Prestroke physical activity and outcomes after intracerebral haemorrhage in comparison to ischaemic stroke: protocol for a matched cohort study (part of PAPSIGOT)

Adam Viktorisson, Dongni Buvarp, Katharina S Sunnerhagen

ABSTRACT

Introduction Piling evidence suggests that a higher level of prestroke physical activity can decrease stroke severity, and reduce the risk of poststroke mortality. However, prior studies have only included ischaemic stroke cases, or a majority of such. We aim to investigate how premorbid physical activity influences admission stroke severity and poststroke mortality in patients with intracerebral haemorrhage, compared with ischaemic stroke. A prespecified analysis plan counteract some inherent biases in observational studies, and promotes transparency.

Methods and analysis This is a statistical analysis protocol for a matched cohort study, including all adult patients with intracerebral haemorrhage, and matched ischaemic stroke controls, treated at Sahlgrenska University Hospital in Sweden between 1 November 2014 and 30 June 2019. All patients have been identified in the Väststroke register, and the data file has been sent for merging with national registries. The follow-up of time for survival will be approximately 2–7 years. The sample size calculation indicates that a minimum of 628 patients with intracerebral haemorrhage is needed for power of 80% at an alpha level of 0.01. Multiple imputation by chained equations will be used to handle missing data. The entire cohort of patients with intracerebral haemorrhage will be matched with consecutive ischaemic stroke controls (1:3 ratio) using nearest neighbour propensity score matching. The association between prestroke physical activity and admission stroke severity will be evaluated using multivariable ordinal regression models, and risk for all-cause mortality will be analysed using multivariable Cox proportional-hazards models. Potential confounders include age, ethnicity, income, educational level, comorbidity, medical treatments, alcohol-related disorders, drug abuse and smoking.

Ethics Data collection for the Physical Activity Pre-Stroke in GOTHenbug project was approved by the Regional Ethical Board on 4 May 2016. An additional application was approved by the National Ethical Review Authority on 7 July 2021.

INTRODUCTION Observational studies are of great value in cases when a randomised controlled trial would be unethical, impractical or untimely. Observational studies are, however, subjected to several types of selection and reporting bias, which can lead to a lack in reproducibility. The reliability of observational data can be enhanced by emulating principal design aspects of randomised trials. A prespecified eligibility criteria, a priori sample size estimation, definition of exposure and outcome measures, a standardised follow-up period, and a statistical analysis plan may counterbalance some inherent limitations of the observational study design and promote transparency.

Physical activity is an example of exposure that is hard to study in a randomised, controlled setting. The relationship between premorbid physical activity and admission stroke severity, as well as poststroke mortality has been analysed in several observational studies, with varying results. The majority of prior studies have found that prestroke physical activity reduce stroke severity, cardiovascular mortality, in-hospital mortality, and all-cause mortality. However, some studies...
found no associations between physical activity and stroke severity or poststroke mortality after covariate adjustment. These studies differ in data collection strategies, sample sizes, physical activity assessments and statistical analyses. Moreover, none of the prior studies have investigated the association between prestroke physical activity and stroke severity or mortality in the subgroup of patients with intracerebral haemorrhage (ICH).

Here, we present a statistical analysis plan, and give the rationale for an observational, matched cohort study part of the Physical Activity Pre-Stroke In GOTHenburg (PAPSIGOT) project. We also provide a narrative discussion for potential confounding factors to the effect of prestroke physical activity on outcomes after ICH. This study protocol follows the guidelines for reporting of statistical analysis plans in clinical trials, published in the Journal of the American Medical Association.

Research objectives
(1) To investigate the relationship between prestroke physical activity and admission stroke severity in patients with ICH, compared with patients with ischaemic stroke.
(2) To investigate how prestroke physical activity influences the risk of mortality after ICH, in comparison to ischaemic stroke.

METHODS
This is a statistical analysis protocol for a matched cohort study. Prestroke physical activity is the main independent variable. Admission stroke severity and all-cause mortality are the outcome variables of interest. All adult patients with ICH and ischaemic stroke, treated at Sahlgrenska University Hospital between 1 November 2014 and 30 June 2019 have been identified in the Väststroke register. The comprehensive stroke units at Sahlgrenska University Hospital has a catchment area of 700 000 people, and patients who had a stroke in nearby regions requiring thrombectomy or neurosurgical monitoring are transported to this hospital. All persons with stroke in Sweden have equal access to tax-funded healthcare services, which ensures the representativeness of this sample.

Data collection and management
Prestroke physical activity and acute stroke severity assessments were collected from the Väststroke register, and cross-references with medical records to minimise missing data. Physical activity was assessed using the four-level Saltin-Grimby Physical Activity Level Scale (SGPALS). Acute stroke severity was assessed using the National Institutes of Health Stroke Scale (NIHSS). The data collected from the Väststroke register has been sent for merging with national registries. Information regarding prestroke conditions and treatments will be collected from the Swedish National Stroke Registry (Riksstroke). Country of birth will be collected from the Swedish Multi-Generation register, and socioeconomic variables will be collected from the Longitudinal Integration Database for Health Insurance and Labour Market Studies held by Statistics Sweden. Comorbidities will be collected from the National Patient Registry, and mortality rates will be collected from the Swedish Cause of Death registry, held by the National Board of Health and Welfare. All-cause mortality and cerebrovascular mortality will be recorded for all patients until October 2021, which allow for a minimum follow-up of 2 years.

The data management started on 17 May 2021 after submission of this protocol. The final data was received on 19 October 2021. Patients without a Swedish personal identification number, and patients with non-specified aetiology will be excluded. The final sample is will consist of 763 patients with ICH.

Continuous variables will be either mean-centred or dichotomised when appropriate. For ordinal variables, categories with limited observations will be merged on the base of clinical reasoning. All variables and data sources are listed in table 1. The categories of SGPALS will be collapsed into inactive (SGPALS level 1) and physically active (SGPALS levels 2–4) for all analyses. NIHSS scores will be categorised based on distributional assumptions. The weighted Charlson Comorbidity Index will be calculated, and applied as a four-level ordinal variable using the ICD-10 diagnoses specified in table 1. The 1-year mortality risk described by Charlson et al was 12% for patients with 0 points, 26% for 1–2 points, 52% for 3–4 points and 85% for ≥5 points.

Sample size calculation
The sample size was estimated using the formula for ordered categorical data suggested by John Whitehead, stated below. The primary outcome measure for the power calculation is the difference in stroke severity on a three-level scale between inactive and physically active patients. Control probabilities were collected from previous research (table 2). We expect a fraction of 60% physically inactive patients (SGPALS level 1). The sample size estimate was calculated to detect an OR of 2.0 at an alpha level of 0.01 with a power of 80. Based on the power calculation, we need a minimum of 628 patients with ICH, which is within the expected number of cases.

$n = 6 \left( \frac{-\frac{2}{\pi} + Z_\alpha + Z_\beta}{\log (OR)^2 \left(1 - \sum_{i=1}^{3} \left(\pi_{iE} + \pi_{IC} \right) / 2 \right)^2} \right)$

Statistical analyses
Data will be analysed using R, V.4.0.2 (R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria.), and SPSS Statistics for Windows, V27.0 (released on 2018; IBM). All statistical testing will be interpreted at a two-tailed significance level of 0.01 to adjust for multiple comparisons.

Missing data
Missing observations for each variable will be explored, and variables with more than 30% will be excluded from the data set. Individuals with missing values on more than
20% of the remaining variables will be excluded from analyses. The number of missing observations in the total data set is expected to be below 3%. Multiple imputation by chained equations will be applied to handle missing data. We will use the predictive mean matching method, originally proposed by Rubin, implemented for multiple imputation. The method generates a separate model for each variable with missing data, and creates multiple imputations for each value. The predictive mean algorithm preserves the original data distribution and allows

---

**Table 1** Registers and variables included in the study

<table>
<thead>
<tr>
<th>The väststroke register</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Mean centred or dichotomised.</td>
</tr>
<tr>
<td>Sex</td>
<td>Female/male.</td>
</tr>
<tr>
<td>Physical activity assessment (SGPALS)</td>
<td>Mostly sedentary (level 1); light physical activity such as walking ≥4 hours/week (level 2); moderate physical activity such as running ≥2 hours/week (level 3); hard physical training for competition sports several times per week (level 4).</td>
</tr>
<tr>
<td>Stroke severity assessment (NIHSS)</td>
<td>Mild stroke (NIHSS 0-5), moderate stroke (NIHSS 6-14), and severe stroke (NIHSS &gt;14).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The Riksstroke register</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior stroke</td>
<td>Yes/no.</td>
</tr>
<tr>
<td>Prior transient ischaemic attack</td>
<td>Yes/no.</td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>Yes/no.</td>
</tr>
<tr>
<td>Diabetes</td>
<td>Yes/no.</td>
</tr>
<tr>
<td>Regular smoking the year before stroke</td>
<td>Yes/no.</td>
</tr>
<tr>
<td>Antihypertensive treatment</td>
<td>Yes/no.</td>
</tr>
<tr>
<td>Statin treatment</td>
<td>Yes/no.</td>
</tr>
<tr>
<td>Antiplatelet treatment</td>
<td>Yes/no.</td>
</tr>
<tr>
<td>Anticoagulant treatment</td>
<td>Yes/no.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The LISA register</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Income</td>
<td>Mean centred or dichotomised.</td>
</tr>
<tr>
<td>Educational level</td>
<td>Primary school (&lt;10 years), secondary school (10–12 years) and postsecondary or university education (&gt;12 years).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The NPR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Charlson Comorbidity Index: myocardial infarction, heart failure, peripheral vascular disease, dementia, pulmonary disease, connective tissue disease, peptic ulcer disease, liver disease, diabetes, Hemiplegia or paraplegia, renal disease, cancer, HIV/AIDS</td>
<td>Low (0 points), moderate (1–2 points), high (3–4 points) and very high (≥5 points) risk. ICD-10 codes: I21, I22, I25.2, I42, I11.0, I13.0, I13.2, I50, I25.5, I65, I70-174, I77, I73.9, I79.0, R02, F00-F03, F05.1, J40-J47, J60-J67, M05.0-M05.3, M05.8-M06.0, M06.3, M06.9, M32, M33.2, M34, M35.3, K25-K28, K70.2, K70.3, K71.7, K73, k74.0, K74.2-K74.6, K72.9, K76.6, K76.7, K72.1, E10.1, E10.5, E10.9, E11.1, E11.5, E11.9, E13.1, E13.5, E13.9, E14.1, E14.5, E14.9, E10.2-E10.4, E11.2-E11.4, E13.2-E13.4, E14.2-E14.4, G80.0, G80.2, G81, G82, N01-N08, N11, N14-N19, N25, C00-C76, C90, C97, C91-C95, C81-C86, C88, C77-C80, B20-B24</td>
</tr>
<tr>
<td>Depression</td>
<td>Yes/no. ICD-10 codes: F32-F33</td>
</tr>
<tr>
<td>Psychotic diseases</td>
<td>Yes/no. ICD-10 codes: F20-F29</td>
</tr>
<tr>
<td>Alcohol related disorders</td>
<td>Yes/no. ICD-10 codes: F10, K70.1-K70.4, K70.9, K74.0</td>
</tr>
<tr>
<td>Drug abuse</td>
<td>Yes/no. ICD-10 codes: F11, T40.0-T40.4, F12, T40.7, F14, T40.5, F15, T43.6, F16, T40.6, F19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The Swedish cause of death register</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>All-cause mortality</td>
<td>–</td>
</tr>
<tr>
<td>Cerebrovascular mortality</td>
<td>ICD-10 codes: I60-I69</td>
</tr>
</tbody>
</table>

ICD-10, International Classification of Diseases, Tenth Revision; LISA, Longitudinal Integration Database for Health Insurance and Labour Market Studies; NIHSS, National Institutes of Health Stroke Scale; NPR, National Patient Registry; SGPALS, Saltin-Grimby Physical Activity Level Scale.
for imputation of continuous, binary and categorical data. The imputation will be performed separately for patients with ICH and ischaemic stroke, and ordered from the variable with least missing values (monotone). Only variables with a minimum correlation of 0.1 and at least 50% data will be used for prediction. The number of multiple imputations will be set to m=20.

Propensity scores and matching
To achieve equipoise between patients with ICH and ischaemic stroke, propensity score matching will be applied. We will include the entire cohort of patients with ICH and match them with consecutive ischaemic stroke patients, in a 1:3 ratio. The propensity score will be based on sex and potential confounding variables. The rational for the selection of confounders is given below. Here, we assume that confounders for the effect of physical activity on stroke severity will be analogous for ICH and ischaemic stroke. We will conduct generalised boosted regression modelling to calculate the propensity scores. We intend to estimate the average effect of being physically active, that is, the average treatment effect for the treated will be applied. The number of trees, shrinkage and interaction depth will be tuned by using 10-fold cross-validation. We will use the stopping rule es.mean, which uses the absolute standardised bias as balance metric and the mean of the covariate balance metrics to summarise across variables.

The propensity score matching will be conducted using the nearest neighbour matching method, with a logistic distance measure and without replacement.

Regression models
We will conduct separate models for haemorrhagic and ischaemic stroke cases, respectively. All models will be adjusted for potential confounders. The association between prestroke physical activity and admission stroke severity will be evaluated using multivariable ordinal regression models. For the ordinal regression, NHISS scores will be trichotomised. A subgroup analysis will be performed for patients with first-time ICH. Prior to the analyses, absence of multicollinearity, and the proportional odds assumption will be checked. Diagnostics will be performed using surrogate residuals.

The association between prestroke physical activity and all-cause mortality will be evaluated using multivariable Cox proportional-hazards models. Additional models will be fitted adjusted for admission stroke severity. Model diagnostics will be applied to evaluate violations of the proportional hazards assumption, random censoring and functional forms of continuous covariates. Furthermore, outlying and influential observations will be checked for all models.

Confounding variables
There is limited evidence for the relationship between physical activity and previously known determinants of outcomes following ICH. However, several factors may confound or mediate the potential impact of physical activity. The covariates included as potential confounders in this study are based on theoretical reasoning, and current evidence. The relationship between prestroke physical activity and the main groups of confounders is portrayed in figure 1.

Degenerative vascular changes are frequent in the ageing population. In addition to being a risk factor for all stroke, higher age has been identified as an important predictor of disability, and decreased long-term survival in ICH. A person’s age also influences the probability of being physically active, and should therefore be considered a confounding factor. Sex, however, did not significantly influence the severity of haemorrhagic stroke in a meta-analysis of eight population-based stroke incidence studies.

Socioeconomic status-related differences influence haemorrhagic stroke severity, but could partly be explained by differences in comorbidities, disparity in health and care services, ethnicity, stress, job strain and life style factors. Income, education and living alone have been associated with mortality following the acute phase of stroke. Low income was found to increase the absolute risk of death at 3 months after ICH, and the effect

Table 2  Frequencies of stroke severity using the National Institutes of Health Stroke Scale (NIHSS), in physically inactive and active patients with stroke (Reinholdsson et al)

<table>
<thead>
<tr>
<th>Stroke Severity</th>
<th>Physical Activity</th>
<th>Cumulative OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mild stroke</td>
<td>πC  73.6%</td>
<td>πE  86.5%</td>
</tr>
<tr>
<td>Moderate stroke</td>
<td>πC  20.4%</td>
<td>πE  10.8%</td>
</tr>
<tr>
<td>Severe stroke</td>
<td>πC  6.0%</td>
<td>πE  2.7%</td>
</tr>
</tbody>
</table>

Figure 1  Directed acyclic graph for confounders to the effect of physical activity on outcomes after intracerebral haemorrhage (stroke severity and poststroke mortality). Solid lines indicate the relationships between covariates, and the dotted line indicate the causal effect of physical activity.
was largely mediated by initial stroke severity.\textsuperscript{38} In adulthood, socioeconomic variables are associated with the propensity of being physically active.\textsuperscript{39} Among young people, a recent study found that haemorrhagic stroke survivors of African and Hispanic ethnicity have better outcomes compared with those of Caucasian descent.\textsuperscript{40} In the current study, country of birth will be used as a surrogate measure for ethnicity.

Active cigarette smoking has been associated with increased stroke severity,\textsuperscript{41, 42} and premorbid alcohol consumption has been identified as independent determinant of functional status and mortality after ICH.\textsuperscript{43} Cigarette smoking and physical activity are intuitively incongruent behaviours, and both positive and negative associations has been described.\textsuperscript{44} Likewise, alcohol consumption has in some studies been found to be a contributor to sedentary behaviour.\textsuperscript{45}

Patients with severe comorbidity may experience worse outcomes following ICH, although this has not been thoroughly investigated. One study found that dialysis patients with a medical history of stroke, diabetes and malignancy had a higher risk of mortality following ICH.\textsuperscript{46} In another study, global comorbidity evaluated by the Charlson Comorbidity Index was associated with increased risk of 90-day mortality, but not with the length of hospital stay, among immobile patients below 50 years of age.\textsuperscript{47} Severe disease may limit a person’s ability to be physically active, and thereby confound the effect of prestroke physical activity on outcomes following ICH.

Psychiatric comorbidity has been associated with stroke severity and functional outcomes in a population with both haemorrhagic and ischaemic stroke cases.\textsuperscript{48} Physical activity decrease the risk of depression, and a majority of people with depression report sedentary behaviour.\textsuperscript{49, 50} Similarly, patients with psychotic diseases, such as Schizophrenia, generally have a very low level of physical activity.\textsuperscript{51}

Hyperglycaemia and diabetes have been related to worse outcomes after ICH. Diabetes and high blood glucose in nondiabetic patients have been associated with cerebral complications, and are independent predictors of 30-day mortality, and 3-month mortality following ICH.\textsuperscript{52} Hyperglycaemia has also been found to decrease short term survival after ICH.\textsuperscript{53, 54} The presence of hyperglycaemia in severe stroke may, however, be a result of an extensive cerebral lesion, causing a physiologic increment of cortisol and catecholamines. Regular physical activity promotes blood glucose control in diabetic patients.\textsuperscript{55} Still, a large survey of the US population found that diabetic adults generally engaged in less physical activities compared with non-diabetic adults.

Atrial fibrillation is present in about one third of patients 3 months after ICH,\textsuperscript{56} and premorbid atrial fibrillation has been found to predict a higher risk of death 3 months after ictus.\textsuperscript{57} However, the risk increment seemed to be mediated by the higher prevalence of anti-thrombotic treatment in the group with atrial fibrillation. There is a complicated relationship between physical activity and the risk of atrial fibrillation. While a moderate amount of physical activity is protective for atrial fibrillation, vigorous physical activity increase the risk of atrial fibrillation in men.\textsuperscript{58}

Coagulopathy caused by anticoagulants increase the risk of haematoma expansion, mortality and disability after ICH. Warfarin usage has been associated with increased risk of in-hospital haematoma expansion,\textsuperscript{59} and a doubled risk of mortality at 3 months after ICH.\textsuperscript{60} In an international multicentre pooled analysis, haemorrhagic volume on admission, haematoma expansion, functional status at discharge and 90-day mortality were similar in patients treated with warfarin and non-vitamin K antagonist oral anticoagulation (NOAK), meaning that the risk increment is substantial for treatment with NOAK as well.\textsuperscript{61} The propensity of being physically active may be connected to anticoagulant treatment through differences in the prevalence of atrial fibrillation and previous strokes.

The use of lipid-lowering drugs has been controversial in patients with ICH as hypocholesterolaemia has been identified as a risk factor for ICH and haematoma expansion.\textsuperscript{62, 63} Low levels of serum lipids were also independently associated with a poor 3 months prognosis after ICH in women.\textsuperscript{64} The use of statins has, however, only been correlated to an increased risk of ICH in patients with a prior haemorrhagic or lacunar ischaemic stroke. On the other hand, several observational studies have reported better outcomes after ICH among patients medicated with statins. Premorbid statin usage was associated with a better long-term outcome at 12 months after ICH.\textsuperscript{65} Another study found that statins reduced mortality and disability without any negative effect on haematoma growth, but the association disappeared when adjusted for potential indication bias.\textsuperscript{66} Regular physical activity positively alter the levels of blood lipids, and may thereby decrease the probability of receiving statin treatment.\textsuperscript{67}

A previous stroke or transient ischaemic attack (TIA) can potentially precondition the brain to endure ischaemia, and having gone through a TIA has been associated with lower ischaemic stroke severity.\textsuperscript{3} Although ischaemic tolerance may play an important role in haemorrhagic stroke severity, the influence of a preceding stroke or TIA has not yet been evaluated. Patients diagnosed with stroke or TIA are, however, likely to receive recommendations regarding physical activity in accordance poststroke guidelines.

In conclusion, the following potential confounders will be collected, and adjusted for: age, ethnicity, income, educational level, comorbidity, medical treatments, alcohol related disorders, drug abuse and smoking. The comorbidities accounted for are prior TIA, prior stroke, diabetes, psychiatric disorders and Charlson Comorbidity Index. The medical treatments accounted for are anticoagulant treatment, antihypertensive treatment, antiplatelet treatment and statin treatment.
Ethics
Data collection for the PAPSGOT project was approved by the Regional Ethical Board of Gothenburg on 4 May 2016 (346-16). An additional application was approved by the Swedish Ethical Review Authority on 7 July 2021 (2021-03324). Research within the framework of quality registers does not require written informed consent from included patients. The Personal Data Act (Swedish law #1998:204, issued 29 April 1998) allows data from medical records to be collected for quality control without receiving a written informed consent.

Contributors
AV has taken part in the planning and design of this study. AV will collect the data, take part in data analysis and write the manuscript. DB has taken part in the design of this study. DB will take part in the data analysis and revise the manuscript. KSS has led the planning and design of this study. KSS will assist in the interpretation of data and revise the manuscript.

Funding
The study was funded by the Swedish Research Council (VR2017-00946), the Swedish Heart and Lung Foundation, the Swedish Brain Foundation, the Swedish Stroke Association, Promobilia, and the Swedish state under an agreement between the Swedish government and the county councils: the ALF agreement (ALFGBG-71980).

Competing interests
None declared.

Patient and public involvement
Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication
Not applicable.

Provenance and peer review
Not commissioned; externally peer reviewed.

Open access
This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: https://creativecommons.org/licenses/by/4.0/.

ORCID iDs
Adam Viktorsson http://orcid.org/0000-0002-4659-2406
Katharina S Sunnerhagen http://orcid.org/0000-0002-5940-4400

REFERENCES
2 Thomas L, Peterson ED. The value of statistical analysis plans in observational research: defining high-quality research from the start. *JAMA* 2012;308:773–4.
Race/ethnicity influences and short term survival in primary intracerebral hemorrhage:


