Plan Overview

A Data Management Plan created using DMPonline

Title: Focus groups with vaccine hesitant people

Creator: Stefan Gaillard

Principal Investigator: Stefan Gaillard, Stefan van Geelen, Marcel Hobma

Data Manager: John Meeuwsen

Project Administrator: Stefan van Geelen

Contributor: Willem Sanders

Affiliation: Utrecht University

Template: UMC Utrecht DMP with DPIA V.3.0

ORCID iD: 0000-0003-1956-7325

ORCID iD: 0000-0001-8216-4268

Project abstract:

In this study, we explore why different groups of people have doubts about mainstream healthcare, particularly vaccinations. What concerns do they have, and how do they feel about them? We are especially interested in how skepticism relates to trust in the government and its institutions, as well as what participants believe the government could do better.

Different groups seem to have different reasons for questioning healthcare and vaccinations. Our research focuses on three groups: people with a migration background in Kanaleneiland, residents of the Friesland region, and highly educated individuals critical of mainstream health policies.

This study is commissioned by the Ministry of the Interior and Kingdom Relations in response to declining vaccination rates in the Netherlands. The findings will be compiled into a report for the Ministry of Health, Welfare, and Sport, to be published in mid-2025.

ID: 173010

Start date: 31-03-2025

End date: 13-06-2025

Last modified: 31-08-2026

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

Focus groups with vaccine hesitant people

1. General features

1.1. Acronym/short study title

Focus groups with vaccine hesitant people

1.2 Division of Principal Investigator

- Onderwijscentrum (Education Center)

1.3 Department

De Nieuwe Utrechtse School

1.4 Path of the Research Folder

\\ds\data\DOO\Onderzoek

1.5 WMO/DEC

- non-WMO

1.6 Research type(s)

- Qualitative

Focus groups

1.7 Research design(s)

- Observational
- Other

Focus groups

1.8 Mono or multicenter study (one choice)

- Monocenter

1.10 Which organization is the sponsor of the study?

Ministerie van Binnenlandse Zaken

1.11 Name of datamanager consulted

John Meeuwsen

1.12 Last check date by datamanager

2025-04-08

1.13 Indicate which laws and regulations are applicable for the project (please check all that apply)

- Richtlijn Kwaliteitsborging Mensgebonden Onderzoek (Quality Assurance for Research Involving Human Subjects)
- Nederlandse gedragscode wetenschappelijke integriteit
- Algemene Verordening Gegevensbescherming (AVG) or General Data Protection Regulation (GDPR)

2. Data Collection

2.1 Give a short description of the research data.

We are organizing max. four focusgroups with 8-12 people that will be recorded and produce audio files. After storing these audio files on protected servers, we will transcribe the audio files (as for now, without the use of any software). These transcripts will be anonymized text documents (DOC), that we will use for thematic analysis.
During the thematic analysis we will keep track of how the analysis is conducted, and choices that are being made, ensuring replicability of our results.

Subjects	Volume	Data Source	Data Capture Tool	File Type	Format	Personal data involved?
Humans	4 focus groups x 8-12 people (audio recording)	Focus group	Audio recorder	Audio	WAV	Yes
Humans	4 x 8-12 people (anonymized transcript)	Focus group	Audio recorder	Qualitative (text document)	DOC	No

2.2 Describe the flow of the data (name systems used and/or third parties, recipients) <add link to location where diagram is stored in RFS>

1: The data is recorded on an offline audio-recorder.
2: The data is uploaded to the protected Research Folder of UMC Utrecht, where it remains for 10 years.
3: The audio data will be deleted from the audio recorder after uploading is complete
4: The data is transcribed on a personal computer that has access to the protected Research Folder (computer of S.G., transcription mostly by M.H.). Personal information, including but not limited to name, age, address, nationality, ethnicity and other specific details of one's life are anonymized.
5: The transcription will be stored using One Drive.

2.3 Estimated storage space for your project

- < 250 GB (e.g. questionnaires, textfiles, datasets)

2.4 Can you reuse existing data? If so, list the data source(s)

- No, please specify below

Focus groups will be new and no previous data will be used.

2.5 Describe how you will take care of good data quality.

#	Question	Yes	No	N/A
1.	Do you use a GCP-compliant Data Capture Tool or Electronic Lab Notebook?	x		
2.	Have you built in skips and validation checks?			x
3.	Do you perform repeated measurements?			x
4.	Are your devices calibrated?			x
5.	Are your data (partially) checked by others (4 eyes principle)?	x		
6.	Are your data fully up to date?			x
7.	Do you lock your raw data (frozen dataset)	X		
8.	Do you keep a logging (audit trail) of all changes?	X		
9.	Do you have a policy for handling missing data?			x
10.	Do you have a policy for handling outliers?	X		

Explanation:
1: We will make use of a GCP-compliant audio recorder.
5: For our thematic analysis, which will be conducted by Marcel Hobma, the other researchers (notably Stefan Gaillard) will cross-check thematization, and check the work for potential biases and misinterpretation.
8: We will keep track of the process of analysis, so future researchers can replicate our findings.
10: Possible deviant data (outliers) will be removed from the analysis and results, but mentioned and acknowledge in the discussion of our project.

2.6 Specify data management costs and how you plan to cover these costs.

#	**Type of costs**	**Division** ("overhead")	**Department**	**Funder**	**Other** (specify)
1.	Time of datamanager	x			
2.	Storage	x			
3.	Archiving	x			
4.					
5.					

2.7 Please give some more details on other centers and organizations involved. What are the roles of the other centers and organizations involved? (What research activity does this organization carry out in relation to the study and the data?)

Organization	Role/research activity
Marcel Hobma (freelance journalist and researcher)	Principle investigator (MH will conduct the focus groups and the majority of the thematic analysis)
Ministry of the Interior and Kingdom Relations	Our employer for this research project. The Ministry will not have access to our data (audio or transcriptions).
Ministry of Health, Welfare, and Sport	Recipient of the rapport that will discuss the results of our focus groups

2.8 Which contracts are in place?

Organization	Contract Type with UMCU	JOIN number
Marcel Hobma (freelance journalist and researcher)	Affiliate Researcher of De Nieuwe Utrechtse School	nvt

2.9 State how ownership of the data and intellectual property rights (IPR) to the data will be managed

UMC Utrecht is and remains the owner of all collected data for this study.

2.10 Use of new technology. Does your study involve the implementation of a technology that has not been used before at UMC Utrecht?

- No

2.12 Will the study need/use personal data (directly or indirectly identifying)? For example, from the Electronic Patient Files (EPD; HiX), DNA, body material, images or any other form of personal data?"

- Yes. You have indicated that you are using personal data in your project. The following chapter is the Data Protection Impact Assessment (DPIA) for research data. It is derived from the full DPIA, in accordance with the privacy office of UMC Utrecht. Answering questions in this chapter helps to determine the risk of processing the personal data and what measures to take to minimize these risks.

We will mention the socio-economic background of our participants.

3. Data Protection Impact Assessment (DPIA)

3.1 Describe the recipients outside the UMC Utrecht to whom the personal data are provided, what their role is (controller or processor) and where they are located.

- All systems and service providers involved are mentioned in question 2.1 and 2.2. All of them are already contracted by UMC Utrecht. I do not share personal data with other organisations.

3.4 What type of sensitive personal data will be used?

- Political opinions
- Religious or philosophical beliefs
- Racial or ethnic background

Focus groups will be with people from a specific socio-economic and in some cases ethnic background. We will ask them about their vaccine hesitancy, in relation to their trust in various institutions such as the government. During their answers, they might reveal religious or philosophical beliefs as well as political opinions.

3.5 What type of directly or indirectly identifying personal data will be used? Indicate why you need this data. Is this truly necessary?

Only socio-economic background and ethnicity. This is needed for our analysis of the dominant reasons for vaccine hesitancy among various subgroups. We believe there might be a link between these characteristics and vaccine hesitancy.
We will also collect names and emails for initial contact and follow-ups. Names are not connected to research data.

3.6 Select any vulnerable groups from which you will collect data.

- Other --> describe

3.7 Which legally prescribed personal number will be used? Note: it is NOT allowed to use BSN (or its international counterpart) for scientific research purposes.

- None

3.8 Can the purpose of the study be achieved with anonymous or pseudonymized data?

- Yes, I collect pseudonymized data from subjects in a questionnaire, I do not need to know the identity of the subject

Participants of the focus group will be pseudonymized (Participant 1, participant 2) and the pseudonymized transcripts will be analyzed. There is no questionnaire but rather a focus group.

3.9 Which measures are taken to prevent the data from being traceable to the natural person? Also consider the measures taken to prevent data breaches.

- Clear retention period(s)
- 2FA/MFA before access to (health) data
- Minimalization of collected data points
- Pseudonymization of data

3.10 Does the reuse of the data fit within the purpose for which they were originally collected?

- Not applicable, we will not reuse data

3.11 Are data subjects contacted and included only after informed consent?

- Yes, we ask study-specific or other type of Informed consent (e.g. broad consent, deferred consent).

3.16 What type of consent for using personal data is obtained?

- The informed consent includes the option to share data with third parties for reuse.
- Study-specific or other type of Informed consent (e.g. broad consent, deferred consent, explain).

3.17 Is there a dispute settlement or a party where the subject can go to with questions or complaints about the processing of personal data?

- Subjects are provided contact information whom and how to contact the study team via the PIF. Also, subjects are informed about their possibility to contact the data protection officer (DPO) or supervisory authority (Autoriteit Persoonsgegevens).

3.18 Describe how you manage your data to comply to the rights of study participants.

- Right to be forgotten is limited (e.g. because this will have a major (bias) impact on the research which cannot be corrected for)
- Access rights are limited (e.g. because we do not store the key table)
- A subject can object to processing of their personal data or withdraw consent
- We inform the subjects about their rights of access, rectification and deletion of their data. In the information provision we describe the contact information in case a subject wants to exercise their rights,

In the case a participant wants to access, rectify or delete their data, we will comply and delete the audio file, or - if the file hasn't been transcribed yet - censor the sections where the participant can be heard.

If the recording is already transcribed, the sections of the transcription involving the participant will be deleted where possible. We will ask the participant which section or citation they would like to see removed from the transcript.

Because of the pseudonymization process, it might not be possible to trace back all quotes of the participant. Full removal of the citation might therefore not be possible, although we will try to comply to the participant's wish. Due to this practical issue, the access rights of the participants will be slightly limited. We will inform the participants about this in our information letter.

3.19 Does the data collected concern data from which behavior, presence or performance (profiling) can be measured when this is not the purpose of the research?

- No

3.20 Are automated (i.e. without any human intervention) decisions made about the subjects based on the data?

- No

3.21 Describe the tools, procedures and transport methods that you use to ensure that only authorized people have access to personal data

- We use the secured Research Folder Structure that ensures that only authorized personnel has access to personal data, including the key table that links personal data to the pseudoID

The key table and the audio files can only be accessed by the data manager (John Meeuwsen), the Principal Investigator who has access to the UMC folder structure (Stefan Gaillard) and the project coördinator (Stefan van Geelen). It can be found in the secure research folder structure of UMC Utrecht.

3.22 Describe your backup strategy or the automated backup strategy of your storage locations.

- All (research) data is stored in the RFS on UMC Utrecht networked drives from which backups are made automatically twice a day by the division IT (dIT).

3.23 Describe who will have access to which data during your study.

Stefan Gaillard and Stefan van Geelen will have direct access to the data in the Research Folder Structure. The principal investigator (Marcel Hobma) will only have access to the pseudonymized transcripts and the raw audio-files (before deletion, within two weeks of recording).

Type of data	Who has access
Direct identifying personal data	Datamanager and coördinator (Stefan Gaillard and Stefan van Geelen)
Key table linking study specific IDs to Patient IDs	Datamanager and coördinator (Stefan Gaillard and Stefan van Geelen)
Pseudonymized data	PI, Datamanager, project coördinator, intern

MB: can you say the data is pseudonymized for the PI when he does have access to the raw audio-files?

3.24 Indicate the ISO who was consulted for this DPIA and what advice follows from this?

- Positive, conditional (describe)

1. Please check again with the datamanager if OneDrive or perhaps SurfDrive of Stefan Gaillard could be used for the analysis of the transcripts by Marcel Hobma
2. Ensure that the information letter is sufficiently clear about where data subjects' rights are restricted by pseudonymisation

5. Metadata and Documentation

5.1 Describe the metadata that you will collect and which standards you use.

We will collect a minimal amount of metadata, and will not use this in our study. For example: the time of recording.

5.2 Describe your version control and file naming standards.

Our file naming standard is based on date and name of the subgroup: for example: 20250311 Friesland Group 1

We will distinguish versions by indicating the version in the filename of the master copy by adding a code after each edit, for example V1.1 (first number for major versions, last for minor versions). The most recent copy at the master location is always used as the source, and before any editing, this file is saved with the new version code in the filename. The file with the highest code number is the most recent version and older versions are moved to a folder OLD. The major versions will be listed in a version document (projxVersDoc.txt), stating the distinguishing elements per listed version.

6. Data Analysis

6 Describe how you will make the data analysis procedure insightful for peers.

- It is anticipated that we are going to write a paper and publish it, which will make the research accessible to peers.
- Other

After transcription and thematic analysis of the focus groups audio files, we will write out the results, discussion and results. These sections of our research will become part of a public report on medical misinformation in the Netherlands.
Possible, we will also make the research accessible to peers by publishing the whole study as an academic article.

7. Data Preservation and Archiving

7.1 Describe which data and documents are needed to reproduce your findings.

Audio files will be retained for 10 years. In addition, our anonymized transcripts and log used in our thematic analysis will be available for 15 years.

7.2 Describe which archive or repository (include the link!) you will use for long-term archiving of your data and whether the repository is certified.

- After finishing the project, the data package will be stored at the UMC Utrecht Research Folder Structure and is under the responsibility of the Principal Investigator of the research group. The (meta)data will be published in DataverseNL, the preferred UMCU repository.

7.3 Give the Persistent Identifier (PID) that you will use as a permanent link to your published dataset.

- A PID will be generated when a data package is published on DataverseNL. This PID will be updated when available in the additional comment area of this plan.
- I cannot publish the dataset in an external repository. Therefore, I do not have a PID.

8. Data Sharing Statement

8.1 Describe what reuse of your research data you intend or foresee, and what audience will be interested in your data.

We intend to reuse our research data for an academic publication following the publication of our report.

8.2 Are there any reasons to make part of the data NOT publicly available or to restrict access to the data once made publicly available?

- Yes (please specify)

Data consists of non-pseudonymized audio files which could identify participants based on for example the sound of their voice.

8.3 Describe which metadata will be available with the data and what methods or software tools are needed to reuse the data.

All metadata (date of recording) will be available with the data.
Including, but not limited to:
a timestamp
demographic of people interviewed
name of researcher involved
actual situations happening around the time of the focusgroup.

8.4 Describe when and for how long the (meta)data will be available for reuse

- (Meta)data will be available upon completion of the project

8.5 Describe where you will make your data findable and available to others.

N/A

Planned Research Outputs

Journal article - "Medical mis- and disinformation: Underlying Factors"

Report - "Medische mis- en desinformatie: handelingsopties"

Planned research output details

Title	DOI	Type	Release date	Access level	Repository(ies)	File size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Medical mis- and disinformation: Underlying Factor ...		Journal article	Unspecified	Open	None specified		None specified	None specified	No	No
Medische mis- en desinformatie: handelingsopties		Report	Unspecified	Open	None specified		None specified	None specified	No	No