0. Proposal name

Implementation of a Culturally Tailored Decentralization Programme for Snakebite Treatment in Indigenous Communities in the Brazilian Amazonia

1. Description of the data

1.1 Type of study

Data will be collected in a CFIR-based formative intervention adaptation and evaluation (Aim 1), a pragmatic cluster randomized implementation / effectiveness trial (Hybrid Type 2) (Aim 2), and an interrupted time-series and cost-effectiveness quasi-experimental design (Aim 3). AJURI Protocol will be registered at ***ISRCTN Registry*****.**

1.2 Types of data

In Aim 1, we will generate quantitative data from the formative evaluation survey and qualitative data from the co-design sessions. In Aim 2, we will generate quantitative data from the cluster randomized clinical trial and qualitative data from content analysis from interviews from all projected 90 participants. Qualitative data will include implementation outcomes such as acceptability of seeking care, adoption, implementation, and sustainability. In Aim 3, we will have participants' demographics and clinical data from a secondary and public data repository, the SINAN database (part of DATASUS). All data will be deidentified prior to receipt by the repository, but the information needed to generate a global unique identifier for the technologies like the NIMH Data Archive (NDA) will be collected for each subject. Qualitative: co-design sessions, focus groups, and interviews audio recordings and transcribed transcripts. Quantitative: clinical measurements, sociodemographics, content validity survey, IAM, FIM, and AIM tools, and secondary, public data.

1.3 Format and scale of the data

All qualitative data will be audio recordings as MP3 files and transcripts as Word documents in English and Portuguese. We anticipate at least nine co-design sessions of 5-10 participants each (Aim 1); 90 patient interviews (Aim 2); and a maximum of 35 focus groups of 5 participants each (Aim 2). All primary quantitative data will be collected via REDCAP. We anticipate at least 100 responses for the content validity survey and IAM, AIM, and FIM tools (Aim 1) and 192 clinical measurements and sociodemographics (Aim 2). Secondary quantitative data will be downloaded from the public sources described with ~500 patient records per year. Quantitative data will be analyzed using R Language for Statistical Software. Qualitative data will be analyzed with Nvivo. Georeferenced data will be analyzed using R Language for Statistical Computing and ArcGIS. All codes for developing the questionnaires and analysis scripts will be shared using Git, through our GEMINI Gitlabs website. Information can be obtained on our GEMINI Center website. All analytical codes will be commented to ensure reproducibility of the analysis, and version control will also be maintained in our Git channel.

2. Data collection / generation

2.1 Methodologies for data collection/generation

Patient privacy will be maintained by keeping all records in an anonymous fashion. Data collection will occur in a quiet, private area at the care unit. No results from screening or this study will be shared with anyone other than the treating health professional. Aim 1: Qualitative data will be collected from co-design sessions as audio recordings and transcripts, and quantitative data generated from the IAM, AIM, and FIM tools and a content validity survey. Aim 2: Quantitative data will be collected using digital tools, such as redcap,hosted at the FMT-HVD. Aim 3: Quantitative public, secondary data from CNES, IBGE, SINAN, and SINITOX will be collected as well as literature based cost-effectiveness values. All data will be collected following the data standards suggested by the SBE Core Outcomes set, and will be harmonized using either the NINDS Common Data Elements, the NIMH Data Repository, the WHO injury minimal set, MRC Research Data Gateway, or the BioPortal repository for biomedical ontologies. All analysis will be conducted via R Language for Statistical Computing or Python, and codes stored in a safe repository with version-control (git-based).

2.2 Data quality and standards

Quality Assurance will occur at three levels: 1) at the research team level, 2) at the data entry level, and 3) at the data system level. First, at the research assistant level, after data collection, the project coordinator on-site at FMT-HVD will be responsible for performing a quality check daily on data collected. They will mark unfilled, incomplete, or nonsensical data for reevaluation

by the CHC nurse who collected the data. This will occur on a daily basis at the end of each day of data collection. Next, during data entry into our online systems, all data will be evaluated for completeness and appropriateness. Any erroneous data will be updated and we will attempt to identify potential systematic errors or process challenges in order to standardize our process to remove data entry errors. After initial data entry, our project coordinator will check 10% of the participants' data to ensure appropriate and complete data entry. Finally, at the data system level, we will have weekly quality assurance evaluations where we can determine the amount of missing data and find any responses which lie outside a priori established validation parameters or are outside the expected range of valid responses. These data evaluations will occur weekly, and will provide useful feedback to our research team at FMT-HVD to perform rapid quality improvement and protocol changes as needed. Data quality and process improvement checks will be led by Dr. Vissoci and conducted weekly. Quality and improvement sessions will occur in two separate phases: an unblinded discussion in which Dr. Vissoci will discuss unblinded issues, and a blinded discussion in which the entire team will discuss issues and processes during trial development. Advisory Board meetings will happen quarterly to discuss improvements and updates on the trial development. DSMB meetings will happen yearly as the trial unfolds. Automated data quality will be conducted using a data visualization dashboard that will monitor data for completeness, allocation errors, enrollment rates, and follow up rates. Automated emails will be triggered when a data quality issue is identified and a report sent to the study team.

3. Data management, documentation and curation

3.1 Managing, storing and curating data

Data will be entered, managed and curated via REDCAP after de-identification according to the ICO Anonymisation: managing data protection risk code of practice.

3.2 Metadata standards and data documentation

All data collected will be harmonized to open repository metadata standards. Starting with NDA, we will harmonize our data prior to data collection. In case NDA is not sufficient, we will use NINDS repository, HL7 FHIR, Semantic ISO, and others. A data dictionary will include data definitions and standards. Along with the data described above, we will share analytical codes to ensure reproducibility of the analysis.

3.3 Data preservation strategy and standards

We aim to preserve all data listed in the research protocol needed to ensure reproducibility of the results as well as to inform potential secondary analysis.

4. Data security and confidentiality of potentially disclosive information

4.1 Formal information/data security standards

We will use the NDA GUID tool to generate completely de-identified data. No personal information will be shared with NDA. NDA's GUID tool is guided by a Certificate of Confidentiality.

4.2 Main risks to data security

All information collected for this proposal will be kept in a confidential manner separate from any identifying information in order to ensure patient confidentiality and avoid potential data security risks. FDG participants could be concerned about anonymity, even though no identifiers will be collected. It is theoretically possible to identify a respondent based on the recordings. After transcriptions and validation of transcript content, every recording will be destroyed to avoid any data leakage. All paper data logs and records will be kept in an anonymous fashion in a locked drawer or a locked cabinet in the locked FMT-HVD research office. Only the research team members will have keys to the drawers and cabinets in the office. Paper will only be used when digital data collection is not possible. All paper documents will be entered into an online database (REDCap) in a de-identified fashion. This online repository is located behind the institution firewall and limits access to investigators listed on the regulatory paperwork. Only the PI and Analysis leader will have 'download' rights to further protect access to this data. Audio recordings, transcriptions, and translations of the interventions will be located in an online data repository (Box) where investigators with regulatory clearance can be added to access these materials. Both of these electronic data repository systems are approved by all regulatory bodies.

5. Data sharing and access

5.1 Suitability for sharing

All data will be collected following the data standards suggested by the SBE Core Outcomes set, and will be harmonized using either the NINDS Common Data Elements, the NIMH Data Repository, the WHO injury minimal set, MRC Research Data Gateway, or the BioPortal repository for biomedical ontologies. All analysis will be conducted via R Language for Statistical Computing or Python, and codes stored in a safe repository with version-control (git-based).

5.2 Discovery by potential users of the research data

Data will be findable for the research community through the AJURI Repository, following the procedures established by similar repositories such as the NDA/NINDS CDE Collection. The AJURI repository will be established during the first year after the application is funded. For all publications, an open repository study ID (such as Figshare) will be created. Each of those studies is assigned a digital object identifier (DOI). This data DOI will be referenced in the publication to allow the research community easy access to the exact data used in the publication.

5.3 Governance of access

To request access to the data, researchers will use the standard processes similar to the ones applied at NDA/NINDS CDE and any additional processes requested by the Brazilian regulations, and the Data Access Committee will decide which requests to grant The Data Access Committee will be composed by members of the research group and the community. The standard data access process allows access for one year and is renewable.

5.4 The study team's exclusive use of the data

All data will be deposited to AJURI Repository starting 12 months after the award begins and will be deposited every six months thereafter following the usual NDA/NINDS CDE data submission dates. The research community will have access to data when the award ends. Studies will also be created that contain the data used for every publication. Those studies will be shared when the pre-print is available. NDA studies have digital object identifiers (DOI) to aid in findability. We will include that DOI in relevant publications. NDA will make decisions about how long to preserve the data, but DNA has not deleted any deposited data up to now.

5.5 Restrictions or delays to sharing, with planned actions to limit such restrictions

All research participants will be consented for broad data sharing. Since this data is being collected in Brazil, we will also pursue a data sharing agreement that allows for broad sharing in compliance with Brazilian regulations. For secondary data, it is unlikely we will experience data access challenges given we have previously obtained the publicly available DATASUS data to be used and conducted preliminary analyses. We have conducted distance and coverage determination for other topics with DATASUS data demonstrating feasibility for this project.

5.6 Regulation of responsibilities of users

Data to be shared will be approved by the regulatory authorities in Brazil, US, and UK, and specified in the data sharing agreements. We will share qualitative data in the form of de-identified codes and coding trees to support secondary evaluation of qualitative data.

6. Responsibilities

In addition to Dr. Monteiro, the Data and Analytics team, led by Dr. Vissoci, will be responsible for study-wide data management, metadata creation, data security, and quality assurance.

7. Relevant institutional, departmental or study policies on data sharing and data security

Data Management Policy & Procedures	*ICO Anonymization; NDA Guid Tool*
Data Security Policy	*Proteção de Dados Pessoais (LGPD) (Brazilian General Law of Data Protection)*
Data Sharing Policy	*NDA/NINDS CDE*
Institutional Information Policy	
Other:	*ISRCTN Registry*

8. Authors of this Data Management Plan (Name) and, if different to that of the Principal Investigator, their **telephone & email contact details**
Dr. Wuelton Monteiro