Plan Overview

A Data Management Plan created using DMPonline

Title: Spousal survival Advantage by Marrying into longevity Enriched families (SAME)

Creator: Niels Berg

Principal Investigator: Niels van den Berg

Affiliation: Leiden University Medical Center

Funder: Netherlands Organisation for Scientific Research (NWO)

Template: NWO Template

ORCID iD: 0000-0001-7169-9881

Project abstract:

In the Netherlands, persons with low socioeconomic status (SES) live 8.5 years shorter and spend 24 more years in poor health compared to those with high SES, and these differences are strongly concentrated in families. This project aims to understand how partnership (e.g. marriage or cohabiting) into healthy aging long-lived families affects health and survival within and between generations. Understanding the mechanisms that promote health/survival could lead to novel strategies for reducing health inequalities. Partners may exhibit similarities in health through various mechanisms, such as exchanging resources and behavior or through a like-seeking-like principle. However, these mechanisms have not yet been fully tested. I propose a completely new, family-centered, approach where I will leverage Dutch/Swedish nationwide-registry data to reconstruct two-generation family-networks and study the lifecourses of all family members. This enables longitudinal analysis of socioeconomic resources and health outcomes and within-family comparisons to address causality. By integrating molecular/genetic cohort data, I can study health before disease onset, and apply genetic methods for causal inference. I will first explore the age/sex-specific family dynamics of partnership into long-lived families to test literature/theory-based covariates/confounders and establish novel confounders, such as shared genetic disease predisposition. Second, I will leverage the registry and cohort data to explain the mechanisms by which partnership into long-lived families affects health and survival. I will apply a within-family and genetic approach to assess causality and take shared genetic disease predisposition into account as a confounder. Finally, I will use the two-generational genetic data to study how partnership causally affects socioeconomic resources and health in the next generation. This design allows me to shed light on how health shapes socioeconomic resources, compared to socioeconomic resources shaping health. This project improves our understanding of how resources and behaviors are exchanged between partners and how these exchanges influence multigenerational health outcomes.

ID: 186930

Start date: 05-09-2025

End date: 05-01-2030

Last modified: 11-10-2025

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan 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

Spousal survival Advantage by Marrying into longevity Enriched families (SAME)

General Information

Name applicant and project number

Name: Niels van den Berg
Project number: VI.Veni.241S.034

Name of data management support staff consulted during the preparation of this plan and date of consultation.

Kristel Schaap/Petra van Overveld
LUMC central data stewardship experts

1. What data will be collected or produced, and what existing data will be re-used?

1.1 Will you re-use existing data for this research?

If yes: explain which existing data you will re-use and under which terms of use.

- Yes

In this project I will re-use data from 1) the Leiden Longevity Study, 2) Statistics Netherlands(NL), 3) Statistics Sweden(SW), and 4) Lifelines.

The Leiden Longevity data is available in-house, to me as part of the LLS research team, and as such no constraints apply to my project. Researchers who want to replicate my results can request access to the data by contacting dr. Marian Beekman (datarequest_LLS@lumc.n). At the moment access and usage of the Leiden Longevity Study is free of charge for all scientific researchers. Terms of use: https://leidenlangleven.nl/data-access/.

Data from Statistics-NL is protected by Dutch privacy legislation and Statistics-NL law (Centraal Bureau voor de Statistiek; CBS law), which is more strict than the standard privacy legislation. I already have a running Statistics-NL project, but researchers who want to replicate my results will need to request a Statistics-NL project by following the Statistics-NL microdata procedures: https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen. Terms of use: https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research/rules-and-sanctioning-policy

Similarly, for this project access to LifeLines data is already arranged. To replicate project findings the data access procedures of LifeLines have to be followed: https://www.lifelines-biobank.com/researchers/working-with-us/step-1-prepare-and-submit-your-application. Terms of use: https://www.lifelines-biobank.com/researchers/working-with-us.

Through my position at Lund University I have access to the nationwide Swedish register data

(Statistics Sweden). Similar to Statistics Netherlands, also Statistics Sweden has microdata available. If the published statistics in the statistical database do not provide sufficient information, they are able to produce statistics according to request. For this, and thus to replicate my findings based on the Swedish register data, one must contact them. Terms of use: https://www.scb.se/en/services/ordering-data-and-statistics/.

1.2 If new data will be produced: describe the data you expect your research will generate and the format and volumes to be collected or produced.

New data

No new data will be collected. Over the past years I have invested a lot of time to collect novel data and gain access to existing data in order to carry out my project proposal. Hence, for this project everything is now well-arranged and there is no need to collect additional data.

Generated data

Family networks will be generated by combining multiple datasets from Statistics Netherlands and for all analyses, summary statistics will be generated. Both the family networks and the summary statistics will be saved, if possible, in preferred file formats.

1.3. How much data storage will your project require in total?

- 10 – 100 GB

2. What metadata and documentation will accompany the data?

2.1 Indicate what documentation will accompany the data.

Metadata, including data dictionaries and scripts, necessary to use the generated family network data will be provided in line with CBS guidelines and in accordance with ODISSEI regulations. For all other datasets, documentation will follow the established data management and metadata standards of the respective host institutes.

After the project has been completed, metadata documentation will be saved from the used equipment and/or created, specifying all relevant information needed to replicate my studies. I will also include the necessary software and tools needed for reuse and state whether embargoes, licenses, commercial objectives or other conditions have been imposed on the reuse of data. In line with the Netherlands Code of Conduct for Research Integrity, I will store our raw data for a period of at least 10 years.

2.2 Indicate which metadata will be provided to help others identify and discover the data.

1. In collaboration with CBS, an instruction document containing detailed feature descriptions will be prepared and added to the CBS Microdata Catalog (https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen).
2. In collaboration with ODISSEI, a comparable instruction document will be created and made available through the ODISSEI Data Portal (https://portal.odissei.nl/).

project-level and dataset-level metadata will use the metadata standards of CBS and ODISSEI.

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

3.1 Describe where the data and metadata will be stored and backed up during the project.

- Other (please specify)

All cohort data will be stored at the respective hosting institutes in compliance with national and institutional legislation and regulations, with secure backup provided by the institutes' IT services according to their data management protocols.

CBS register data will be stored and accessed exclusively within the secure CBS environment, where backup and preservation are managed by CBS under strict legal and technical safeguards. The datasets concerned are: (1) the Leiden Longevity Study, (2) Statistics Netherlands (CBS), (3) Statistics Sweden, and (4) Lifelines.

The Leiden Longevity Study data are available in-house and therefore pose no storage constraints for this project. Data from Statistics Netherlands are stored on CBS's secure servers and are subject to Dutch privacy legislation as well as the CBS Law, which imposes stricter requirements than standard privacy regulations. Data from Statistics Sweden are stored on secure servers managed by Swedish authorities, with backups maintained according to national regulations for sensitive research data. Lifelines data are stored and backed up on secure servers at the Lifelines research institute, following institutional policies and national Dutch data protection laws.

No summary statistics, output or other files generated *during* the study, such as documentation, scripts, will be stored locally/at the host institute level. Instead, such output and data will be stored within the CBS environment. When possible such source data will be published on gitlab alongside the publication of a scientific article.

3.2 How will data security and protection of sensitive data be taken care of during the research?

- Additional security measures (please specify)

Cohort data will be managed according to the hosting institutes' rules and comply with Dutch and European privacy legislation (AVG/GDPR). CBS data are subject to additional safeguards under the Dutch CBS Law: all data are fully anonymized, only summary statistics can leave the secure CBS environment, and uploads/extractions are manually checked for disclosure risk by two CBS employees. More information on CBS regulations is available here: https://www.cbs.nl/nl-nl/deelnemers-enquetes/bedrijven/meer-over-cbs-enquetes/handhaving/wet-regelgeving. Swedish registry data is subject to statistics Sweden GDPR legislation which is similar to the Dutch law, with the exception of manual output control; in Sweden disclosure risk during output is the responsibility of the researchers themselves.

4. How will you handle issues regarding the processing of personal information and intellectual property rights and ownership?

4.1 Will you process and/or store personal data during your project?

If yes, how will compliance with legislation and (institutional) regulation on personal data be ensured?

- Yes

Yes: we will adhere to all mandatory safety, security and legal requirements needed to be able to analyse these types of data. For the terms of use per cohort, see question 1.1.

4.2 How will ownership of the data and intellectual property rights to the data be managed?

Re-used data: Data and intellectual property will remain the property of the owner of the cohort.
Data and intellectual property will remain the property of the researchers and their institutions, in line with Dutch law and institutional policies. Access, use, and sharing of the data will follow FAIR principles and agreements recorded in a data management plan and, where applicable, consortium or collaboration agreements.

5. How and when will data be shared and preserved for the long term?

5.1 How will data be selected for long-term preservation?

- All data resulting from the project will be preserved for at least 10 years

All data analyses resulting from the project will be preserved for at least 10 years. For this, we will follow the LUMC guideline for research data stewardship.
For the Statistics Netherlands financial costs will be made to store data and scripts for the legally required period of 10 years.

5.2 Are there any (legal, IP, privacy related, security related) reasons to restrict access to the data once made publicly available, to limit which data will be made publicly available, or to not make part of the data publicly available?

If yes, please explain.

- Yes

Our own analyses will be available/published after consulting our partners from the used registry if we are allowed to do so in line with their terms of use.
Re-used data: see question 1.1, that also lists all terms of use per registry that we use data from.

5.3 What data will be made available for re-use?

- Other (please specify)

see question 5.2.

5.4 When will the data be available for re-use, and for how long will the data be available?

- Data available upon completion of the project

After obtaining data access following CBS and institution procedures data will be available either directly after article publication or, regarding the family networks, upon completion of the project. These data will be available/published if we are allowed to do so in line with the terms of use of the registries.
An embargo period might be necessary to allow publication of the data and time needed to pursue IP-protection. Furthermore, an embargo period may also be needed to establish collaborations with enterprises which could transform the discovery into a clinically applicable intervention. If the experiments lead to patentable discoveries, advice on how best to proceed will be asked from the technology transfer office (LURIS).

5.5 In which repository will the data be archived and made available for re-use, and under which license?

As described above, data is available through the concerning institutes as well as CBS, following the institutes and CBS data access procedures.

5.6 Describe your strategy for publishing the analysis software that will be generated in this project.

Following standard LUMC procedures and GRP guidlines, I will publish all software and code software needed for reuse and state whether embargoes, licenses, commercial objectives or other conditions have been imposed on the reuse of data.
Code will be deposited on the LUMC github as well as CBS and ODISSEI portal
Software, if relevant, will be published on gitlab

6. Data management costs

6.1 What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

€4,000. For further details, see the project budget.