Plan Overview

A Data Management Plan created using DMPonline

Title: The effect of study protocol preregistration on animal studies transparency and quality

Creator:Julia Menon

Principal Investigator: Julia Menon, Judith de Haan, Kim Wever, Annemarie Scholman-Végh

Data Manager: Julia Menon

Project Administrator: Julia Menon

Contributor: Mira van der Naald

Affiliation: Utrecht University

Template: Utrecht University DMP

ORCID iD: 0000-0002-3467-1908

Project abstract:

Preregistration, the act of registering a study protocol before the start of the study, is applied in several fields to improve transparency. It enables to state a priori the hypothesis, experimental design and analysis plan. By allowing a comparison between the final manuscript and the planned approach, preregistration is believed to reduce reporting biases (in particular selective outcomes reporting and publication bias), questionable practices such as HARKing, and prevent the unnecessary duplication of studies. At present, preregistration is not yet a common standard in animal research, and little support exists to facilitate it, but previous literature shows promising results from other fields (e.g. clinical research, psychology). No research on its effect in animal studies has yet been formally conducted. Therefore, this project team aims to evaluate if preregistration improves the quality and transparency of animal studies. It involves the scoring of 1) the ARRIVE guideline essential 10 (reporting), 2) risk of bias (using the Syrcle's risk of bias tool) (quality), and 3) consistency between protocol and manuscript (reporting).

ID: 113319

Start date: 09-01-2023

End date: 30-06-2024

Last modified: 21-12-2022

The effect of study protocol preregistration on animal studies transparency and quality

Data Collection

1.1 Will you re-use existing data ?

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

- No, I will be collecting/generating new data

1.2 Describe your data.
Fill the table below with a brief description of the data, including the type, format and volume.

#	Data Description	Data Type	Format	Total Volume
1	Raw answers from scoring ARRIVE essential 10	Tabular	.xls	60 Ko
2	Cleaned answers from scoring ARRIVE essential 10	Tabular	.csv	40 Ko
3	Raw answers from scoring risk of bias	Tabular	.xls	100 Ko
4	Cleaned answers from scoring risk of bias	Tabular	.csv	70 Ko
5	Raw answers from scoring consistency	Tabular	.xls	TBD
6	Cleaned answers from scoring consistency	Tabular	.csv	TBD

Data Documentation

2.1 Describe the documentation and metadata that you will use to to make your data reproducible and interoperable.
Describe which files you will provide, along with a brief description of the information they will contain, to make your data reproducible and interoperable. Describe the information that you will provide to make the data items in questions 2.1 reusable and interoperable. If using a specific metadata standard, please mention this below.

For each phase of scoring, i.e. scoring ARRIVE essential 10, scoring risk of bias and scoring protocols-manuscripts consistency, a .csv file will be created (file level):
A tab called "readme" in each corresponding .csv file will describe the data provided (file level). This tab will contain information on the methodological procedure used to collect and analyse the data. Another tab "code tab", will show

explanations of the terms used in the datasheet (i.e. code book). In particular, it will explain clearly how the scoring was set. Accompanying methodology files will include the study protocol (.pdf file), R analysis setting and codes (.HTML files).

Each line within the csv. sheets correspond to 1 included study (item level). For each study, we will provide metadata on:

-The date and time the study was scored (for the files generated with Google form)

-The manuscript ID (the ID is a random number)

-The initials of the scorer

-Answers to the scoring questions (the set-up of the scoring questions and their scoring will be explained in the study protocol and read-me tab).

To our knowledge, no standard exists for "meta-research on preregistration". Hence we will not use existing controlled vocabulary or specific metadata standards. However, we will provide a "code tab" (as mentioned above) to ensure that others can understand and reuse all variables.

All files will be uploaded to the Open Science Framework, with links to each other using the DOI format. The given DOI will be mentioned in the final paper. Using this DOI system makes our data Findable and Accessible (as the OSF is an open platform).

An overview .pdf file will give an inventory of the files provided and the meaning of the different variables.

2.2 Describe the folder structure you will provide to make your data reproducible and interoperable.

Describe the folder structure, naming conventions and/or version control you will use for this project.

Our folder structure will be broken down into the steps of the research project i.e. 1) preparation, 2) collection, 3) scoring, 4) analysis. The files will be divided by raw data and cleaned data.

>Project Folder

>> Scoring

>>> Raw data

e.g. ARRIVE10_scoring_V#, Riskofbias_scoring_V#

>>>Cleaned data

We will keep track of versions by adding V# at the end of filenames when there is the possibility of new versions (i.e. filename_V1)

Data Storage

3.1. Select the storage solution where you will store and back-up your data.

Select the locations where your data will be stored. You may select more than one. Please describe the storage solution and the backup strategy of your storage solution if it does not appear in the list below.

- Other (please specify below)

Data collection sheets are shared amongst contributors and will be filled in online using Google Forms and Google Sheets. The resulting data (raw, cleaned and analysed) .csv files will be stored long-term and backed up by the Netherlands Heart Institute "Own Cloud" Service.

For sharing publicly, we will upload each file to the OSF repository (in our existing project).

Data Privacy and Security

4.1 Will you be collecting or using personal data ?

Personal data is any data which, alone or in combination with other information, can identify a living person. Such data must abide by the GDPR and requires additional safeguards and documentation to be processed lawfully.

- No, I will not collect and/or use personal data

4.2 How will ownership and intellectual property rights of the data be managed?
Describe who controls access to the data and who determines what is done to the data.

The principal investigators will create the data and have access to the raw data to edit, update and/or delete if necessary.
The project administrator/main principal investigator and the statistician allocated to this project will have access to and handle the clean data and analysed data (edit/update/delete). Read-only and edit-granted access can be provided to other project members (also members located outside the Netherlands) by requesting it from the Project Administrator. Access will not be granted to other individuals until the project is finalised and published.
Intellectual property rights are shared amongst the different institutes involved in this project: i.e. Utrecht University and the Netherlands Heart Institute (for creating the data). To note: project members outside the Netherlands, i.e. Center for Open Science and the German Centre for the Protection of Laboratory Animals (Bf3R), actively participate in the selection of manuscripts/controls used to create the data but do not create the data themselves.
When the data is shared on the OSF, it will be freely available under a CC-BY license.

Data Selection, Preservation & Sharing

5.1 Describe the data you will be preserving and the storage solution where it will be preserved?
Describe which data will be preserved under long-term storage. You may refer back to the data described in question 1.2 to specify which data will be preserved. Explain where you will preserve your data, and how procedures are applied to ensure the survival of the data for the long term.

All collected data will be preserved for at least 10 years.
The latest version of the data sets (raw and cleaned) for each stage stated in 1.2 will be uploaded to the OSF and the Netherlands Heart Institute "Own Cloud" as well as on a physical media (principal investigator's hard drive) for long-term preservation. The aim is to have several copies to have a backup in case one of the copies becomes faulty/corrupted.
The main investigator for this study will handle the governance of the data and ensure that the files get updated to a proper format if necessary. Refreshing will also be performed by placing the files into a proper new media every 4 years to prevent media defects.
The OSF project will be accessible to all; the Own Cloud is accessible to the project members.

5.2 Describe the data you will be sharing and the repository where it will be shared?
Describe which data you will be sharing. Select where you will make your data findable and available to others. If selecting "Other" please specify below which repository and provide a URL.

Please also write below if you will apply any conditions to the re-use of your data. (i.e. Creative commons license or Data Transfer Agreement).

- Open Science Framework

A complete data package will be uploaded to the Open Science Framework and be re-usable under a CC-By Attribution 4.0 International. It will include
-All data sheets (raw and cleaned) with accompanying files (mentioned below);
-R code and analysis output (in HTML)
-protocol (.pdf) (preregistered)
-Preparation files used in the selection process of protocols/manuscripts/controls (.csv)
-An overview file explaining the package (.pdf)

They will also be referred to in the final paper (using the DOI provided by the OSF).

5.3 Are specialized, uncommon or expensive software, tools or facilities required to use the data?

Please list any specialized, uncommon or expensive software, tools or facilities that are absolutely required to obtain, use or handle your data, if any.

No, all the data can be accessed by free, open-source or non-proprietary software.

Data Management Costs and Resources

6.1 What are the foreseeable research data management costs and how do you expect to cover them ?
Please specify the known and expected costs involved in managing, storing and sharing your data. Also explain how you plan to cover these costs.

The data we will generate represents a small volume and should not lead to substantial costs.
We will use the existing Own Cloud subscription held by the Netherlands Heart Institute, which will not yield additional costs.
We will upload our data and accompanying files to the OSF for free, as we will not exceed the 50 GB OSF public storage limit.
In the upcoming year, the main investigator will buy with her own money a new, more effective hard drive (for about 50-100 euros).

6.2 Who will be responsible for data management?
Please specify who is responsible for updating the DMP and ensuring it is being followed accordingly.

The main investigator, Julia Menon, will be responsible for maintaining the DMP up to date. She will also be responsible for granting permissions and ensuring the data is deposited in the OSF, backed up in the Own Cloud and her personal hard drive.

6.3 State if you contacted an RDM consultant from Utrecht University to help you fill out your DMP.

Please list their name and date of contact.
This is mandatory for NWO grants.

We received guidance from an RDM consultant from the Durrer Center (Evelien van der Schaaf) in May 2023.
No, we did not receive help from an RDM consultant from Utrecht University.