Plan Overview

A Data Management Plan created using DMPonline

Title: Developing and Validating a Generalized Structural Equation Model for Social Network Data

Creator: Terrence Jorgensen

Principal Investigator: Terrence D. Jorgensen

Data Manager: Aditi Manoj Bhangale

Project Administrator: Terrence D. Jorgensen

Affiliation: University of Amsterdam (Universiteit van Amsterdam)

Funder: Netherlands Organisation for Scientific Research (NWO)

Template: Data Management Plan NWO (September 2020)

ORCID iD: 0000-0001-5111-6773

Project abstract:

Accurate theories of psychological processes must include interpersonal interactions, as measured by social-network variables. How to incorporate their complex structure into familiar theory-testing frameworks (e.g., structural equation modeling: SEM) is a topic on the methodological cutting-edge. I propose a two-stage estimation approach. First, estimate correlations of social-network data with person-level (e.g., body image) or group-level (e.g., school climate) covariates, then use SEM to explain those correlations. This innovation would empower researchers to improve their theoretical models of psychosocial processes (e.g., how peer conversations affect disordered eating) and identify effective interventions (e.g., whether antibullying programs work on specific victim–bully pairs).

ID: 119561

Start date: 10-04-2023

End date: 12-04-2024

Last modified: 16-03-2023

Grant number / URL: 406.XS.01.078

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

Developing and Validating a Generalized Structural Equation Model for Social Network Data

General Information

Name applicant and project number

Terrence D. Jorgensen, project number 406.XS.01.078

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

Emma Schreurs (27 February 2023), as part of ERB application.

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

Denissen et al. (2011) have agreed to share their data with us for reanalysis, given than the analyses can be reproduced by publicly posting the data and R syntax. They gathered the data at Utrecht University prior to the establishment of an ethical review board (ERB), so we completed an application to the ERB at the University of Amsterdam to inquire whether any special concerns need to be addressed. We are still awaiting their reply.

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.

The only new data we will produce are simulated data generated for Monte Carlo simulation studies. The simulated data need not be saved because they can be reproduced using the R syntax files that we will use to generate the data. However, we will save the Monte Carlo results, to make it easier for others to replicate our analyses without having to run the entire Monte Carlo study again. The results will be saved as an R workspace (i.e., an *.RData file), and the R syntax will be saved in plain-text files.

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

- 0 – 10 GB

2. What metadata and documentation will accompany the data?

2.1 Indicate what documentation will accompany the data.

For the Monte Carlo simulation studies, R syntax files will be heavily annotated to include instructions for generating the simulated data, as well as for analysis of Monte Carlo results. All files will be made publicly available on the Open Science Framework (OSF).

For the reanalysis of Denissen et al.'s (2011) human-subjects data, we will make a PDF that includes the list of variable names with their descriptions (e.g., interpretation, scale of measurement, missing-data codes). Pending ERB approval, all files will be made publicly available on the Open Science Framework (OSF).

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

The name of the PI and the project (number) will be saved at the top of every plain-text R syntax file. Wherever data are provided, a README text file will include the same information.

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)

Project data will be stored remotely on the PI's secure SURFdrive account, where the PI and data manager can securely access the data with their password-protected laptops.

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

- Default security measures of the institution networked research storage

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?

- No

Simulated data are not derived from human-subjects, so they pose no security concerns.
The human-subjects data shared by Denissen et al. (2011) will not be linked to any personal information, so any data posted publicly or used during the project will be pseudonymized.

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

The Denissen et al. (2011) data will be owned by the original authors, who consent to it being posted publicly on the OSF.
The simulated data will be publicly available via a Creative Commons license (CC-BY).

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?

- Other (please specify)

All simulated data can be simulated again by anyone using R. Files will be available indefinitely via the OSF. Likewise, any Denissen et al. (2011) variables re-used for example analyses will be posted via OSF.

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

If the ERB places restrictions on posting Denissen et al.'s (2011) human-subjects data on OSF, we will abide by those restrictions.

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

- All data resulting from the project will be made available

Upon ERB approval (see 5.2).

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

- Data available as soon as article is published

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

The Open Science Framework (OSF) under a Creative Commons license (CC-BY Attribution 4.0 International)

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

R syntax files will be provided via the OSF to accompany any data analyzed for any published paper.

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)?

Data management is part of the research-assistant's job, which is part of the budget for this NWO-funded project.