

Plan Overview

A Data Management Plan created using DMP Tool

DMP ID: <https://doi.org/10.48321/D13K8S>

Title: PERSIST study

Creator: Emily Ricotta - **ORCID:** [0000-0001-9928-8275](https://orcid.org/0000-0001-9928-8275)

Affiliation: National Institutes of Health (nih.gov)

Principal Investigator: Emily Ricotta

Data Manager: Emily Ricotta

Funder: National Institutes of Health (nih.gov)

Grant: AI001335

Template: NIH-GEN DMSP (Forthcoming 2023)

Project abstract:

This prospective cohort study will assess the pre- and post-vaccination immune responses in individuals with select immunodeficiencies and immune dysregulations compared to healthy volunteers who receive a coronavirus disease 2019 (COVID-19) vaccine, as well as any adverse events (AEs) experienced after vaccination. All required study visits for this protocol may be conducted remotely; in-person visits at the NIH are optional. Subjects who have not yet been vaccinated will undergo baseline blood sampling using finger stick microsamplers and/or venous blood draw within 7 days prior to receiving the vaccine. Additional samples will be requested from participants approximately 14-21 days after dose 1 and 21-28 days after dose 2 (if applicable). Optional samples may be collected at 6, 12, and 24 months post-vaccination. If subsequent booster doses are received while a participant is still on study, blood samples will again be requested approximately 28 days after each booster dose, through the 5th COVID-19 vaccine dose received, and then participants may proceed with the optional 6-, 12-, and 24-month follow-up sample collection. Participants who are able to attend in-person visits at NIH will have optional on-site blood draws 1 and 3 days after doses 1 and 2 (as applicable). Research evaluations will include baseline severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) antibody titers to the spike (S), nucleocapsid (N), and receptor binding domain (RBD) proteins, to assess pre-vaccination SARS-CoV-2 exposure and evaluate responses to vaccination. Additional immune markers of interest may include presence of autoantibodies, transcriptomic profiling, T-cell receptor (TCR) repertoire, among others. Participants who only submit finger stick home microsamplers at the timepoints listed above will be evaluated for SARS-CoV-2 antibody titers and autoantibodies only. All subjects will be asked at baseline about prior COVID-19 diagnosis, symptoms, and severity, and will be asked additional questions at follow-up timepoints (including after additional booster doses) about vaccine AEs using standardized questionnaires.

Start date: 04-01-2021

End date: 06-30-2025

Last modified: 01-07-2025

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

PERSIST study

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)

Data types generated

- Demographic/patient-level data
 - A. Age, gender, state of residence
 - B. Healthy or immune disorder
 - C. Medications/medical history
- Survey data
 - A. Medical history
 - B. COVID-19 illness symptom reporting
 - C. COVID-19 vaccine-related adverse events
 - D. Medical change questionnaire
 - E. COVID-19 preventive behaviors
- Laboratory findings
 - A. ELISA optical density and titer concentration (COVID-19 Spike and RBD IgG, IgM, IgA titers; Nucleocapsid IgG presence/absence)
 - B. Mesoscale Discovery V-PLEX Panels 3 and 23 (COVID-19 Spike IgG, IgM, IgA concentration: Wuhan, B.1.1.529, AY.4.2, AY.4, P.1, B.1.1.7, B.1.351, B.1.617.2; ACE2 percent binding inhibition : Wuhan, B.1.1.529, AY.4.2, AY.4, P.1, B.1.1.7, B.1.351, B.1.617.2; Seasonal respiratory virus IgG concentrations: HCoV-229E Spike, Flu A/Hong Kong H3, Flu A/Michigan H1, Flu A/Shanghai H7, Flu B/Brisbane HA, Flu B/Phuket HA, HCoV-HKU1 Spike, HCoV-OC43 Spike, HCoV-NL63 Spike, RSV Pre-Fusion F, SARS-CoV-1 Spike, SARS-CoV-2 N, MERS-CoV Spike, SARS-CoV-2 S1 RBD) collected per protocol, up to 2 years post-dose 5
 - C. T cell receptor repertoire DNA sequences
 - D. Breakthrough infection counts, RBD sequences

Amount of data generated

- Demographic data is provided by all enrolled participants (enrollment ceiling 600).
- Survey data
 - A) Collected once at baseline
 - B) Collected once at baseline, then after each reported COVID-19 illness episode
 - C) Collected after each vaccination (up to 5 times)
 - D) Collected after each vaccination (up to 5 times)
 - E) Collected once per participant
- Laboratory findings
 - A) ELISA data collected per protocol, up to two years post-dose 5
 - B) Mesoscale Discovery *data* collected per protocol, up to two years post-dose 5
 - C) T cell receptor repertoire collected from Spike IgG non-responders, hyper-responders (relative to healthy controls), and age-matched healthy controls, up to two years post-dose 5
 - D) Breakthrough infections and sequences collected per protocol, up to 6 months post-dose 5

File types generated

- Demographic data is stored in CRIMSON or in .csv (comma separated values) or .xlsx (Excel) files in Box
- Survey data is stored in REDCap or in .csv files in Box
- Raw ELISA and Mesoscale Discovery data are stored as .xlsx files. After data processing they are stored as RData files
- T cell receptor data will be received as .tsv (tab-separated values)
- Breakthrough infection incidence is stored in CRIMSON or as .csv in Box. Sequences are stored in raw format in FASTA files and variant calls in .csv files in Box.

Raw data will be transformed using the appropriate software tools (e.g. R, interARTIC) and the subsequent processed dataset used for statistical analysis. To protect research participant identities, de-identified data will be made available for sharing with certain patient-level variables anonymized through data masking, perturbation, or generalization as appropriate.

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.*

Deidentified data will be deposited at ImmPort (<https://www.immport.org>) under study accession SDY1234. All laboratory findings will include unique, anonymized patient-level identifiers, patient age category, and immune disorder group (as defined in protocol).

High-quality (>90% coverage) COVID-19 genetic sequences will be deposited into the Global Initiative on Sharing All Influenza Data (GISAID) repository. Lower coverage sequences will be deposited into Zenodo or another general-purpose open data repository.

Raw data will be retained in secure, NIH-approved storage platforms including CRIS (medical records), CRIMSON (patient demographics, medications, etc.), REDCap (participant surveys), and BSI systems (biospecimens). Data for analysis will be stored in NIH Box. This data contains PII and PHI and will not be shared outside of NIH. Per the informed consent document, de-identified raw data may be shared internally for research purposes as deemed appropriate by the principal investigator.

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 of the data, the following metadata will be shared and associated with the relevant datasets:

Study protocol and informed consent documents will be posted to ClinicalTrials.gov (NCT04852276)

Survey tools and code can be found at [DOI: 10.5281/zenodo.7449282](https://doi.org/10.5281/zenodo.7449282)

Metadata for data shared to ImmPort includes:

- Basic study, Subjects (Human), Clinical lab test, Intervention, ELISA Assays, Neutralizing Antibody Titer Assays, Other Assays. See <https://www.immport.org/resources/dataTemplates> for details
- Assay standards used

Metadata for shared genetic sequences will be completed in accordance with GISAID metadata requirements. Additional metadata requirements for Zenodo (DataCite Metadata Schema) will also be completed.

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.

Relevant documentation (including survey instruments), code, and simulated data are shared under an open-source license on Github, which can be found at [DOI: 10.5281/zenodo.7449282](https://doi.org/10.5281/zenodo.7449282).

T cell receptor repertoire sequences were analyzed using Adaptive Biotechnologies ImmunoSEQ Analyzer.

COVID-19 raw sequence data were processed using interARTIC to generate consensus sequences and variant calls. SARS-CoV-2 lineages were determined using these consensus sequences and the NextClade and Pangolin platforms.

All other data will be analyzed using R and RStudio.

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

Whenever possible, we will use the following data standards to structure and organize our data:

- Clinical data:
 - Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) or other recognized international standard as deemed appropriate by the data repository
- Laboratory data:
 - Results
 - ImmPort data standards, see <https://www.immport.org/resources/dataTemplates>
 - GISAID data standards
 - T cell receptor repertoire: identification of specific V, D, and J genes is carried out according to the definitions established by the International ImMunoGeneTics (IMGT®) collaboration
 - Experimental
 - ELISA: human chimeric anti-RBD IgG, IgM, IgA (Gen-Script, Cat# A02038, A02046, A02071)
 - Mesoscale Discovery: National Institute for Standards and Biological Control anti-SARS-CoV-2 Ab 20/130 (Spike, RBD), 20/124 (Nucleocapsid)
 - T cell receptor repertoire: Adaptive Biotechnologies synthetic immune repertoire representing all VJ-gene combinations
 - COVID-19 sequences: Tiled amplicon libraries: Midnight panel and Rapid barcoding kit RBK-004 (Oxford Nanopore technologies); SARS-CoV-2 sequences: interARTIC consensus sequences, NextClade and Pangolin platforms for variant calls

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 dataset(s) that can be shared will be deposited in ImmPort as appropriate (ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. Sci Data. 2018 Feb 27;5:180015. doi: 10.1038/sdata.2018.15.). Data can be found using study accession study_1285.

Sequences will be deposited in GISAID (GISAID: from vision to reality. EuroSurveillance, 22(13) doi: 10.2807/1560-7917.ES.2017.22.13.30494) or Zenodo (10.5281/zenodo.2842909)

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.

ImmPort provides metadata and assigns a digital object identifier to all data shared in the repository and enables long-term data access. This repository is supported by the National Institute of Allergy and Infectious Diseases and datasets are available under a limited license, and the data provider maintains ownership of all deposited data.
<https://docs.dev.immport.org/home/agreement/>

GISAID provides metadata and each record receives a unique and permanent identifier. This data is available long-term. While Data in GISAID are publicly accessible, Submitters do not forfeit their rights (IPR) to the Data they deposit in GISAID. All rights are explicitly preserved and may not be altered under the license provided through GISAID's Terms of Use.
<https://gisaid.org/terms-of-use/>

Zenodo provides metadata and assigns a digital object identifier to all data shared in the repository and enables long-term data access. Zenodo is funded by the European Commission through CERN and OpenAIRE. Users must specify a license for all publicly available files. Licenses for closed access files may be specified in the description field. All meta data is openly available under CC0 license. By uploading content, no change of ownership is implied and no property rights are transferred to CERN. All uploaded content remains the property of the parties prior to submission. <https://about.zenodo.org/policies/>

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.

Scientific data will be made available no later than the time of publication, when possible.

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.

All finalized data will be shared in an uncontrolled fashion once deposited in the appropriate repository. To protect research participant identities, de-identified data will be made available for sharing with certain patient-level variables anonymized through data masking, perturbation, or generalization as appropriate.

Raw data contains personally identifiable information and protected health information and will not be shared outside of NIH under any circumstances, in accordance with ethical and legal considerations. Per the informed consent document, de-identified raw data may be shared internally for research purposes as deemed appropriate by the principal investigator.

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).

Access to ImmPort research and clinical data is available to any user after a brief registration and approval process. Users must accept a data sharing and access agreement before approval is granted.

All users of GISAID databases are issued personal access credentials after having provided their identity and agreed to terms of use that govern the GISAID sharing mechanism. This requirement is not only essential to help uphold the integrity of the GISAID user community, but necessary to enforce the GISAID sharing mechanism that assures reciprocity of the data for future generations

Files may be deposited in Zenodo under closed, open, or embargoed access. Files deposited under closed access are protected against unauthorized access at all levels. Access to metadata and data files is provided over standard protocols such as HTTP and OAI-PMH. Users may deposit restricted files with the ability to share access with others if certain requirements are met. These files will not be made publicly available and sharing will be made possible only by the approval of depositor of the original file.

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 data, describing the motivation for sharing, and explaining that personal identifying information will be removed.

All records will be kept confidential to the extent provided by federal, state, and local law. Authorized representatives of NIAID may inspect all documents and records required to be maintained by the investigator, including but not limited to, medical records. Records will be kept locked and data will be coded. Any personally identifiable information maintained for this study will be kept on restricted-access computers and networks. Personally identifiable information will only be shared with individuals authorized to receive it under this protocol. Individuals not authorized to receive personally identifiable information will be provided with coded information only, as needed. Clinical information will not be released without written permission of the participant, except as necessary for monitoring by the IRB, NIAID, and the Office for Human Research Protections (OHRP).

To further protect the privacy of study participants, a Certificate of Confidentiality has been issued by the NIH. This certificate protects identifiable research information from forced disclosure. It allows the investigator and others who have access to research records to refuse to disclose identifying information on research participation in any civil, criminal, administrative, legislative, or other proceeding, whether at the federal, state, or local level. By protecting researchers and institutions from being compelled to disclose information that would identify research participants, Certificates of Confidentiality help achieve the research objectives and promote participation in studies by helping assure confidentiality and privacy to participants.

Coded specimens and data will be stored at the NIH indefinitely for future research after the study is complete. Human genetic testing may be performed. Plans for future use of specimens and data is described in the informed consent document. Only investigators or their designees will have access to the samples and data. Other investigators (at NIH and elsewhere) may wish to study these specimens and data. If the planned research falls within the category of “human subjects research” on the part of the investigators, NIH IRB review and approval will be obtained. This includes the investigators sending out coded and linked specimens or data and getting results that they can link back to their participants.

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 following individuals will be responsible for data collection, management, storage, retention, and dissemination of project data, including updating and revising the Data Management and Sharing Plan when necessary:

- Emily Ricotta, Principal Investigator, Division of Intramural Research, NIAID, emily.ricotta@nih.gov
- Anita Ginigeme, Study Coordinator, Division of Intramural Research, NIAID, anita.ginigeme@nih.gov
- Mackenzie Zendt, Epidemiologist, Division of Intramural Research, NIAID, mackenzie.zendt@nih.gov
- Fausto Bustos, Data Science Fellow, Office of Data Science and Emerging Technologies, NIAID, mackenzie.zendt@nih.gov
- Heather Kalish, Chemist, Trans-NIH Shared Resource on Biomedical Engineering and Physical Science, NIBIB, heather.kalish@nih.gov
- Neeltje van Doremalen, Associate Staff Scientist, Division of Intramural Research, NIAID, neeltje.vandoremalen@nih.gov
- Pavel Khil, Genomics Scientist, National Institutes of Health Clinical Center, pavel.khil2@nih.gov

This plan will be monitored annually to ensure accuracy and compliance. Review will be conducted by members of the study team and members of the Office of Data Science and Emerging Technologies, if necessary.

Planned Research Outputs

Dataset - "Characterization of the anti-spike IgG immune response to COVID-19 vaccines in people with a wide variety of immunodeficiencies"

This data was collected as part of the PERSIST cohort study (NCT04852276), which aimed to assess the immune response to COVID-19 vaccines in people with immune disorders and healthy volunteers. We collected and analyzed data from April 2021 through April 2022 for this study. Laboratory methods that describe the assay specifics can be found in the associated publication, "Characterization of the anti-spike IgG immune response to COVID-19 vaccines in people with a wide variety of immunodeficiencies" in Science Advances.

DOI: <https://doi.org/10.5061/dryad.6hdr7sr68>

Software - "ericotta/PERSIST_Study: PERSIST study metadata v2.1"

The R scripts can be found at <https://doi.org/10.5281/zenodo.8428160>. Note that all errors or warnings elicited by running the R code can be safely ignored. Most occur when R is asked to statistically compare two groups in a multi-panel plot and in at least one panel, such a comparison is not possible. For example, it is not possible to compare the IgG concentrations of healthy volunteers to participants with immunodeficiencies at the 6 months post-dose 3 timepoint because there were no healthy volunteer samples at this time point.

Data paper - "Characterization of the antispike IgG immune response to COVID-19 vaccines in people with a wide variety of immunodeficiencies"

Research on coronavirus disease 2019 vaccination in immune-deficient/disordered people (IDP) has focused on cancer and organ transplantation populations. In a prospective cohort of 195 IDP and 35 healthy volunteers (HV), antispoke immunoglobulin G (IgG) was detected in 88% of IDP after dose 2, increasing to 93% by 6 months after dose 3. Despite high seroconversion, median IgG levels for IDP never surpassed one-third that of HV. IgG binding to Omicron BA.1 was lowest among variants. Angiotensin-converting enzyme 2 pseudo-neutralization only modestly correlated with antispoke IgG concentration. IgG levels were not significantly altered by receipt of different messenger RNA-based vaccines, immunomodulating treatments, and prior severe acute respiratory syndrome coronavirus 2 infections. While our data show that three doses of coronavirus disease 2019 vaccinations induce antispoke IgG in most IDP, additional doses are needed to increase protection. Because of the notably reduced IgG response to Omicron BA.1, the efficacy of additional vaccinations, including bivalent vaccines, should be studied in this population.

DOI: [10.1126/sciadv.adh3150](https://doi.org/10.1126/sciadv.adh3150)

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Characterization of the anti-spike IgG immune resp ...	Dataset	2023-10-09	Open	DRYAD	14 MB	Creative Commons Zero v1.0 Universal	Dryad Metadata Application Profile	No	No
ericotta/PERSIST_Study: PERSIST study metadata v2. ...	Software	2023-10-09	Open	Zenodo	1 MB	BSD 3-Clause "New" or "Revised" License	None specified	No	No
Characterization of the antispoke IgG immune respo ...	Data paper	Unspecified	Open	None specified		None specified	None specified	No	No

Related Works

Articles

- Zendt, Mackenzie, Fausto A. Bustos Carrillo, Sophie Kelly, Taylor Saturday, Maureen DeGrange, Anita Ginigeme, Lurline Wu, et al. 2023. "Characterization of the Antispike IgG Immune Response to COVID-19 Vaccines in People with a Wide Variety of Immunodeficiencies." [Article]. Science Advances 9 (41). <https://doi.org/10.1126/sciadv.adh3150>.

Datasets

- Ricotta, Emily. 2023. "Characterization of the Anti-Spike IgG Immune Response to COVID-19 Vaccines in People with a Wide Variety of Immunodeficiencies." [Article]. Dryad. <https://doi.org/10.5061/dryad.6hdr7sr68>.

Preprints

- Zendt, Mackenzie, Fausto Andres Bustos Carrillo, Viviane Callier, Maureen DeGrange, Anita Gingimeme, Lurline Wu, Bevin Manuelpillai, Ana Ortega-Villa, and Emily Ricotta. 2024. "Breakthrough SARS-CoV-2 Outcomes in Immune-Disordered People During the Omicron Era: A Prospective Cohort Study." [Article]. <https://dx.doi.org/10.2139/ssrn.5008376>.