	1,045 (24.3%)
Graduate degree/professional	960 (22.4%)
BMI (SD), kg/m^2	26.4 (± 4.5)
Hypertension, N (%)	1,950 (45.4%)
Cholesterol (SD), mg/dL	
Total	212.3 (± 39.0)
HDL	54.4 (± 15.9)
LDL	130.4 (± 35.4)
Diabetes, N (%)	
None	3,101 (72.2%)
Impaired fasting glucose	592 (13.8%)
Known or new	598 (13.9%)
Smoking, N (%)	
Never	1,971 (45.9%)
Former	1,841 (42.9%)
Current	481 (11.2%)
Alcohol use, N (%)	
Never	1,961 (45.7%)
Occasional	1,729 (40.3%)
Frequent	591 (13.8%)
FVIII:C (IQR), %[b]	116 (95–141)
CRP (IQR), mg/L	2.4 (1.2–4.2)
Fibrinogen (IQR), mg/dL	311 (270.0–361)
Max Common CIMT (IQR), mm	0.96 (0.86–1.08)
Max Internal CIMT (IQR), mm	1.33 (0.90–1.90)
3MSE score (IQR)[c]	94 (89–97)
DSST score (IQR)[d]	42 (35–51)

Abbreviations: GED = General Educational Development (high school equivalency diploma); BMI = body mass index; HDL = high-density lipoprotein; LDL = low-density lipoprotein; FVIII:C = coagulation factor VIII activity; CRP = C-reactive protein; CIMT = carotid intima-media thickness; 3MSE = Modified Mini-Mental State examination; DSST = Digit Symbol Substitution Test; IQR = interquartile range.

[a] Owing to missing data, percentages may not total 100. All variables have <2% missing values except 3MSE (6.8%), and DSST (7.9%).

[b] percentage of normal pooled plasma

[c] out of 100 possible points, measured during 1991–92 study year

[d] out of 90 possible points, measured during 1991–92 study year

https://doi.org/10.1371/journal.pone.0242062.t001

than participants with no or low WMG, no strong association was observed between high FVIII:C and WMH burden in the cross-sectional analyses. Furthermore, high FVIII:C did not appear to be linked to worsening of the WMG between MRI scans performed about 5 years apart after adjustment for confounding variables.

Table 2. FVIII:C and burden of white matter hyperintensities on initial cranial MRI scan (cross-sectional).

	WMG 0–1 (n = 995)	WMG 2–3 (n = 1,371)	WMG 4–9 (n = 369)	OR1[b]	95% CI	OR2[c]	95% CI	OR3[d]	95% CI
FVIII:C[a] groups									
≤ p25	285	337	97	1.06	(0.89–1.26)	1.03	(0.86–1.23)	1.03	(0.85–1.23)
p25-p75	522	690	182	1	ref	1	ref	1	ref
> p75	188	344	90	1.37	(1.11–1.68)	1.22	(1.02–1.46)	1.20	(0.99–1.45)
Continuous									
per SD increase of FVIII:C[e]				1.15	(1.07–1.24)	1.09	(1.01–1.17)	1.08	(0.99–1.17)

Abbreviations: FVIII:C = coagulation factor VIII activity; WMG = white matter grade; MRI = magnetic resonance imaging; OR = odds ratio; CI = confidence interval; p25 = 25th-percentile; p75 = 75th-percentile; ref = reference category; SD = standard deviation

[a] percentage of normal pooled plasma

[b] unadjusted model

[c] model adjusted for demographic risk factors (age, gender, ethnicity, education level)

[d] model additionally adjusted for cardiovascular risk factors (hypertension, smoking status, diabetes, alcohol use, BMI, HDL cholesterol, LDL cholesterol, fibrinogen, log-transformed C-reactive protein, log-transformed maximum common carotid intima-media thickness, log-transformed maximum internal carotid intima-media thickness, and occurrence of stroke or TIA prior to initial MRI scan)

[e] one standard deviation increase in FVIII:C corresponds to 36 units

https://doi.org/10.1371/journal.pone.0242062.t002

We additionally investigated the relationship between high FVIII:C and cognitive ability as measured by two different scores. Although crude analyses showed an association, no meaningful influence of high FVIII:C on 3MSE scores was observed in cross-sectional analyses after adjusting for potential sources of confounding [27–30]. FVIII:C showed a small association with DSST scores after full adjustment; on average, low FVIII:C (≤p25) levels were associated with scores 0.85 points higher (95%CI 0.06 to 1.64) and high levels (>p75) were associated with scores 0.69 points lower (95%CI -1.52 to 0.13) than the reference group (p25-p75). A dose response was observed across FVIII:C quintile groups and DSST scores in crude models, though the effect sizes attenuated with adjustment. No meaningful relationship was observed

Table 3. FVIII:C and white matter hyperintensity grade worsening on MRI scans over 5 years; results from ordinal logistic regression models.

	No change (n = 1,098)	1 grade worse (n = 367)	2+ grades worse (n = 62)	OR1[b]	95% CI	OR2[c]	95% CI	OR3[d]	95% CI
FVIII:C[a] groups									
≤ p25	318	99	20	1.02	(0.78–1.33)	1.03	(0.79–1.35)	1.01	(0.76–1.33)
p25-p75	552	185	25	1	ref	1	ref	1	ref
> p75	228	83	17	1.16	(0.88–1.55)	1.15	(0.87–1.54)	1.18	(0.87–1.59)
Continuous									
per SD increase of FVIII:C[e]				1.05	(0.93–1.18)	1.04	(0.93–1.18)	1.07	(0.94–1.22)

Abbreviations: FVIII:C = coagulation factor VIII activity; WMG = white matter grade; MRI = magnetic resonance imaging; OR = odds ratio; CI = confidence interval; p25 = 25th-percentile; p75 = 75th-percentile; ref = reference category; SD = standard deviation

[a] percentage of normal pooled plasma

[b] model adjusted for time interval (in years) between MRI scans

[c] model additionally adjusted for demographic risk factors (age, gender, ethnicity, education level)

[d] model additionally adjusted for cardiovascular risk factors (hypertension, smoking status, diabetes, alcohol use, BMI, HDL cholesterol, LDL cholesterol, fibrinogen, log-transformed C-reactive protein, log-transformed maximum common carotid intima-media thickness, log-transformed maximum internal carotid intima-media thickness, and occurrence of stroke or TIA during follow-up period before second MRI scan

[e] one standard deviation increase in FVIII:C corresponds to 36.3 units

https://doi.org/10.1371/journal.pone.0242062.t003

Table 4. FVIII:C and cognitive test scores: Cross-sectional results using linear regression.

3MSE scores (100 points maximum), study year 1991–92 (n = 4,005)						
	β_1 [b]	95% CI	β_2 [c]	95% CI	β_3 [d]	95% CI
FVIII:C[a] percentile groups						
≤ p25	0.61	(0.22 to 1.00)	0.33	(-0.03 to 0.69)	0.33	(-0.04 to 0.71)
p25-p75	0	ref	0	ref	0	ref
> p75	-0.55	(-0.95 to -0.15)	-0.16	(-0.53 to 0.21)	-0.06	(-0.45 to 0.32)
Continuous						
per SD increase of FVIII:C[e]	-0.40	(-0.56 to -0.23)	-0.16	(-0.31 to -0.01)	-0.14	(-0.30 to 0.03)
FVIII:C quintile groups						
Q1 (low)	0.22	(-0.24 to 0.68)	0.22	(-0.24 to 0.68)	0.22	(-0.25 to 0.69)
Q2	-0.24	(-0.70 to 0.22)	-0.24	(-0.70 to 0.22)	-0.25	(-0.72 to 0.22)
Q3	0	ref	0	ref	0	ref
Q4	-0.39	(-0.86 to 0.07)	-0.39	(-0.86 to 0.07)	-0.34	(-0.82 to 0.13)
Q5 (high)	-0.28	(-0.75 to 0.19)	-0.28	(-0.75 to 0.19)	-0.18	(-0.67 to 0.31)
DSST scores (90 points maximum), study year 1991–92 (n = 3,954)						
	β_1	95% CI	β_2	95% CI	β_3	95% CI
FVIII:C percentile groups						
≤ p25	1.58	(0.71 to 2.44)	0.90	(0.13 to 1.68)	0.85	(0.06 to 1.64)
p25-p75	0	ref	0	ref	0	ref
> p75	-2.05	(-2.94 to -1.16)	-1.01	(-1.81 to -0.22)	-0.69	(-1.52 to 0.13)
Continuous						
per SD increase of FVIII:C[e]	-1.16	(-1.53 to -0.80)	-0.54	(-0.87 to -0.21)	-0.37	(-0.72 to -0.02)
FVIII:C quintile groups						
Q1 (low)	1.37	(0.26 to 2.48)	0.66	(-0.33 to 1.65)	0.53	(-0.48 to 1.53)
Q2	0.86	(-0.26 to 1.97)	0.57	(-0.43 to 1.56)	0.47	(-0.54 to 1.47)
Q3	0	ref	0	ref	0	ref
Q4	-0.86	(-1.99 to 0.27)	-0.21	(-1.21 to 0.80)	-0.12	(-1.14 to 0.89)
Q5 (high)	-1.84	(-2.98 to -0.70)	-0.80	(-1.81 to 0.22)	-0.42	(-1.46 to 0.63)

Abbreviations: FVIII:C = coagulation factor VIII activity; 3MSE = Modified mini-mental state examination; DSST = Digit Symbol Substitution Test; CI = confidence interval; p25 = 25th-percentile; p75 = 75th-percentile; ref = reference category; SD = standard deviation.

β coefficients were calculated using linear regression in three models.

[a] percentage of normal pooled plasma

[b] unadjusted model

[c] model adjusted for demographic risk factors (age, gender, ethnicity, education level)

[d] model additionally adjusted for cardiovascular risk factors (hypertension, smoking status, diabetes, alcohol use, BMI, HDL cholesterol, LDL cholesterol, fibrinogen, log-transformed C-reactive protein, log-transformed maximum common carotid intima-media thickness, and log-transformed maximum internal carotid intima-media thickness).

[e] one standard deviation increase in FVIII:C corresponds to 36.3 units

https://doi.org/10.1371/journal.pone.0242062.t004

between high FVIII:C at baseline and cognitive worsening over time as measured by the 3MSE or DSST after full confounding adjustment.

Although the relationship between WMH and cognitive outcomes is well described in the literature [2–6], few studies have investigated the potential relationship between FVIII:C levels and WMH. An earlier CHS paper with exploratory aims included FVIII:C as one of 70 potential predictors for WMH worsening [22]. However, these exploratory analyses were distinct from the goals and corresponding approaches in our paper, as we aimed to obtain a confounding-adjusted estimates of the specific effect of FVIII:C on WMG as well as FVIII:C on cognitive

Table 5. FVIII:C and cognitive test scores: Longitudinal results using linear mixed-effects regression.

3MSE scores (100 points maximum)						
	β_1 [b]	95% CI	β_2 [c]	95% CI	β_3 [d]	95% CI
FVIII:C[a] percentile groups						
≤ p25	0.29	(-0.25 to 0.83)	-0.27	(-0.76 to 0.21)	-0.40	(-0.89 to 0.09)
p25-p75	0	ref	0	ref	0	ref
> p75	-1.09	(-1.64 to -0.54)	-0.34	(-0.83 to 0.15)	-0.07	(-0.58 to 0.44)
Model variance [e]	*46.7*	*(44.4 to 49.1)*	*35.7*	*(33.9 to 37.6)*	*34.9*	*(33.1 to 36.8)*
Continuous						
per SD increase of FVIII:C[f]	-0.51	(-0.73 to -0.28)	-0.02	(-0.22 to 0.19)	0.15	(-0.06 to 0.37)
DSST scores (90 points maximum)						
	β_1	95% CI	β_2	95% CI	β_3	95% CI
FVIII:C percentile groups						
≤ p25	1.33	(0.49 to 2.17)	0.50	(-0.21 to 1.21)	0.40	(-0.32 to 1.13)
p25-p75	0	ref	0	ref	0	ref
> p75	-1.87	(-2.73 to -1.01)	-0.69	(-1.42 to 0.04)	-0.22	(-0.97 to 0.53)
Model variance	*119.5*	*(114.1 to 125.1)*	*83.4*	*(79.5 to 87.4)*	*80.3*	*(76.5 to 84.3)*
Continuous						
per SD increase of FVIII:C	-1.11	(-1.46 to -0.76)	-0.37	(-0.67 to -0.06)	-0.11	(-0.43 to 0.22)

Abbreviations: FVIII:C = coagulation factor VIII activity; 3MSE = Modified mini-mental state examination; DSST = Digit Symbol Substitution Test; CI = confidence interval; p25 = 25th-percentile; p75 = 75th-percentile; ref = reference category; SD = standard deviation.

β coefficients were calculated using linear mixed-effects regression with random intercepts in three models.

[a] percentage of normal pooled plasma

[b] unadjusted model

[c] model adjusted for demographic risk factors (age, gender, ethnicity, education level)

[d] model additionally adjusted for cardiovascular risk factors (hypertension, smoking status, diabetes, alcohol use, BMI, HDL cholesterol, LDL cholesterol, fibrinogen, log-transformed C-reactive protein, log-transformed maximum common carotid intima-media thickness, log-transformed maximum internal carotid intima-media thickness, and the occurrence of stroke or TIA during follow-up).

[e] variance between individuals for each model for the primary exposure categorization (percentile groups)

[f] one standard deviation increase in FVIII:C corresponds to 36.3 units

https://doi.org/10.1371/journal.pone.0242062.t005

function. In the present study, potential sources of confounding were selected *a priori*, and the methodology was not dependent on the combined explained variance of a regression model.

Although not in line with our original hypotheses, these results do align with some findings from previous publications. For example, results from the ARIC study indicated that FVIII might have another role in brain morphology that is not specifically related to white matter lesion development; for instance, multiple midlife systemic inflammatory markers (including FVIII) were associated with brain volume late in life [31]. Another recent publication from the same cohort examined the relationship between an "inflammation composite score" that included factor VIII in middle-aged individuals and 20-year cognitive decline [32]. The authors found that each SD increase in the score was associated with additional decline, and that the highest scores were associated with steeper declines over the 20-year period [32]. Taken together, it is feasible that FVIII exerts an effect over a longer period than we captured in our study on cognitive function and by other pathways than via WMH development.

Results from the REGARDS study showed an association between high FVIII levels and cognitive impairment in crude analyses, which were attenuated after multivariable adjustment, similar to our findings [33,34]. These findings may indicate that high FVIII is a marker of one or more underlying processes that play a causal role in cognitive impairment with increasing

age, but is perhaps itself not a cause. Although our study did not investigate dementia diagnosis as an outcome, an earlier prospective study of middle-aged men found an association between factor VIII levels and other coagulation markers on vascular dementia after 17 years of follow-up (HR, 1.79; 95% CI, 1.09–3.00) [11]. More recently, a systematic review and meta-analysis synthesized findings from 32 identified studies on the role of the hemostatic system (including some coagulation factors) in cognitive decline in older persons. Among individuals with vascular dementia, higher levels of FVIII were observed together with other coagulation factors [35].

Study strengths and limitations

Strengths of this study include its population-based, prospective design, large number of older participants, and multiple outcomes and time points of their assessment. Information about many potential sources of confounding was collected and very few values were missing at baseline (<2% of all potential confounding variables). Furthermore, starting with the sixth follow-up, telephone-based estimates could be used to estimate missing 3MSE scores when in-person visits were not possible.

While we were able to control for confounding by including a larger number of sociodemographic and cardiovascular risk factors, we acknowledge that some residual confounding may still be present. However, adjusting for well-known, potentially strong sources of confounding, such as age, did not change our results materially, suggesting that confounding due to weaker unmeasured sources would not have substantially altered our results. It is also possible that Model 3 resulted in some overadjustment, as the occurrence of a stroke or TIA during follow-up could be both a source of time-dependent confounding and an intermediate on the causal path.

Our sensitivity analysis revealed only small differences in the point estimates when this variable was removed from the model, which would not change our interpretation of the results.

Readers should also consider some limitations with respect to the FVIII:C measurements when interpreting our findings. In addition to its implication in thrombosis, FVIII is also a known marker of relevant biological mechanisms such as inflammation and is an acute phase reactant. Though the measurements in the CHS were taken when participants were healthy at baseline, subclinical disease cannot be ruled out. The laboratory measurements of FVIII:C were somewhat variable in the CHS study (coefficients of variation 9% and 10%), however, we do not anticipate this variability in the laboratory measurements had any meaningful impact on our results with such a large sample size. Furthermore, in this secondary data analysis, only one FVIII:C measurement (taken at baseline) was available. When possible, future studies on this topic should consider repeating these measurements to better understand how changes in FVIII:C over time may impact WML and cognitive test performance in the healthy general population.

With respect to our outcome assessments, given the non-linear nature of the WML grading scale, we chose *a priori* to perform ordinal analyses to use all ordered information captured by WMG. However, due to small group size in some WMG categories, which was anticipated among the older general population, we collapsed the higher categories, thereby reducing the resolution of the data. Furthermore, for the longitudinal analyses of worsening WMG, only survivors remaining in the study through the eighth follow-up could be included due to the need for the follow-up MRI, a selection that inherently leads to selection bias. As shown previously, participants who underwent at least one MRI scan were healthier than those who were never scanned, and those with both the initial and follow-up scans were healthier than those who only had the initial MRI scan [17], however, this is a limitation common to all studies

involving MRI, and part of our motivation to also look at cognitive function measures. Future studies with more frequent scans over a longer time period may provide valuable insights, however, feasibility of such studies is remains a large obstacle.

We emphasize that these findings from an older, general population cohort should not be extrapolated to younger individuals or specific chronically ill populations. We also cannot rule out that the weak cross-sectional associations observed between FVIII:C and WMG may be explained by reverse causation. WML, as an inflammatory stimulus, could also be a contributing cause of higher FVIII:C levels.

With regard to the cognition outcomes, we acknowledge that the true relationship between high FVIII:C and 3MSE scores in both cross-sectional and longitudinal analyses may be stronger than what we observed due to the ceiling effect of the 3MSE and the high number of participants with high scores. However, upon comparing these results with a second measurement of cognitive function, the DSST (without a ceiling effect), we observed no material differences. We did not adjust any of our longitudinal models for baseline cognitive test scores, as this procedure introduces bias under some circumstances [26]. It is well known that individuals with lower cognitive function are less likely to remain in studies, therefore, attrition was likely differential with regard to the outcomes in the analyses using cognitive test scores. Therefore, the cognitive decline of study participants is likely greater than what we observed in the participants who remained in the study. Although we had two measures of cognitive function, we further acknowledge that neither assesses any one cognitive domain in depth, and readers should interpret our results with this limitation in mind.

Conclusions

Our findings from this large, population-based study indicate no strong relationship between FVIII:C and WMH burden or WMH worsening over time in the older general population following adjustment for demographic and cardiovascular factors. Furthermore, FVIII:C levels do not appear to be associated with cognitive worsening, as measured using serial 3MSE and DSST assessments. Though an important player in hemostatic balance and implicated in overt vascular disease, this study suggests that FVIII:C is not strongly related to covert, subclinical brain lesion development, and subsequent cognitive decline in older adults.

Acknowledgments

We are indebted to the participants of the Cardiovascular Health Study for their time and commitment. JLR is a PhD candidate at the Charité –Universitätsmedizin Berlin, and this submitted work is in partial fulfillment of her PhD requirements. BMP served as the CHS sponsor to JLR and BS.

Author Contributions

Conceptualization: Jessica L. Rohmann, W. T. Longstreth, Jr., Mary Cushman, Annette L. Fitzpatrick, Susan R. Heckbert, Kenneth Rice, Frits R. Rosendaal, Colleen M. Sitlani, Bruce M. Psaty, Bob Siegerink.

Data curation: W. T. Longstreth, Jr., Mary Cushman, Annette L. Fitzpatrick, Susan R. Heckbert, Bruce M. Psaty.

Formal analysis: Jessica L. Rohmann, Bob Siegerink.

Methodology: Jessica L. Rohmann, Kenneth Rice, Colleen M. Sitlani, Bob Siegerink.

Project administration: Bruce M. Psaty, Bob Siegerink.

Resources: W. T. Longstreth, Jr., Mary Cushman, Annette L. Fitzpatrick, Susan R. Heckbert, Bruce M. Psaty.

Software: Jessica L. Rohmann.

Supervision: Bruce M. Psaty, Bob Siegerink.

Writing – original draft: Jessica L. Rohmann.

Writing – review & editing: Jessica L. Rohmann, W. T. Longstreth, Jr., Mary Cushman, Annette L. Fitzpatrick, Susan R. Heckbert, Kenneth Rice, Frits R. Rosendaal, Colleen M. Sitlani, Bruce M. Psaty, Bob Siegerink.

References

1. de Leeuw FE, de Groot JC, Achten E, Oudkerk M, Ramos LM, Heijboer R, et al. Prevalence of cerebral white matter lesions in elderly people: a population based magnetic resonance imaging study. The Rotterdam Scan Study. J Neurol Neurosurg Psychiatry. 2001; 70: 9–14. https://doi.org/10.1136/jnnp.70.1.9 PMID: 11118240
2. Tracy RP, Bovill EG, Fried LP, Heiss G, Lee MH, Polak JF, et al. The distribution of coagulation factors VII and VIII and fibrinogen in adults over 65 years. Results from the Cardiovascular Health Study. Ann Epidemiol. 1992; 2: 509–519. https://doi.org/10.1016/1047-2797(92)90100-5 PMID: 1342301
3. Chang TR, Albright KC, Boehme AK, Dorsey A, Sartor EA, Kruse-Jarres R, et al. Factor VIII in the setting of acute ischemic stroke among patients with suspected hypercoagulable state. Clin Appl Thromb Hemost. 2014; 20: 124–128. https://doi.org/10.1177/1076029613488936 PMID: 23677913
4. Maino A, Rosendaal FR, Algra A, Peyvandi F, Siegerink B. Hypercoagulability Is a Stronger Risk Factor for Ischaemic Stroke than for Myocardial Infarction: A Systematic Review. PLoS One. 2015; 10: e0133523. https://doi.org/10.1371/journal.pone.0133523 PMID: 26252207
5. Borissoff JI, Spronk HMH, ten Cate H. The hemostatic system as a modulator of atherosclerosis. N Engl J Med. 2011; 364: 1746–1760. https://doi.org/10.1056/NEJMra1011670 PMID: 21542745
6. Siegler JE, Samai A, Albright KC, Boehme AK, Martin-Schild S. Factoring in Factor VIII With Acute Ischemic Stroke. Clin Appl Thromb Hemost. 2015; 21: 597–602. https://doi.org/10.1177/1076029615571630 PMID: 25669199
7. Koster T, Blann AD, Briët E, Vandenbroucke JP, Rosendaal FR. Role of clotting factor VIII in effect of von Willebrand factor on occurrence of deep-vein thrombosis. Lancet. 1995; 345: 152–155. https://doi.org/10.1016/s0140-6736(95)90166-3 PMID: 7823669
8. Kamphuisen PW, Eikenboom JC, Bertina RM. Elevated factor VIII levels and the risk of thrombosis. Arterioscler Thromb Vasc Biol. 2001; 21: 731–738. https://doi.org/10.1161/01.atv.21.5.731 PMID: 11348867
9. Bath PMW, Anderton PR, Ankolekar S. Hemostasis and vascular dementia. Arterioscler Thromb Vasc Biol. 2010; 30: 461–463. https://doi.org/10.1161/ATVBAHA.109.200824 PMID: 20167665
10. Quinn TJ, Gallacher J, Deary IJ, Lowe GDO, Fenton C, Stott DJ. Association between circulating hemostatic measures and dementia or cognitive impairment: systematic review and meta-analyzes. J Thromb Haemost. 2011; 9: 1475–1482. https://doi.org/10.1111/j.1538-7836.2011.04403.x PMID: 21676170
11. 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. Arterioscler Thromb Vasc Biol. 2010; 30: 599–604. https://doi.org/10.1161/ATVBAHA.109.197368 PMID: 19965782
12. Tracy RP, Arnold AM, Ettinger W, Fried L, Meilahn E, Savage P. The relationship of fibrinogen and factors VII and VIII to incident cardiovascular disease and death in the elderly: results from the cardiovascular health study. Arterioscler Thromb Vasc Biol. 1999; 19: 1776–1783. https://doi.org/10.1161/01.atv.19.7.1776 PMID: 10397698
13. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al. The Cardiovascular Health Study: design and rationale. Ann Epidemiol. 1991; 1: 263–276. https://doi.org/10.1016/1047-2797(91)90005-w PMID: 1669507
14. Cushman M, Cornell ES, Howard PR, Bovill EG, Tracy RP. Laboratory methods and quality assurance in the Cardiovascular Health Study. Clin Chem. 1995; 41: 264–270. PMID: 7874780
15. O'Leary DH, Polak JF, Wolfson SK Jr, Bond MG, Bommer W, Sheth S, et al. Use of sonography to evaluate carotid atherosclerosis in the elderly. The Cardiovascular Health Study. CHS Collaborative

Research Group. Stroke. 1991; 22: 1155–1163. https://doi.org/10.1161/01.str.22.9.1155 PMID: 1926258

16. Psaty BM, Kuller LH, Bild D, Burke GL, Kittner SJ, Mittelmark M, et al. Methods of assessing prevalent cardiovascular disease in the Cardiovascular Health Study. Ann Epidemiol. 1995; 5: 270–277. https://doi.org/10.1016/1047-2797(94)00092-8 PMID: 8520708

17. Longstreth WT Jr, Arnold AM, Beauchamp NJ Jr, Manolio TA, Lefkowitz D, Jungreis C, et al. Incidence, manifestations, and predictors of worsening white matter on serial cranial magnetic resonance imaging in the elderly: the Cardiovascular Health Study. Stroke. 2005; 36: 56–61. https://doi.org/10.1161/01.STR.0000149625.99732.69 PMID: 15569873

18. Longstreth WT Jr, Manolio TA, Arnold A, Burke GL, Bryan N, Jungreis CA, et al. Clinical correlates of white matter findings on cranial magnetic resonance imaging of 3301 elderly people. The Cardiovascular Health Study. Stroke. 1996; 27: 1274–1282. https://doi.org/10.1161/01.str.27.8.1274 PMID: 8711786

19. Manolio TA, Kronmal RA, Burke GL, Poirier V, O'Leary DH, Gardin JM, et al. Magnetic resonance abnormalities and cardiovascular disease in older adults. The Cardiovascular Health Study. Stroke. 1994; 25: 318–327. https://doi.org/10.1161/01.str.25.2.318 PMID: 8303738

20. Teng EL, Chui HC. The Modified Mini-Mental State (3MS) examination. J Clin Psychiatry. 1987; 48: 314–318. PMID: 3611032

21. Arnold AM, Newman AB, Dermond N, Haan M, Fitzpatrick A. Using Telephone and Informant Assessments to Estimate Missing Modified Mini-Mental State Exam Scores and Rates of Cognitive Decline. Neuroepidemiology. 2009; 33: 55–65. https://doi.org/10.1159/000215830 PMID: 19407461

22. Salthouse TA. The role of memory in the age decline in digit-symbol substitution performance. J Gerontol. 1978; 33: 232–238. https://doi.org/10.1093/geronj/33.2.232 PMID: 637915

23. Elkins JS, O'Meara ES, Longstreth WT Jr, Carlson MC, Manolio TA, Johnston SC. Stroke risk factors and loss of high cognitive function. Neurology. 2004; 63: 793–799. https://doi.org/10.1212/01.wnl.0000137014.36689.7f PMID: 15365125

24. Williams R. Generalized Ordered Logit/ Partial Proportional Odds Models for Ordinal Dependent Variable. The Stata Journal. 2006; 6: 58–82.

25. Williams R. Understanding and interpreting generalized ordered logit models. J Math Sociol. 2016; 40: 7–20.

26. Glymour MM, Weuve J, Berkman LF, Kawachi I, Robins JM. When is baseline adjustment useful in analyses of change? An example with education and cognitive change. Am J Epidemiol. 2005; 162: 267–278. https://doi.org/10.1093/aje/kwi187 PMID: 15987729

27. Warsch JRL, Wright CB. The aging mind: vascular health in normal cognitive aging. J Am Geriatr Soc. 2010; 58 Suppl 2: S319–24. https://doi.org/10.1111/j.1532-5415.2010.02983.x PMID: 21029061

28. Knecht S, Oelschläger C, Duning T, Lohmann H, Albers J, Stehling C, et al. Atrial fibrillation in stroke-free patients is associated with memory impairment and hippocampal atrophy. Eur Heart J. 2008; 29: 2125–2132. https://doi.org/10.1093/eurheartj/ehn341 PMID: 18667399

29. Black SE. Vascular cognitive impairment: epidemiology, subtypes, diagnosis and management. J R Coll Physicians Edinb. 2011; 41: 49–56. https://doi.org/10.4997/JRCPE.2011.121 PMID: 21365068

30. Newman AB, Fitzpatrick AL, Lopez O, Jackson S, Lyketsos C, Jagust W, et al. Dementia and Alzheimer's disease incidence in relationship to cardiovascular disease in the Cardiovascular Health Study cohort. J Am Geriatr Soc. 2005; 53: 1101–1107. https://doi.org/10.1111/j.1532-5415.2005.53360.x PMID: 16108925

31. Walker KA, Hoogeveen RC, Folsom AR, Ballantyne CM, Knopman DS, Windham BG, et al. Midlife systemic inflammatory markers are associated with late-life brain volume: The ARIC study. Neurology. 2017; 89: 2262–2270. https://doi.org/10.1212/WNL.0000000000004688 PMID: 29093073

32. Walker KA, Gottesman RF, Wu A, Knopman DS, Gross AL, Mosley TH Jr, et al. Systemic inflammation during midlife and cognitive change over 20 years: The ARIC Study. Neurology. 2019; 92: e1256–e1267. https://doi.org/10.1212/WNL.0000000000007094 PMID: 30760633

33. Alexander KS, Zakai NA, Gillett S, McClure LA, Wadley V, Unverzagt F, et al. ABO blood type, factor VIII, and incident cognitive impairment in the REGARDS cohort. Neurology. 2014; 83: 1271–1276. https://doi.org/10.1212/WNL.0000000000000844 PMID: 25209581

34. Gillett SR, McClure LA, Callas PW, Thacker EL, Unverzagt FW, Wadley VG, et al. Hemostasis biomarkers and incident cognitive impairment: the REGARDS study. J Thromb Haemost. 2018; 16: 1259–1267. https://doi.org/10.1111/jth.14138 PMID: 29733497

35. Loures C de MG, Duarte RCF, Silva MVF, Cicarini WB, de Souza LC, Caramelli P, et al. Hemostatic Abnormalities in Dementia: A Systematic Review and Meta-Analysis. Semin Thromb Hemost. 2019; 45: 514–522. https://doi.org/10.1055/s-0039-1688444 PMID: 31096308

14. Curriculum Vitae

My curriculum vitae does not appear in the electronic version of my work for reasons of data protection.

15. Complete list of publications

1. Kurth T, **Rohmann JL**, Shapiro RE. Migraine and risk of cardiovascular disease. BMJ. 2018 Jan 31;360:k275. doi: 10.1136/bmj.k275. PMID: 29386182

2. Siegerink B, **Rohmann JL**. Impact of your results: Beyond the relative risk. Res Pract Thromb Haemost. 2018 Sep 30;2(4):653-657. doi: 10.1002/rth2.12148. PMID: 30349882

3. **Rohmann JL**, de Haan HG, Algra A, Vossen CY, Rosendaal FR, Siegerink B. Genetic determinants of activity and antigen levels of contact system factors. J Thromb Haemost. 2019 Jan;17(1):157-168. doi: 10.1111/jth.14307. PMID: 30288888

4. Liefting M, **Rohmann JL**, Le Lann C, Ellers J. What are the costs of learning? Modest trade-offs and constitutive costs do not set the price of fast associative learning ability in a parasitoid wasp. Anim Cogn. 2019 Sep;22(5):851-861. doi: 10.1007/s10071-019-01281-2. PMID: 31222547

5. Nyberg J, Gustavsson S, Linde M, Åberg ND, **Rohmann JL**, Åberg M, Kurth T, Waern M, Kuhn GH. Cardiovascular fitness and risk of migraine: a large, prospective population-based study of Swedish young adult men. BMJ Open. 2019 Aug 30;9(8):e029147. doi: 10.1136/bmjopen-2019-029147. PMID: 31473616

6. Sperber PS, Siegerink B, Huo S, **Rohmann JL**, Piper SK, Prüss H, Heuschmann PU, Endres M, Liman TG. Serum Anti-NMDA (N-Methyl-D-Aspartate)-Receptor Antibodies and Long-Term Clinical Outcome After Stroke (PROSCIS-B). Stroke. 2019 Nov;50(11):3213-3219. doi: 10.1161/STROKEAHA.119.026100. PMID: 3152612

7. Piccininni M, **Rohmann JL**, Logroscino G, Kurth T. Blockchain-based innovations for population-based registries for rare neurodegenerative diseases. Frontiers in Blockchain. 2020;3:20. doi: 10.3389/fbloc.2020.00020

8. Padberg I, Schneider A, **Rohmann JL**, Kelley SW, Grittner U, Siegerink B. Impact of COPD and anemia on motor and cognitive performance in the general older population: results from the English longitudinal study of ageing. Respir Res. 2020 Feb 3;21(1):40. doi: 10.1186/s12931-020-1305-6. PMID: 32013976

9. **Rohmann JL**, Rist PM, Buring JE, Kurth T. Migraine, headache, and mortality in women: a cohort study. J Headache Pain. 2020 Mar 17;21(1):27. doi: 10.1186/s10194-020-01091-9. PMID: 32183686

10. Piccininni M*, **Rohmann JL***, Huscher D, Mielke N, Ebert N, Logroscino G, Schäffner E, Kurth T. Performance of risk prediction scores for cardiovascular mortality in older persons: External validation of the SCORE OP and appraisal. PLoS One. 2020 Apr 9;15(4):e0231097. doi: 10.1371/journal.pone.0231097. PMID: 32271825 (*shared first authorship)

11. Piccininni M, **Rohmann JL**, Foresti L, Lurani C, Kurth T. Use of all cause mortality to quantify the consequences of covid-19 in Nembro, Lombardy: descriptive study. BMJ. 2020 May 14;369:m1835. doi: 10.1136/bmj.m1835. PMID: 32409488

12. Kurth T, Piccininni M, Loder EW, **Rohmann JL**. A parallel pandemic: The crush of COVID 19 publications tests the capacity of scientific publishing. The BMJ Opinion. 2020 May 26. https://blogs.bmj.com/bmj/2020/05/26/a-parallel-pandemic-the-crush-of-covid-19-publications-tests-the-capacity-of-scientific-publishing/

13. Piccininni M, Konigorski S, **Rohmann JL**, Kurth T. Directed acyclic graphs and causal thinking in clinical risk prediction modeling. BMC Med Res Methodol. 2020 Jul 2;20(1):179. doi: 10.1186/s12874-020-01058-z. PMID: 32615926

14. Meier K, Glatz T, Guijt MC, Piccininni M, van der Meulen M, Atmar K, Jolink AC, Kurth T, **Rohmann JL**, Zamanipoor Najafabadi AH; COVID-19 Survey Study group. Public perspectives on protective measures during the COVID-19 pandemic in the Netherlands, Germany and Italy: A survey study. PLoS One. 2020 Aug 5;15(8):e0236917. doi: 10.1371/journal.pone.0236917. PMID: 32756573.

15. **Rohmann JL**, Huo S, Sperber PS, Piper SK, Rosendaal FR, Heuschmann PU, Endres M, Liman TG, Siegerink B. Coagulation factor XII, XI, and VIII activity levels and secondary events after first ischemic stroke. J Thromb Haemost. 2020 Dec;18(12):3316-3324. doi: 10.1111/jth.15092. PMID: 32935900

16. Al-Hassany L, Haas J, Piccininni M, Kurth T, Maassen Van Den Brink A, **Rohmann JL**. Giving Researchers a Headache - Sex and Gender Differences in Migraine. Front Neurol. 2020 Oct 22;11:549038. doi: 10.3389/fneur.2020.549038. PMID: 33192977

17. Huo S, Kränkel N, Nave AH, Sperber PS, **Rohmann JL**, Piper SK, Heuschmann PU, Landmesser U, Endres M, Siegerink B, Liman TG. Endothelial and Leukocyte-Derived Microvesicles and Cardiovascular Risk After Stroke: PROSCIS-B. Neurology. 2021 Feb 9;96(6):e937-e946. doi: 10.1212/WNL.0000000000011223. PMID: 33184230

18. **Rohmann JL**, Longstreth WT Jr, Cushman M, Fitzpatrick AL, Heckbert SR, Rice K, Rosendaal FR, Sitlani CM, Psaty BM, Siegerink B. Coagulation factor VIII, white matter hyperintensities and cognitive function: Results from the Cardiovascular Health Study. PLoS One. 2020 Nov 16;15(11):e0242062. doi: 10.1371/journal.pone.0242062. PMID: 33196677

19. Erdur H, Siegerink B, Leithner C, Franke C, Lorenz-Meyer I, Theen S, Angermaier A, Kinze S, Weber JE, **Rohmann JL**, Scheitz JF, Nolte CH, Endres M, Audebert HJ. Stroke Admissions, Stroke Severity, and Treatment Rates in Urban and Rural Areas During the COVID-19 Pandemic. Front Neurol. 2021 Jan 6;11:607193. doi: 10.3389/fneur.2020.607193. PMID: 33488501

20. Ebinger M, Siegerink B, Kunz A, Wendt M, Weber JE, Schwabauer E, Geisler F, Freitag E, Lange J, Behrens J, Erdur H, Ganeshan R, Liman T, Scheitz JF, Schlemm L, Harmel P, Zieschang K, Lorenz-Meyer I, Napierkowski I,

Waldschmidt C, Nolte CH, Grittner U, Wiener E, Bohner G, Nabavi DG, Schmehl I, Ekkernkamp A, Jungehulsing GJ, Mackert BM, Hartmann A, **Rohmann JL**, Endres M, Audebert HJ; Berlin_PRehospital Or Usual Delivery in stroke care (B_PROUD) study group. Association Between Dispatch of Mobile Stroke Units and Functional Outcomes Among Patients With Acute Ischemic Stroke in Berlin. JAMA. 2021 Feb 2;325(5):454-466. doi: 10.1001/jama.2020.26345. PMID: 33528537

21. Kruppa J, **Rohmann JL**, Herrman C, Sieg M, Rubarth K, Piper S. What statistics instructors need to know about concept acquisition to make statistics stick, Journal of University Teaching & Learning Practice. 2021 Feb 25; 18(2). https://ro.uow.edu.au/jutlp/vol18/iss2/02

16. Acknowledgments

While working on my doctoral project, I experienced periods of growth sprinkled with setbacks. Reflecting upon the set of underlying causes that ultimately led to this final draft reaching a form worthy of submission, I want to express my utmost gratitude to the driving forces contributing to my success and helping me cross the finish line. My list of acknowledgments is by no means exhaustive. I want to extend my deepest, heartfelt thanks to everyone who contributed even a small slice to my dissertation's causal pie.

I will never forget the moment Bob Siegerink answered my email expressing my interest to work with him and his group at the Center for Stroke Research. As my daily supervisor, he introduced me to the world of coagulation factors and challenged me to be a better problem solver. Bob taught me how to think like an epidemiologist, ask better questions, and stay spontaneous in an academic setting. Perhaps most importantly, Bob understood that my own path in academia was not a traditional or linear one. During my PhD journey, he kept an open ear and open office door. His abilities as an educator, both inside and outside the classroom, inspired me to be a better teacher and leader. I am forever grateful for his mentorship, his persistence in helping me hone these important career skills. It took some time, but in the end, he also managed to convince me that every food is better deep-fried. He may be a short guy, but he left me awfully big shoes to fill at the Center for Stroke Research. I hope I can make him proud.

Andreas Meisel did not know me well when he agreed to serve as my primary dissertation supervisor, but looking back now, am I sure glad he said "yes". I am deeply grateful for Andreas' willingness to explore and embrace new ideas like DAGs and the importance of asking good casual questions in stroke research. I always felt energized after talking with Andreas, and thank him for always believing in me. Andreas' support throughout my time as a doctoral student was indispensable, and will continue to be instrumental for me as I embark on this new journey as junior group leader in the next chapter in my academic story.

Tobias Kurth will forever remember me as the student who walked into his not-yet-unpacked office just days after he relocated to Berlin. He listened patiently as I laid out a comprehensive, step-by-step plan to modernize the epidemiology scene in Berlin, and then asked him to be my third PhD supervisor. Lucky for me, he was very receptive on all counts and asked me to stay to chat over coffee while he shared his ideas on the topics. Thereafter, I quickly learned Tobias never runs out of energy to improve existing structures and create new things, which has been so inspiring to me. I have yet to meet a stronger advocate for early career researchers. Somehow, Tobias always finds a way to provide unconditional support to his students even while juggling a million other responsibilities. After a few years of observation, I still haven't quite worked out how he

does it. Thankfully, our collaborations will continue, and I look forward to the opportunity to collect further observations and learn more from this great role model.

Furthermore, I want to thank each of the brilliant methodologists who served as rich sources of inspiration to me and who donated their time to present at the Berlin Epidemiological Methods Colloquium. My amazing Society of Epidemiological Research mentorship group (Robert Platt, Eileen Ming, Erin Bowles and Vanessa Miller) was also an indispensable resource and provided monthly confidence boosts. I thank these brilliant, talented individuals for their openness and willingness to share their experiences with me. On the whole, I feel so fortunate to have been surrounded by so many great role models in science during my time as a researcher-in-training.

Regrettably, I did not have space in the Conclusion section to address it, but one of my biggest takeaways from my doctoral studies is that other people not only make research better but also way more fun. In this vein (pun, of course, intended), hats off to my CSB 2.0 & 3.0 colleagues who put up with me, even after pressuring them to join my lunchtime strength training, yoga and line dancing sessions. An equally large thank you to my IPH CONVERGE colleagues; our weekly online Gartic Phone games were so grounding during the toughest phases of the pandemic and gave me the much-needed boost to keep our research projects moving along.

And in regard to my “other” career, I can only express my immense gratitude to the Berlin-based ComedyShorts group full of talented, wonderful improvisers for welcoming me with open arms into their community. Time spent with these folks both on-and off-stage has taught me so many important, transferrable career and life skills. Thank you for all the pick-me-ups sprinkled with good old-fashioned “yes, and”s and for always reminding me to find my breath, even in between big belly laughs.

Next, I want to thank my family and friends on the other side of the pond for their continued support during this whole process. First, many thanks to my parents, Gayle and Jeff, for listening to weekly digests of my cumulative frustrations with great patience and advice, for never questioning why I wanted to take German instead of Spanish in the first place, and for single-handedly boosting my Altmetric score by sharing my articles like crazy on social media. Second, to Kelsey, for the many helpful sister pep talks and for mailing me the comfiest writing clothes ever. Third, to Karina, Andrew, Jill, and Carolyn for so many years of great friendship, for open ears, and for their willingness to install multiple messaging apps to stay in touch with me - that’s real love.

And where would I be without my Berlin “family”? In the order I met them: Guido, Ulla, Lydia, Robertson, Mary, Toni, Marco, Ilaria and Toivo. This city feels home to me thanks to each of you. Thank you for being such grounding and stabilizing forces for me during my PhD studies. I will forever value our stimulating discussions interlaced with

welcomed distractions. You have each held my hand on this PhD rollercoaster - now we can laugh and look back fondly at the silly pictures! Ready for the next one?

Marco and Ilaria deserve an extra special acknowledgement for their indispensable encouragement during the writing “crunch-time”. Marco wanted to learn epidemiology from me, but in the end, he taught me to rethink everything I thought I knew. Marco, thinking hard about big questions with you gives me so much energy and reminds me why I wanted to become a scientist in the first place. I am eternally grateful for the way you have consistently pushed me to stay curious, especially during this formative phase of my life. I also want to thank Ilaria, an incredibly good listener, for her deep friendship, deep-fried Italian delicacies, and deep-seated support. My appreciation for the way you have opened your heart and your home to me is without bounds.

And to Guido, my life partner, my “reference group”: I am forever grateful to and for you. Thanks in no small part to your unconditional love, patience, flexibility, and understanding, we have successfully navigated yet another big challenge together. The printing of this doctoral thesis represents a celebratory conclusion of this journey, along which you have believed in me and cheered me on as my self-proclaimed “biggest fan” every single step of the way. This milestone would not have been possible without your tireless support and dedication in every imaginable form.

Lastly, I want to thank the participants in the three studies from which I obtained data. Particularly for stroke patients, study participation can be incredibly taxing. I am incredibly grateful for their contributions to research and to my project, in particular. I can only hope that my past, current, and future research might contribute to a better understanding of this disease and serve to help alleviate suffering.