Plan Overview

A Data Management Plan created using DMP Tool

DMP ID: https://doi.org/10.48321/D12BA8af71

Title: Increasing Trial Enrollment of Rural Residents in ARDS Clinical Trials

Creator: Dustin Krutsinger - **ORCID:** 0000-0001-7362-7304

Affiliation: University of Nebraska Medical Center (unmc.edu)

Principal Investigator: Dustin Krutsinger

Funder: National Institutes of Health (nih.gov)

Funding opportunity number: RFA-HL-22-011

Template: NIH-Default DMSP

Project abstract:

A major limitation of evidence-based medicine is that patients who carry the greatest burdens of illness are often systematically underrepresented in randomized clinical trials (RCTs). The common and often deadly Acute Respiratory Distress Syndrome (ARDS) exemplifies how this lack of representation affects patients. ARDS outcomes are worse among marginalized groups, including rural residents, who face inordinate structural barriers to healthcare (and therefore research) access. Many leading government and research organizations have identified rural residents as a priority population. Yet, critical knowledge gaps for rural healthcare remain, including identifying mechanisms for inequitable research representation and developing interventions to increase rural participants in ARDS RCTs. The proposed study's broad objective is to identify inequitable ARDS trial participation mechanisms and pilot a novel intervention to increase rural patient enrollment. The proposal builds upon the candidate's previous work focused on improving trial enrollment. The candidate has conducted both web-based and in-person trials of interventions aimed at improving RCT enrollment. He also helped design the protocol for the RETAIN study, which showed the effectiveness and ethics of financial incentives for trial participation. He led qualitative work that identified barriers and facilitators of ICU-based RCT enrollment. The proposed study builds upon these findings and seeks to 1) conduct a network analysis of the existing critical care transfer network to identify where rural residents with ARDS receive care, 2) identify mechanisms of and interventions to improve the underrepresentation of

rural residents in ARDS trials, and 3) identify the most promising implementation strategies to increase enrollment into ARDS trials among rural residents. The identification of the most promising intervention will be guided by the Consolidated Framework for Implementation Research (CFIR) and will be achieved through stakeholder engagement and implementation mapping methods. Completion of this research will build upon the candidate's past training, which includes a Masters of Science in Clinical Epidemiology obtained with NHLBI T32 support. It will provide the experience, education, and mentorship to allow the candidate to become a fully independent investigator. The candidate's rigorous training plan focused on obtaining advanced skills in implementation science, network analysis, and community-based participatory research methods, will allow him to submit a successful R01 or equivalent applications testing interventions developed directly from the results of this work. The candidate's primary mentor, co-mentors, collaborators, and advisors will ensure adherence to the proposed timeline and goals. Along with the candidate's strong institutional support, this proposal will be fundamental in the candidate's transition to becoming an independent acute respiratory failure trialist and an expert in the efficient, ethical, and equitable conduct of clinical trials.

Start date: 12-01-2024

End date: 12-01-2029

Last modified: 07-08-2024

Copyright information:

Increasing Trial Enrollment of Rural Residents in ARDS Clinical Trials

Data Type

Types and amount of scientific data expected to be generated in the project: *Summarize the types and estimated amount of scientific data expected to be generated in the project.*

Describe data in general terms that address the type and amount/size of scientific data expected to be collected and used in the project (e.g., 256-channel EEG data and fMRI images from ~50 research participants). Descriptions may indicate the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing that has occurred (i.e., how raw or processed the data will be)

This project will produce three distinct sets of data. Dataset A consists of a network analysis administrative databases and publicly available data collected from clinicaltrials.gov. Dataset B will consist of qualitative data collected using focus group interviews. Audio recordings will be taken of the interviews, and transcriptions generated from the audio recordings. Dataset C consist of qualitative data collected during stakeholder taskforce implementation mapping meetings. Audio recordings will be taken of the meetings, and transcriptions generated from the audio recordings.This project is estimated to generate 250 Gigabytes of data. Raw data will be transformed by masking methods to make all publicly available data de-identified. To protect research participants' identities, de-identified data will be made available for sharing.

Scientific data that will be preserved and shared, and the rationale for doing so: *Describe which scientific data from the project will be preserved and shared and provide the rationale for this decision.*

Data will be de-identified, creating an anonymous dataset for future research, and made available for re-use. We will preserve and share each of the three datasets described above with the generalist repository Dryad. Dryad allows for datasets to be archived and preserved on its servers indefinitely. Due to ethics and privacy issues associated with the personal, sensitive nature of the qualitative data, all identifiable data will not be shared with users outside of the research team. De-identified, summarized data will be shared in a structured format using comma separated value (.csv files) datasets.

Metadata, other relevant data, and associated documentation: Briefly list the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data.

To facilitate interpretation and subsequent analyses of our scientific data, project-specific metadata (e.g., variable descriptors, data source types, unit of observation) and corresponding data dictionaries will accompany the final analytic dataset(s) along with supporting documentation, including all survey guides used in our focus groups. Each transcript (both clean and coded and preserved as a TEXT file) and audio recording (preserved in .wav files, used and processed in .mp3 files) will be uploaded to the participant study record in REDCap. The final codebook, analytic memos, and field notes will be uploaded to the File Repository in REDCap, then transferred into the generalist repository Dryad for reuse. Audio recordings will not be shared publicly due to the possibility they will include protected

private health information. We will use Dryad's internal metadata schema and provide the necessary information for optimal Findability, Accessibility, Interoperability, and Re-use.

Related Tools, Software and/or Code

State whether specialized tools, software, and/or code are needed to access or manipulate shared scientific data, and if so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.

REDCap has the capability to export the collected data to comma separated value (.csv) files and Stata files, the software programs that will be used in this study for analysis. Transcription of the audio interviews will be completed by using software for audio segmenting, labeling, and transcription software called NVivo. All transcribed data will be translated into .csv files for re-usability. Transcripts will be uploaded to qualitative software on the local system for analysis. During the qualitative phase of our analysis, a codebook will be developed. All datasets from the qualitative research portion will be available in Dryad in .csv format.

Standards

State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources, and provide the name(s) of the data standards that will be applied and describe how these data standards will be applied to the scientific data generated by the research proposed in this project. If applicable, indicate that no consensus standards exist

Dryad uses the DataCite Schema for optimal data searchability. Our project will follow the DataCite Schema, along with a structured and detailed versioning system, to keep records updated, accurate, and flexible.

Versioning of files will be controlled through the following naming system:

[data type]_[data name]_[YYYY-MM-DD]_v[xx.yy]

where [data type] describes the ExportData, Protocol, ICF, DataDictionary, InterviewGuide, CaseBook, etc. where [data name] describes the file such as POP1, POP2, Case4Caring1, Case4Science2,Qual1, Quant2, etc. where YYYY-MM-DD is the four digit year, two digit month, and two digit day the file is created

where 'xx' is the sequential number of major revision and 'yy' is the number of times the file is modified.

Data Preservation, Access, and Associated Timelines

Repository where scientific data and metadata will be archived: Provide the name of the repository(ies) where scientific data and metadata arising from the project will be archived; see Selecting a Data Repository)

All data sets that can be shared will be deposited in Dryad, which provides metadata, persistent identifiers (i.e., DOIs), and long-term access. Dryad shares data under a CC0 waiver, which makes the dataset(s) publicly available. Data will be made available as soon as possible or at the time of associated publication. Dryad datasets are backed up to Merritt, the UC's CoreTrustSeal-certified digital repository, for long-term storage and accessibility.

How scientific data will be findable and identifiable: Describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.

Dryad provides metadata, persistent identifiers (DOI), and long-term access. This repository shares data under a CC0 waiver, which makes the dataset(s) publicly available.

When and how long the scientific data will be made available: Describe when the scientific data will be made available to other users (i.e., no later than time of an associated publication or end of the performance period, whichever comes first) and for how long data will be available.

Data will be made available as soon as possible or at the time of associated publication. Datasets will be archived and preserved on Dryad servers indefinitely.

Access, Distribution, or Reuse Considerations

Factors affecting subsequent access, distribution, or reuse of scientific data: NIH expects that in drafting Plans, researchers maximize the appropriate sharing of scientific data. Describe and justify any applicable factors or data use limitations affecting subsequent access, distribution, or reuse of scientific data related to informed consent, privacy and confidentiality protections, and any other considerations that may limit the extent of data sharing. See Frequently Asked Questions for examples of justifiable reasons for limiting sharing of data.

The primary investigator and research team will have exclusive use of the data until publication. Following publication secondary use of the de-identified research data is possible through the data repository without restrictions.

Whether access to scientific data will be controlled: State whether access to the scientific data will be controlled (i.e., made available by a data repository only after approval).

Our de-identified dataset(s) will be made available for public use via direct download from Dryad, where usage will be free of charge to the public. Users who download the data are required to comply with Dryad's terms of service. We do not anticipate additional controls on access as the data will be fully de-identified, thus, the risk of disclosing confidential information of our research participants is minimal.

Protections for privacy, rights, and confidentiality of human research participants: If generating scientific data derived from humans, describe how the privacy, rights, and confidentiality of human research participants will be protected (e.g., through de-identification, Certificates of Confidentiality, and other protective measures).

In order to ensure participant consent for data sharing, IRB paperwork and informed consent documents will include language describing plans for data management and sharing of data, describing the motivation for sharing, and explaining that personal identifying information will be removed. Our data management and sharing plan will be submitted to the University of Nebraska Medical Center Institutional Review Board for review and final approval. Identification will be completed prior to the finalization of the public use data files.

Oversight of Data Management and Sharing

Describe how compliance with this Plan will be monitored and managed, frequency of oversight, and by whom at your institution (e.g., titles, roles).

The principal investigator (PI), Dr. Dustin Krutsinger, will be responsible for the day-to-day oversight of the data management activities and data sharing. Broader issues of DMS Plan compliance oversight and reporting will be handled by the PI and mentorship team as part of the general University of Nebraska Medical Center's stewardship, reporting, and compliance processes.
