Plan Overview

A Data Management Plan created using DMPonline

Title: Per-protocol P2YR

Creator: Hans-Georg Kuhn

Principal Investigator: Hans-Georg Kuhn

Data Manager: Hans-Georg Kuhn

Contributor: Madeleine Hinwood

Affiliation: Göteborgs universitet (Gothenburg University)

Funder: Swedish Research Council

Template: DCC Template

ORCID iD: 0000-0003-4247-5613

Project abstract:

Recent animal studies suggest that P2Y12 receptor (P2Y12R) inhibitors, antiplatelet drugs commonly prescribed after stroke, can suppress the activity of the brain's innate immune system, and potentially affect brain repair processes. Results from preclinical (animal) studies have shown this to be a very real possibility, with animals administered these drugs after stroke or other event which compromises the blood-brain barrier exhibiting worse cognitive outcomes compared with animals not administered these drugs. We conducted an observational target trial emulation study, which was an analogue if an intention-to-treat analysis testing this hypothesis using Swedish National Health and Medical Registers. This study (currently under review and available as a preprint here) demonstrated cognitive safety in terms of incidence of diagnosis with dementia or mild cognitive disorder. However, the effects of sustained treatment adherence on cognitive outcomes remain unclear. This study will evaluate whether maintaining P2Y12R inhibitor treatment for at least one year affects subsequent dementia risk, and will examine whether this relationship is mediated through prevention of recurrent cardiovascular events.

ID: 169361

Start date: 01-07-2025

End date: 31-12-2026

Last modified: 27-05-2025

as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Per-protocol P2YR

Data Collection

What data will you collect or create?

- Pseudonymised, longitudinal health-register data for individuals with incident ischemic stroke (n ≈ 109 855; 1.8 M rows) from Socialstyrelsen, Statistics Sweden and national quality registers.
- Prescription dispensation dates for $P2Y_{12}$ receptor inhibitors.
- Dates and codes for recurrent cardiovascular events (re-stroke, acute coronary syndromes).
- Dates and codes for dementia or mild cognitive disorder diagnoses.
- Dates of death and demographic covariates (age, sex, education, income).
- Derived analysis datasets: exposure-outcome linkage tables; Bayesian multi-state model inputs/outputs (Stan/R objects, MCMC chains, diagnostics).

How will the data be collected or created?

- Data will be extracted by secure linkage from Swedish National Health and Medical Registers (Inpatient Register, Prescribed Drug Register, Cause of Death Register, etc.).
- Data custodians pseudonymise and provide data via secure transfer under ethical permit Dnr 937-18 (addendum 2019-0157).
- Derivative datasets created using R scripts (version-controlled in Git) and Bayesian estimation in Stan.

Documentation and Metadata

What documentation and metadata will accompany the data?

- README files for each dataset outlining variable definitions, formats, and derivation steps.
- Standard Operating Procedures (SOPs) describing data linkage, cleaning workflows, and per-protocol exposure definitions.
- Excel-Codebooks listing ICD-10 and ATC codes, covariate categorizations, and model specifications.

Ethics and Legal Compliance

How will you manage any ethical issues?

- All data use is covered by ethical approval Dnr 937-18 (addendum 2019-0157).
- Only pseudonymised data are used; no direct identifiers.
- Analysis conforms to GDPR and Swedish Patient Data Act requirements
- Pseudonymisation processes is performed by the Swedish National Board of Health and Welfare (Socialstyrelsen)
- No individual-level results will be published.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

- Data provided under a Data Processing Agreement between Göteborgs universitet and Swedish National Board of Health and Welfare/Statistics Sweden.
- No third-party IP restrictions beyond custodial terms; derivative code and aggregated outputs released under CC BY 4.0.

Storage and Backup

How will the data be stored and backed up during the research?

- Raw and pseudonymised master data stored long-term on Göteborgs universitet’s secure research server with nightly, geo-redundant backups.
- Computational copies on UPPMAX/Bianca scratch space for HPC analyses (ephemeral).
- MCMC outputs and post-processing files mirrored weekly from HPC to the institutional server.

How will you manage access and security?

- MFA access via university VPN and additional MFA login for UPPMAX/Bianca.
- Data encrypted at rest (AES-256) and in transit (SSH/SFTP).
- No downloading of raw data to local machines

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

- All pseudonymised raw data necessary to reproduce analyses until project end.
- Aggregated summary statistics, transition probabilities, and counterfactual estimates.
- Analysis code, model scripts, and documentation.

What is the long-term preservation plan for the dataset?

- After analysis completion (expected December 2026), move raw data back to institutional archive.
- Deposit aggregated outputs, code, and metadata in the university’s certified Research Data Archive.
- Retain Stan model objects and diagnostic summaries alongside code in a permanent repository.

Data Sharing

How will you share the data?

- Aggregated results (risk differences, CIFs, mediation estimates) published as open-access journal supplements and in a public GitHub repository.
- Underlying pseudonymised datasets available on managed request via Swedish custodians (Swedish National Board of Health and Welfare/Statistics Sweden) under existing Data Processing Agreement.

Are any restrictions on data sharing required?

- Raw individual-level data not publicly shared beyond custodial channels.
- Requesters must demonstrate ethical approval and DPA coverage for secondary use.

Responsibilities and Resources

Who will be responsible for data management?

- **Principal Investigator**: oversight of ethics, compliance, and final sign-off.
- **Data Manager**: handles secure transfers, backups, metadata records and runs models.
- **Statistician/Analyst**: develops models, documents workflows, performs QC.

What resources will you require to deliver your plan?

- **Compute**: 12 000 core-hours on UPPMAX/Bianca over 12 months.
- **Storage**: 40 GB total (1.2 GB raw; 25 GB MCMC; 6 GB transition analyses; 7.8 GB buffer).
- **Personnel**: Data manager and statistician/analyst salaries
- **Software**: R, Stan