

Plan Overview

A Data Management Plan created using DMPTool

DMP ID: <https://doi.org/10.48321/D1KK84>

Title: Bacterial Pneumonias and Lung Injury: The Myeloid Innate Immune Response in the Lungs

Creator: Claire Doerschuk - **ORCID:** [0000-0003-2638-3321](https://orcid.org/0000-0003-2638-3321)

Affiliation: University of North Carolina School of Medicine (med.unc.edu)

Funder: National Institutes of Health (nih.gov)

Funding opportunity number: PA-20-185

Template: NIH-Default DMSP

Project abstract:

Bacterial pathogens cause pneumonia and lung injury, and *S. pneumoniae* is an important and common cause, particularly of community-acquired pneumonia as well as nosocomial pneumonia and bacterial pneumonia secondary to viral infections. The immune response of the lungs and the functions of myeloid cells (neutrophils, monocytes and macrophages) in the lungs are critical to clearing the bacterial infection and resolving the injury. Using forefront technologies, our studies investigate the mechanisms of the immune response in the lungs in response to *S. pneumoniae*, including the heterogeneity in gene expression over time in neutrophils and lung macrophages, the biological importance of the subsets, and how the heterogeneous response is generated, offering insight into therapeutic opportunities targeting specific myeloid cell subsets not yet appreciated.

Start date: 10-01-2023

Last modified: 03-05-2023

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 customize 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

Bacterial Pneumonias and Lung Injury: The Myeloid Innate Immune Response in the Lungs

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)

In this proposed project, data will be generated via the following methods: flow cytometry, sequencing, ELISAs, assays of cell function. The data will come primarily from in vivo studies using mice. The data will be primarily in a tabular format. The protocols vary with the experiment. We estimate that the total size of data collected will be 1000 GB.

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.*

In this proposed project, the cleaned, item-level spreadsheet data for all variables will be shared openly, along with example quantifications and transformations from initial raw data. Final files used to generate specific analyses to answer the Specific Aims and related results will also be shared. The rationale for sharing only cleaned data is to foster ease of data reuse.

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 the interpretation and reuse of the data, a README file and data dictionary will be generated and deposited into a repository along with all shared datasets. The README file will include method description, instrument settings, RRIDs of resources such as antibodies, model organisms, and other tools. The data dictionary will define and describe all variables in the dataset.

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.

No specialized tools, software, and/or code are needed to access or manipulate shared scientific data.

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

Formal standards for have not yet been widely adopted. However, our data and other materials will be structured and described according to best practices.

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](#))

Gene expression data will be stored at GEO. Any other data will be stored on UNC's institutional repository.

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.

GEO and my institutional repository provide searchable study-level metadata for dataset discovery. Data will be discoverable online through standard web search of the study-level metadata as well as the persistent pointer from the DOI to the dataset or through contact with the investigator.

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.

All scientific data generated from this project will be made available as soon as possible, and no later than the time of publication or the end of the funding period, whichever comes last. The duration of preservation and sharing of the data will be a minimum of 10 years after the funding period.

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.

There are no anticipated factors or limitations that will affect the access, distribution or reuse of the scientific data generated by the proposal.

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).

Controlled access will not be used. The data that is shared will be shared by unrestricted download.

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).

Not applicable.

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).

Lead PI Claire Doerschuk, ORCID: 0000-003-2638-3321, will be responsible for the day-to-day oversight of lab/team data management activities and data sharing. Broader issues of DMS Plan compliance oversight and reporting will be handled by the PI and Co-I team as part of general campus stewardship, reporting, and compliance processes.
