

Example Data Management Plan – complete version with full responses

Understanding patient symptoms, thoughts and experiences of hyperemesis gravidarum during antenatal and postnatal periods



Data Stewardship Wizard (DSW) Example Data Management Plan

Complete version with full responses

Organisation
ELIXIR-UK

Created by



Jamie
McQueen

id [0000-0001-7172-7504](#)

Based on

Based on Standard DMP template for Biotechnology and Biological Sciences Research Council (BB-SRC) 2024



Funded by the [ELIXIR-UK: FAIR Data Stewardship training UKRI award](#) (MR/V038966/1) and BioFAIR

(Q1) Proposal name

Understanding patient symptoms, thoughts and experiences of hyperemesis gravidarum during antenatal and postpartum periods.

(Q2) Description of the data: Type of study

This study aims to develop a greater understanding of the symptoms and experiences of women diagnosed with hyperemesis gravidarum. We will collect data from antenatal and early postnatal questionnaires, followed by interviews at 6 months and 12 months postpartum.

(Q3) Description of the data: Types of data

We will generate semi-quantitative data from HELP (HyperEmesis Level Prediction) Scores. We will also generate qualitative data from the questionnaires, including first-hand descriptions of patient experience. Audio data and downstream transcripts will be generated from those participants who consent to be interviewed. We will also create associated documentation such as electronic participant consent forms, distress protocol(s), question lists, interim reports and analyses.

(Q4) Description of the data: Origin of the data

This study will generate novel primary data. We will use existing related datasets from published literature within the field, where possible, for comparison and verification of findings.

(Q5) Description of the data: Format and scale of the data

We aim to sample approximately 1000 women across the UK via questionnaires and to have follow-up interviews with up to 50. Questionnaire data will be stored within a .xlsx spreadsheet, a standard machine-readable format. Audio data will be stored short-term in .mp4 format but will be automatically transcribed using software and transcripts stored as .doc files following quality control checks. We estimate that this study will generate approximately 50 GB of data.

(Q6) Managing, storing and curating data

Data will be collected and managed in line with the University Research Data Management policy in addition to local institutional policies and SOPs. All data and associated documentation will be stored on the [institutional data storage system]. This is a secure, enterprise-class networked storage system with disaster recovery mechanisms in place including multi-site backup.

Data will be pseudonymised and each participant will be automatically assigned a unique identifier. The key to re-identification will be held by the group leader and securely stored within an encrypted container within [institutional data storage]. This will be in a separate location from the project data and will be inaccessible to anyone other than the project lead. Participants will be asked to provide their date of birth as an initial way to distinguish between individuals with similar names however once the participant questionnaire has been given a pseudonymised ID we will stratify participant age within a band (e.g., 18-24, 25-29 etc).

All data will be stored within a project folder containing sub-folders for each dataset and access will be limited to named individuals working on the project. A consistent file-naming strategy will be established within the group at the outset of the project to allow for efficient organisation and continuous data curation.

(Q7) Metadata standards and data documentation

We will generate and maintain detailed metadata describing all aspects of our methodology including data collection, inclusion/exclusion criteria, data cleansing and statistical analyses. Questionnaire data will use controlled vocabularies and established coding schemas when possible and we will create a data dictionary and user guide to describe all variables. For interview data, we will follow the 'Consolidated criteria for reporting qualitative research' (COREQ) standards to ensure maximal metadata capture.

(Q8) Data preservation strategy and standards

All data and associated documentation will be stored in standard machine-readable formats for long-term preservation. Data will be archived using the University data archive facility and will be curated as per the facility's requirements. All data will be retained for a minimum of 10 years following study completion in compliance with the BBSRC Data Sharing Policy.

Aggregated data, stripped of all identifying information, will be collated and deposited with the UK Data Service (outlined below) and will be available upon request following publication.

(Q9) Where will data be shared?

We plan to share data and associated metadata by depositing with the UK Data Service. All data will be stripped of identifying information. Data will be curated using repository guidance and requirements and will include associated metadata as described above. The data will not be open-access but will instead be 'safeguarded'. This is to prevent the identification of any of the women consenting to follow-up interview(s) as we anticipate a relatively small sample size. By safeguarding the data this way, we will ensure that anyone seeking access to the data will be bound by the terms of the UK Data Service End User Licence Agreement (preventing attempts at identification) in addition to justifying the reasons for their request.

(Q10) When will data be available?

Data will be available upon request at the point of publication to ensure full transparency and to maximise possible reuse by others. We do not anticipate any delays in data sharing and will endeavour to respond to access requests in a timely and responsible manner.

(Q11) How will data be made findable and accessible?

Summary data will be accessible through open-access publication which will include a Data Access statement. Data will be deposited with the UK Data Service and will be assigned a DOI. Whilst the data will not be open-access we will readily provide access to those with valid justification. We will present our findings at national and international conferences and will engage with the University Press Office to create a press release if the findings are of significant public interest.

(Q12) How will data be made reusable?

Data will be deposited with the UK Data Service and will contain comprehensive metadata in line with repository requirements to enable reuse by others.

(Q13) Restrictions or delays to sharing, with planned actions to limit such restrictions

Data will be 'safeguarded' by the UK Data Service and

will require potential users to submit contact information and provide brief justification for their request. We will endeavour to review requests and grant access in a responsible, open-minded and timely manner in line with the BBSRC Data Sharing Policy.

(Q14) Secondary use

We anticipate that this data will be of interest to other researchers working within the field both nationally and internationally. It may also be of interest to external stakeholders such as relevant charities/initiatives and may also help inform policy.

(Q15) Formal information/data security standards

None Identified

(Q16) Main risks to data security

The main risk to data security in this study is the identification of participants. This will be mitigated by the pseudonymisation of participants and the stratification of data into bands such as those for age or geographical location. Once a questionnaire is digitally received it will be automatically assigned a participant number. The identifiable information linking the participant with their unique participant number will be recorded in the master spreadsheet. This will reside in an encrypted container within a folder on [institutional storage] which will be in a completely separate location to the study data. The master spreadsheet will only be accessible to the project lead. An audit trail will be in place to log when the file was accessed and the reasons for access. All analyses will be based on the pseudonymised dataset(s) therefore we do not anticipate that the master spreadsheet will be accessed other than to comply with a participant wishing to withdraw themselves and their data from the study. A full DPIA for the study will be submitted to the University Data Protection Officer for approval before study commencement.

Finalised datasets will be deposited with the UK Data Service and will be 'safeguarded' as outlined above. This is to prevent identification of the small number of women who consent to follow-up interview(s) and ensures that any party seeking to access the data is bound by the UK Data Service End User Licence Agreement which prevents attempts at identification.

(Q17) Capabilities

The University has a Research Data Management Policy and secure data storage infrastructure to allow efficient and optimised research data management. All costs for data storage have been included in this application and we do not anticipate incurring additional costs.

(Q18) Maintaining and implementing the Data Management Plan

This data management plan is a living document and will be reviewed by the PI every 6 months to ensure the project data management is on track and to revise if necessary.

(Q19) Environmental considerations

The University is a signatory of the Concordat for the Environmental Sustainability of Research and Innovation Practice and is committed to net zero by 2040 as part of the Climate Strategy. We will ensure all equipment such as laptops and printers are turned off and unplugged whenever possible and we will join meetings such as conferences virtually wherever possible. We will additionally ensure that all members of the research group spend time on appropriate sustainability training.

(Q20) Responsibilities

All members of the research team, including the group leader, are responsible for data management, maintaining appropriate documentation and quality assurance of the data. Responsibility for data security lies with the group leader as well as the University Information Services and Information Security divisions who are responsible for University-wide security infrastructure.