HEY BABY_DMP

A Data Management Plan created using DMPRoadmap

Affiliation: University of Cape Town (UCT-Generic)

Last modified: 12-05-2021

1. PROJECT NAME

PROPOSAL/PROJECT NAME - Replicate exactly as in the proposal or project that the DMP accompanies.

Helping Empower Youth Brought up in Adversity with their Babies and Young Children (HEY BABY)

2. INTRODUCTION/TYPE OF STUDY

STUDY INTRODUCTION

The HEY BABY study aims to assess resilience-promoting pathways for adolescent parent families living in adversity, including young parents living in resource-constrained, HIV-affected communities. It asks two groups of research questions: (1) what puts adolescent parents and their children at risk of disadvantage? and (2) what services can help adolescent parents and their children? This is a cohort study, which hopes to follow the mother-child dyads for at least two follow-up data points. The HEY BABY Study utilises a mixed methodology to capture the complexity of resilience combining adolescent parent and child data. Qualitative work will inform the design of the questionnaires by combining various research approaches, including in-depth interviews, focus groups discussions, observations, and participatory research. In turn quantitative findings will inform further qualitative exploration of emerging themes. The quantitative arm of the study will utilise longitudinal data from parent-child dyads to conduct quasi-experimental analyses.

3. DESCRIPTION OF THE DATA

EXISTING DATA

The ‘Mzantsi Wakho’ (‘Your South Africa’) study also referred to as ‘MW study’ recruited 1,500 adolescents from the Eastern Cape province, South Africa (Cluver 2015). In three repeat-measure quantitative rounds from 2015 to 2017 (T1-T2 retention rate 94%, T2-T3 retention rate of 97%). At baseline in 2014-2015, the study included 141 adolescents who were pregnant or parents and by the end of the third wave of data collection in 2018, a total of 152 were recruited into the HEY BABY study. During each scheduled follow-up period, the HEY BABY team will continue to recruit adolescent parents from the Mzantsi Wakho cohort. Data from the Mzantsi Wakho study will be combined with the data from HEY BABY study to allow for longitudinal analyses to incorporate before and after pregnancy experiences relating to resilience and adversity.

Data that will be used from the Mzantsi Wakho study include the final data output which is a merged T1, T2 and T3 dataset. Cleaned data (in .dta and .sav format) for each wave was merged together to form a wide format dataset that combines responses from participants for all the three-time points. Patient file data for the participants was also merged into this T1T2T3 dataset, which include clinic information such as biomarkers, type of ART, etc. In this same study we also collected clinic level data based on the Clinic level questionnaire (CLQ) by interviewing clinic stuff where adolescents accessed health. The dataset is available in STATA (.dta) format.

DATA DESCRIPTION

The HB study has two streams: quantitative data collection, and qualitative data collection.

The quantitative arm of the HB study aims to follow up on adolescent parents (n=1046) and their children (<6 years, n=1145) in the Eastern Cape Province, South Africa who were recruited between 2018-2019. During follow-up, the study will also enrol participants from the Mzantsi Wakho cohort who may have become young parents since the previous wave (baseline or first round of follow-up).

The study will combine four different quantitative data sources: (1) Adolescent questionnaires, (2) Child developmental assessment, (3) Child and parent health data from clinical records (where available) including Road to Health cards and (4) Adolescent parent and child laboratory outcomes (HIV test results, viral load, TB tests) from the National Health Laboratory Service (NHLS). Adolescent questionnaires were drafted after consultations with adolescent health researchers and practitioners in the field and piloted with study participant recruited at baseline. Measures use tools validated (where available) in Southern Africa.

1. **The Adolescent Questionnaire:**

This questionnaire focuses on overall adolescent data: socio-demographics, health, school, family, access to services, parenting, and experiences for adolescent parents. This is a self-report questionnaire completed by all study participants. Before COVID-19,. During the first wave of follow-up (2020-2021), data is collected using a combination of remote and face-to-face methods due to the COVID-19 lockdown restrictions and subsequent COVID-19 mitigation efforts which will be put in place due to easing restrictions. At the time of compiling this DMP, it is anticipated that the adolescent questionnaire will be divided into 3-5 parts- referred to by phase number.

- Baseline: the adolescent questionnaire was administered in two parts, each taking about 45 mins. The first part (Interview 1) focused on overall adolescent data: socio-demographics, health, school, family, access to services, etc. The second part (Interview 2) focused on parenting experiences for adolescent parents and data on each of their children. Response from baseline have been used to create a tracking database for follow-up purposes, as an SQL database and as excel rosters.
- Phase 1 (Follow-up) of the adolescent questionnaire includes data on (i) education, employment and income experiences, (ii) health and wellbeing of the adolescent parent and child, (iii) COVID-19/lockdown specific questions, (iv) home and family, and (v) resilience; the questionnaire was made up of 90 questions. The nature of follow-up questions for phase 1 required referral availability for certain services/resources, based on identified needs arising from the questionnaire. These include referrals for adolescent and/or child ARV defaulting, food insecurity, child neglect, counselling, disclosure and grant access.
- Phase 2 (Follow-up) of the adolescent questionnaire includes data on (i) mental health, (ii) exposure to violence, (iii) sexual reproductive health, (iv) parenting experiences and stress, (v) attachment, (vi) stigma and bereavement, (vii) relational resilience, (viii) aspirations, and (ix) self-efficacy; the piloted questionnaire will include 201 questions. The nature of follow-up questions for phase 2 requires referral availability for certain services/resources, based on identified needs arising from the questionnaire. These will include referrals for adolescent counselling, GBV and sexual violence, mental health and substance abuse.
- Phase 3 (Follow-up) of the adolescent questionnaire includes data on the antenatal care experiences for young parents recruited into the study post baseline and for newly recruited participants from the Mzantsi Wakho study.

2. **Child Developmental Assessment:**

The Mullen Scales of Early Learning (MSEL), which is a standardized validated normative measure of child development across five scales was used at baseline. During follow-up, the team will continue to use the MSEL scale in conjunction with Malawi Development Assessment Tool (MDAT) to measure social-emotional adjustment skills of the children. However, due to the COVID-19 pandemic and subsequent move to remote data collection, this has not been done in the current 2 phases of data collection. As we resume face-to-face activities in June 2021, when suitable, we will be conducting child development assessments that measure the same areas as MSEL. The amount of time required to complete the new assessment will be shorter in order minimise the amount of contact time with participants.

3. **Child and parent health data from clinical records:**

During baseline, the data included data extracted from 1052 government-issued home-based children's health records either as a photo of the road to health card or extraction of the maternity case record (MCR) folder. Due to COVID-19

pandemic, extraction of MCR is not scheduled for 200/21 period. However, the study team is considering taking RTH card photos when F2F study activities resume 2021.

4. **Adolescent parent and child laboratory outcomes (HIV test results, viral load, TB tests) from the National Health Laboratory Service (NHLS):**

Most adolescent parents and their children access public health services in one of 70 health facilities that our research team works closely with. These facilities collect and store health information such as access to PMTCT (prevention of mother to child transmission of HIV) care, child HIV testing and results, and child immunization records. During baseline data collection, the study obtained consent from adolescent parents and caregivers to access this information from their electronic patient files. Through partnership with NHLS, we will match – using an algorithm developed and tested by the NHLS team – adolescent parent and child laboratory outcomes (HIV test results, viral load, TB tests) to their interview data. Parameters of matching are based on personal identifiers provided at baseline. The matching will enable the research team to use social science data and methods to identify what can help with health outcomes measured through rigorous biomedical tests.

The qualitative arm of the study includes in-depth semi-structured interviews with health care workers and cognitive interviewing with adolescent participants.

1. **Health Care Worker (HCW) study:**

The HCW study aimed to collect qualitative research data investigating how healthcare providers are adapting provision of HIV and other healthcare services for adolescents and young people living with and affected by HIV during the COVID-19 pandemic. The study was conducted with a purposive sample of health care workers (HCWs) in the Eastern Cape. Data was collected through semi-structured qualitative interviews with 24 health care workers. Study questions involved descriptions of the HCW experiences during the COVID-19 pandemic, their challenges, and perceptions of adolescents' experiences (specifically adolescents living with HIV) during this time. The interviews were conducted remotely by four research assistants. Interviews were done via phone, and recorded using a recording app. In total, 13 semi-structured interviews were completed and recorded. The process of consent was recorded in addition to the interview itself. Interview lengths averaged 45- 60 minutes. Consent recordings when separate from the interview itself were typically less than 10 minutes.

2. **Cognitive interviews**

Cognitive interviews were conducted to inform the adaptation of the Child and Youth Resilience Measure Revised (CYRM-R), a 17-item quantitative instrument designed to be used with youth aged 10-23 years old during follow-up waves of data collection. Cognitive interviewing involves probing the participants' responses and cognitive approaches to answering questions to determine whether they understand the construct meaning of items, as well as specific phrases within the items. Formalized debrief documents and team discussions will be conducted, capturing details of interviews and assess the interviewer's experience and whether any non- verbal language (such as hesitance, pausing or comfortability) were noted during the interview. The aim of the cognitive interviews is to assess, based on the responses during interviews, what needed to be changed and how the scale needed to be adapted to suit the target population within our study context.

4. DATA COLLECTION AND GENERATION

DATA LIFECYCLE

HEY BABY adolescent questionnaire research data is collected through self-reporting adolescent questionnaires in a combination of Face-to-face activities and remotely -via telephone interview under COVID-19 restrictions.

Data collection

For remote data collection, the study team is provided with Android 9 smartphones (or above) for completing data collection remotely. Quantitative responses from the participants are entered into an eCRF designed with Research Electronic Data Capture (REDCap) installed on a tablet device. The study team are trained in how to use the devices, including security features. An Excel spreadsheet with contact details extracted from baseline responses is available for participant tracking purposes. This document is password protected and stored on the HCF, with restricted access,

which requires two-factor authentication for access. The study team updates this document with new contact details received from participants during follow-up interviews.

Face-to-face interviews may begin in mid-2021 for all Phases of the adolescent questionnaire, if COVID-19 restrictions allow. Data will be collected using the same eCRF on REDCap, handled only by the Research team, and following strict health and safety measures.

A participant tracker in the form of an SQL database is used by the study team to track completion of the different phases of follow-up activities. It uses data collected from the adolescent questionnaire, including refusal and recruitment rates, change in contact numbers, referrals. This database is accessed through user login credentials and these credentials are encrypted with a 3-level encryption algorithm.

HCW study
Prior to the commencement of data collection, the team was trained on qualitative research methods, remote data collection and key aspects of the study protocol. In order to conduct remote data collection, they were provided with the abovementioned electronic data collection and recording devices. Training was also then conducted relating to the devices, data security and confidentiality while working remotely. Sampling was purposive and the contact list used was collated by the RAs and then divided amongst the team. The team used rosters to track and monitor calls, progress and participant details during data collection. The team conducted 'practice-calls' with external team members in order to review the research tool, interview techniques and device set up. The first round of data collection involved in-depth exploration as there was no formal piloting process due to the time sensitive nature of the context. The first three interviews were treated as 'semi-pilots'. This involved 'real-time' translations, intense data checking and analysis. This guided the research approach by highlighting wording issues, challenges in understanding of questions or clarifications that needed addressing. After this group review, interviews continued with each interview being recorded, transcribed and translated for quality checking and analysis.

TYPES OF DATA/DATA OUTPUTS

The HEY BABY Study utilises a mixed methodology to capture the complexity of resilience combining adolescent parent and child data. Qualitative work will include in-depth interviews and focus groups, observations, and participatory research. The quantitative arm of the study will utilise self-reported data from young parents and clinic data from health facilities and observational data of child development assessment.

Types of data collected for this study include consent records (which includes participant name, written consent forms, audio-recorded consent, assent forms – for research involving minors), contact details for research purposes only and for future studies, filed notes, questionnaire answers, audio recordings, video recordings, transcripts of audio/video recordings, photographs, information about the health of the participant (including mental health), and physiological test results/measurements. Data collection follows UK GDPR guidelines and South African Protection of Personal Information Act (POPIA) guidelines.

QUALITY CONTROL/ASSURANCE

Quantitative arm:

Database Design: All data will be entered into a password protected REDCap forms using tablets or desktop computers with log-in and screen saver passwords. Participant serial numbers, date of birth and date of administering the questionnaire will act as the unique identifiers for each form. All data collection forms will be programmed to include range and valid response code checks to reduce invalid entries for any given measure. The forms will also be programmed to include skip patterns to facilitate the speed and accuracy of the data entry process. Only skipped questions due to skip patterns in the questionnaires will be allowed to remain blank during data entry. All other questions will require a data entry code, i.e., hard coding of missing/unanswered questions. This approach helps ensure that missing data in the electronic databases are truly missing, rather than data entry error from skipping entering a response.

The study manager and her designees will enter mock data for at least 10 participants for each questionnaire to validate (i.e. pilot) the programming of the form. After the form has been tested by the Study Manager, it will be shared with the data manager at Oxford University or her designee (CUNY) for one final round of programming validation; this process will include entering mock data for three participants, exporting the data into a statistical software program, printing the data as a report, and validating that the report matches what was captured on the mock CRFs. Following completion

of the first 50 interviews of each form, the Study Manager or her designee will run the first data cleaning check, to check for inconsistencies or response variation that are not logical.

Data capture: All study staff involved data collection will be trained on how to complete the questionnaire effectively in accordance with Good Clinical Practice guidelines. Interviewers' training will include mock interviews and role plays to ensure familiarity and comfortability with the questionnaires prior to implementation in the field. As part of training and quality assurance, the data collected during the first week of implementation will be proctored by senior study staff to help ensure competency and ongoing quality assurance. Trained qualitative interviewers will implement the in-depth interviews. These sessions will be audio-recorded using pseudonyms for the participants and will be transcribed verbatim by native isiXhosa-speakers. Following data entry, the research team will indicate form status as complete, the status of the form is then changed to verified after it's been reviewed twice. Only verified forms are to be sent to the REDCap server. In situations where participants' interview cannot be completed on the same day, RAs will continue interview for the participant on a new form (eCRF) on the next interview contact, following the guidelines provided in the RA interview completion SOP (HEY BABY roster updating and data collection standards).

Ongoing quality control: The team will have weekly debriefing to provide the research team an opportunity to share the challenges they face when collecting the data and provide clarity on study measures. Once a month, the data collection team will be asked to record the entirety of their interviews, starting the third week of each month and completed on the last week of the month. The Senior Project Manager will then randomly listen to two interviews from each team member. The Senior Project Manager will write notes for each interview listened and share these with the research team. These notes may indicate additional training needed or a need for a revision of the research tools.

Data is exported from REDCap and reviewed by the research team monthly. The questionnaire includes primarily close-coded questions; these codes will be entered in the REDCap form as recorded on the paper questionnaire. A STATA do-file has been created with syntax for generating queries from the exported data and these queries are reviewed by the fieldwork team for resolution. Monthly quality control sessions are also held with the fieldwork team, overseen by the Senior Project Manager, to review data quality and support the data cleaning process.

Qualitative arm HCW study:

- 1st round of interviews involved in-depth exploration: Data collection began without a formal 'piloting' process due to the time limits and time sensitive nature of the pandemic. The first 3 interviews were treated as 'semi-pilots'. This involved 'real-time' translations, intense data checking and analysis. This guided the research approach by highlighting wording issues, challenges in understanding questions or clarifications that needed to be corrected.
- Quality checking occurred weekly: Each RA would quality check their own work and an interview of a colleague. This would involve spelling, transcribing quality, translation accuracy and queries, recording quality. The team would then document these queries/ checking processes and meet to discuss them as a group.
- Debrief sessions would occur at the end of any workday in which there had been an interview- this would be to discuss the overall experience of the interview, any issues that occurred- technical or otherwise, and any emotional support the team member may need. Queries relating to the research tool were also addressed in these sessions. Feedback then addressed and implemented where necessary.
- Final data set will be checked for naming conventions, quality of recordings, of transcriptions and translations for final certification and ready for analysis.

Monthly quality control sessions are held with the fieldwork team, overseen by the Senior Project Manager, to review data quality and support the data cleaning process.

FILE FORMATS

Quantitative data – at baseline, collected data was submitted to the ODK server, which was exported as .csv files. For follow-up, data collected with REDCap exports as .xlsx files. Format of the final version of processed data is stored in .dta (Stata). Data is rigorously anonymized before sharing with external users and the public at large.
Qualitative data includes HCW and cognitive interviews. The format of this data includes consent records (which includes participant name, written consent forms, audio-recorded consent, etc) contact details, fieldwork notes, audio recordings, transcripts of audio recordings, and translations. The data is stored as Word and Excel documents, as well as audio recordings.

5. DATA MANAGEMENT, DOCUMENTATION AND CURATION

MANAGING, STORING AND CURATING DATA

During remote data collection, the team were provided with suitcases with keys and requested to keep all paper documents, tech resources/ devices and research related items in these suitcases for safe keeping. During office closures (over December) all equipment and paper documents were retrieved from the RAs and securely stored in the office. All informed consent forms and all contact information linking ID numbers to personal identifying information will be stored in locked files/cabinets. The study activities will ensure that research data is stored, backed-up, managed and curated in a secure manner in line with GDPR and/or national data protection legislation (e.g., POPIA in South Africa), as well as local institution data protection policies. Data management will prioritise data confidentiality, security and safety, which is especially important to protect the identity of the data subjects.

Data curation will involve merging data collected during this study using unique identifiers. Collaborators may also develop methodologies/ research questions that might warrant the merging of HEY BABY datasets. During data merging, the research team will prioritise data confidentiality, security and safety, which is especially important to protect the identity of the data subjects. The merged data will be held in STATA, SPSS, CSV or R formats. Curation will also involve recoding the existing data into derived variables that are more appropriate to the research questions asked, or help to maintain longitudinal consistency when phrasing or answer options of questions change in follow up questionnaires. Derived variables may be, for example, dichotomised versions of the original variable or combinations of multiple variables. The derived versions will be indicated as such in the variable names and labels.

All research data will be securely stored and curated in line with General Data Protection Regulation (GDPR) and/or national data protection legislation and local institution data protection policies (POPIA). The anonymised or de-identified data will be stored on the Hub's secure server housed at the University of Cape Town (UCT). The server is located in a protected room with restricted access on a "need-to-access" basis. The server will be managed by the Hub Data Management Team with the technical support of the eResearch and ICTS team In UCT and the Oxford DSPI IT. The server will also be backed up daily to UCT storage array in the Neotel (Liquid Telecom) data centre in Diep River, Cape Town. Overall curation of research data will be conducted at study site with support from the Hub data management team when needed. The server will be backed up daily to UCT storage array in the Neotel (Liquid Telecom) data centre in Diep River, Cape Town.

METADATA STANDARDS AND DATA DOCUMENTATION

Metadata will include methods used to generate the data (research protocol), characteristics of the study population and provenance of data (using anonymised participant identifiers to maintain confidentiality), descriptions of variables, data dictionary/codebook, data inventory type of analysis used and data created.

Sufficient metadata will be made openly available to render the study data discoverable for effective reuse. The metadata standard to be applied to the study will be appropriate, rich, of high-quality and comply with norms and good practice of the research protocol.

DATA PRESERVATION STRATEGY AND STANDARDS

Primary research data will be kept for 5 years after data collection. Informed consent forms and other personal identification information, including the study rosters and cross-referencing identifier, will be destroyed three years after the final study report has been completed for formal de-identification of the data.

Electronic versions of de-identified data used for data cleaning and analyses are in a format suitable for long term storage. Data will be archived by the two PIs on the UCT server. Data will be available to all investigators on this protocol; sharing with external data users requires written permission from both PIs and IRB approval from institutions as applicable.

All files (primary data, papers, anonymised STATA/SPSS/Excel/qualitative data, personal data) will be stored on the Hub central cloud server system at UCT whilst the project is active. Access will be controlled e.g., to specific partners and specific folders, and will be monitored regularly. Once the project is finished these files will be archived in the Hub central server at UCT.

6. DATA SECURITY AND CONFIDENTIALITY OF POTENTIALLY DISCLOSIVE INFORMATION

SECURITY

For in-person research activities, digital devices – 7-inch tablets – are provided to participants for completing the questionnaire. Trained research assistants will sit with adolescents to demonstrate how to use the tablet and to guide the participant where necessary. The participant will then be offered the opportunity to complete the questionnaire autonomously. The name or contact details of the participant will not be included in the questionnaire – each individual will only be identified through a unique participant ID.

For remote data collection, smartphones - Android 9 or above – and encrypted recording devices are provided to research assistants for completing telephone interviews. Research assistants will be trained on how to use the devices, including security features.

Electronic versions of transcripts, recordings, questionnaires and field notes will be kept in a password-controlled electronic file, or password-protected encrypted folder, accessible only to the researchers. Paper-based data will be stored within the secure offices of the Centre for Social Science Research at UCT once completed. Additional secure storage for data will be set up in the study location – Eastern Cape – for anonymised paper-based data such as patient file data extraction forms. Electronic data from tablet-based data collection will be stored in the secure and encrypted server at UCT.

The highly secure Hub server is located within the data and network infrastructure of the Information Communication Technology Services of the University of Cape Town. The Research data and server infrastructure are stored in a secure, biometric controlled data centre on Upper Campus (UCDC). The data centre has redundant power feeds from separate distribution boards, supported by redundant UPS power. Access to the data centres is via biometric control to designated ICTS staff. The data is replicated once a week for disaster recovery purposes to the storage array in a secure data centre in an off-site location. The data are stored on Dell MD3060 storage arrays hosted in the UCDC. A copy of the data is replicated to UCT storage array in the Neotel data centre in Cape Town. Access to research data will be provided via the UCT VPN using UCT credentials for UCT staff and a third-party access credential for non UCT staff. Research data can be shared and accessed by parties remotely using the UCT credentials.

The UCT's network perimeter and data centres are protected and secured from the rest of the internet by Cisco FirePower Firewall Network Security Appliances. It provides powerful Application Control, IPS, Antivirus, Botnet and DDoS protection, Web Filtering and Messaging Security along with centralized network security management and reporting. All computers and servers connected to the UCT network are managed centrally by the enterprise McAfee ePO antivirus system and Microsoft Windows Server Update Services (WSUS) system for security updates and patches. The McAfee ePO antivirus system deploys daily scheduled antivirus updates to all computers on the network. All windows computers and servers receive scheduled monthly updates via the WSUS system. Apple MAC and Linux based computers are updated via the Apple MAC update server and Linux update services. Access to UCT infrastructures such as websites, File systems, research data shares, servers and desktops are via secure UCT Active Directory credentials as stipulated in the UCT Information Security Policy (see below).

Server routine backup strategy

Backup schedule	**Times**	**Detail**
Daily	20h00	Daily incremental backups
Weekly	20h00	Full back-up once a week

ETHICS AND PRIVACY

With respect to privacy, the study team shall always comply with local data protection regulation(s) and shall perform its obligations or activities as stated in the Oxford – UCT Collaboration and Data Sharing Agreement (CDSA). In all prior studies led by the research team, participant privacy has been of high importance, not only for ethical reasons but also to increase the validity of data collected. With this precedent in mind, attempts have been made to maximise privacy for participants during all interviews.

Face-to-face
Interviews take place out of earshot in a range of locations including community centres, clinics, homes, or outside, under a tree. Given the practical reality of data collection, the most private location has to be determined on a case-by-case basis. However, an essential element in ensuring this, both in-person and remotely, is to train RA's in privacy and confidentiality protocol and ethics. They will also learn strategies for ensuring privacy, including how to explain to

interested family members or friends that this is a private interview, and distraction activities (such as paper and colouring pens) to distract small children and allow increased privacy for parents would be provided.

Remote
Remote interviews are done telephonically and take place during preferred times by the adolescents, young people, and healthcare providers. We understand that most of our participants live in intergenerational households and with large families or may be working in high-stress environments (i.e., healthcare providers). This may limit the amount of time they have available to engage with us. Additionally, there may be limited privacy or space to engage with confidential questions. Our RA's will receive training on sensitively responding to these circumstances, confirming participant privacy and offering to reschedule the interview if this adequate privacy is not feasible for the participant at the time. While there may be a risk of indirect disclosure if the participant's phone is answered for them, this will be mitigated as best as possible by clear questioning at the start of the call confirming participant details and identity.

Whether in-person or remote, capacity for confidentiality and private space to speak may be limited during lockdown. To protect the confidentiality of the participant and protect them from the potential stigma associated with HIV/AIDS and other personal topics, each interview will be conducted when the participant is free and able to find a space that is as private as possible. In addition, the study will be presented within families and communities as investigating health and social services experiences in general.

The major ethical issues relating to the study include ensuring participant confidentiality, preventing stigma and unintended disclosure related to HIV/AIDS and minimising potential risk to participants. In the data collection stages of the research, we will consciously introduce measures to avoid stigma associated with child maltreatment, HIV/AIDS and other adversities experienced by adolescent parents and their families. All research and parenting programmes are presented much more broadly as 'family support' and 'teen support' – an approach that has been very successful in preventing stigma or unintended disclosure of adversities in our prior projects.

Participant confidentiality will be ensured through data anonymization. During the study, personally identifiable data will be stored electronically on password-protected tablets and laptops, in password-protected files, as well as, in aggregate form on highly secure online servers. All data analysis will be conducted using anonymised data sets.

7. DATA SHARING AND OPEN ACCESS

INTELLECTUAL PROPERTY RIGHTS

During the study, data is owned by the PIs. Subject to participants' informed consent, completed and fully anonymised datasets will be made available to the public for secondary use by sharing on public repositories in Africa & UK, following their guidelines, RCUK guidelines.

Public-funded research data are a public good, produced in the public interest and will be made openly available to the maximum extent possible, when possible, in a timely and responsible way. Data will be made available per funder requirements i.e. per ERC, MRC, and NIH guidelines

SUITABILITY FOR SHARING

Data will be shared between research investigators and collaborators as is necessary to generate main study findings. All data will be available for sharing after main study findings have been published and allowing for time to fully clean datasets and compile meta-data. Anonymised data sets will be made publicly available in UK and SA data depositories once the main study findings have been published and subject to specific country regulations. Public-funded research data are a public good, produced in the public interest and will be made openly available to the maximum extent possible, when possible in a timely and responsible way.

DISCOVERABILITY BY POTENTIAL USERS OF THE DATA

Prospective users, policymakers/government agencies/researchers (internal/external) will be required to contact the study team to discuss and plan the use of data. Research data will be available on request subject to participant consent and having completed all necessary documentation.

The data access process is as follows:

1. Internal/External researcher expresses interest to use the data through the PI or Co-PIs.
2. PIs and Co-PIs communicate with the data guardians as well as the data managers.
3. Data managers send the relevant data request access forms to the researcher for completion and indicate the type and number of variables needed.
4. Once complete, this information shall be handed to the guardian of the data to prepare the subset and anonymised data for sharing.
5. The Junior Information Officer to send the Hub Account Request form to the researcher to allow for the creation of the HCF linked account for easily monitored data sharing.
6. Once the account is set, the data guardian shall deposit all the relevant data and documentation in the respective folder of the researcher under the Hub Central Filestore on secure SharePoint site.
7. Processed data shared shall be either in in Stata (.dta) or SPSS (.sav) format.

More information on this process is fully documented in the data access and sharing forms.

GOVERNANCE OF ACCESS

Study PI and Co-PIs to provide oversight of data access. Anonymised version of the data to be made available at the *UCT server* and SharePoint site. External users are bound by the Data Access and Sharing guidelines or the Data Use Undertaking signed before access is granted. In case the guardians or data officers or managers are no longer available, the study PI shall ensure a swift, detailed handover of the process and related information to the current person before the outgoing respective individuals leave.

NIH - NIH expects the timely release and sharing of data to be no later than the acceptance for publication of the main findings from the final dataset. NIH continues to expect that the initial investigators may benefit from first and continuing use but not from prolonged exclusive use.

LICENSING/REGULATION OF RESPONSIBILITIES OF USERS

Prospective users, policymakers/government agencies/researchers (internal/external) will be required to contact the study team to discuss and plan the use of data. Research data will be available on request subject to participant consent and having completed all necessary documentation. Beyond the core research team, other research members only have access to the anonymised version of the dataset. Anonymised data sets will be made publicly available in UK and SA data repositories once the main study findings have been published and subject to specific country regulations.

8. RESPONSIBILITIES

0) Principal investigator lead accountable

1) Study-wide data management: Data Manager, Junior Information Officer, Data Management Team

2) Metadata creation: Study data leads

3) Data security: Study level data guardians

4) Quality assurance of data: Study level data guardians and Study Managers

5) Any other data-related roles.

9. RELEVANT INSTITUTIONAL OR FUNDER DOCUMENTS

Policy	URL or Reference
Data Management Policy and Procedure	http://researchdata.ox.ac.uk/university-of-oxford-policy-on-the-management-of-research-data-and-records/ and http://researchdata.ox.ac.uk/university-of-oxford-policy-on-themanagement- of-research-data-and-records/ and https://www.uct.ac.za/sites/default/files/image_tool/images/328/about/policies/TGO_Policy_Research_Data_Management_2018.pdf
Hub Framework Collaboration agreement	https://gcrfhuboxf.sharepoint.com/:b:/s/HUBCentralFilestore/EQnAs-KVa11MtIXtC2Hg_-8BITEQCy0T3jIbZGUoga4sRg?e=tx1wso
Safeguarding policy	http://www.icts.uct.ac.za/sites/default/files/image_tool/images/286/UCT_Information_Security_Policy_PC03_2020.pdf and http://www.admin.ox.ac.uk/dataprotection/policy/ and http://www.it.ox.ac.uk/infosec/ispolicy/
Data Sharing & Access Policy	
Hub's data sharing guidelines (Terms and Conditions)	https://gcrfhuboxf.sharepoint.com/:w:/s/HUBCentralFilestore/Ee6D5qQekpNEsGcRTe3e3uMBH60eZtMCDAv7zD7Q6_4qfQ?e=gsV9HS
UCT Account and Password Policy	http://www.icts.uct.ac.za/sites/default/files/image_tool/images/286/Password-policy.pdf
POPIA	https://popia.co.za/
GDPR	https://gdpr-info.eu/

10. AMENDMENT AND CHANGE PROCEDURES

DMP AUTHORING AND VIEWING RIGHTS

This DMP will be revisited annually by responsible parties within the study team.