NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Onboarding Data and Resource Sharing Informational Document

Table of Contents

Purpose

This document is meant to provide background and information to assist clinical investigators in developing data sharing plans and is to be used along with the accompanying Data Sharing Plan Development Worksheet. This document contains information on data sharing requirements for the NIH Pragmatic Trials Collaboratory, NIH, and medical journals; information on data sharing mechanisms and platforms; and examples from NIH Collaboratory Demonstration Projects.

If you have questions, feedback or suggestions regarding data sharing, please contact us at nih-collaboratory@dm.duke.edu.

Data Sharing Requirements for the NIH Pragmatic Trials Collaboratory, NIH, and Medical Journals

Please note that these policies are current as of the date of this document. Refer to the individual websites for the latest information and full requirements.

NIH Pragmatic Trials Collaboratory Data Sharing Policy

"1. Collaboratory investigators will each share, at a minimum, a final research data set upon which the accepted primary pragmatic trial publication is based.
2. The Collaboratory Steering Committee recognizes that sharing data derived from clinical care in studies performed in partnership with health care systems may, under some situations, require precautions in addition to those regarding patient confidentiality, to protect specific interests of collaborating health care systems, facilities or providers. Precautions such as allowing data sharing in more supervised or restricted settings, such as access to researchers who agree to limited pre-approved research goals, may be appropriate to address these needs in implementing this data sharing policy.
3. Consistent with NIH policy and guidance, Collaboratory investigators will choose the least restrictive method for sharing of research data that provides appropriate protection for participant privacy, health system privacy, and scientific integrity.
4. Collaboratory investigators will work with NIH to implement this data sharing policy, to ensure the appropriate administrative processes and technical infra- structure are in place to support timely data sharing for the Collaboratory."

From: NIH Health Pragmatic Trials Collaboratory Data Sharing Policy

NIH Data Sharing Policy

"Key Points

1. This Policy applies to all human data in the NIH IRP, including the NIH Clinical Center as well as NIH Institutes and Centers.
2. A Data Sharing Plan (PDF File) must be developed for any research involving human data.
3. Data Sharing Plans will be included in the institute scientific review process for research involving human data.
4. The Institute Scientific Director (SD) or their designee is responsible for approving all Data Sharing Plans.
5. All IRP-supported clinical investigators are expected to develop protocols and consent processes/forms to enable broad data sharing for secondary research consistent with this Policy.
6. Sharing data for secondary research purposes shall comply with human subjects research regulations and procedures, if applicable.
7. All IRP investigators are encouraged to deposit data in publicly accessible research repositories for sharing to the extent feasible and appropriate.
8. This Policy is effective as of October 1, 2015. Any intramural research involving human data undergoing scientific review after October 1, 2015 must have a data sharing plan."

From the NIH Intramural Human Data Sharing Policy (updated December 2015). For more information, see NIH Data Sharing Policy and Implementation Guidance.

Medical Journal Data Sharing Requirements

The International Council of Medical Journal Editors (ICMJE) requires that 7 key elements be addressed in the data sharing statement:

1. "Will individual participant data be available (including data dictionaries)?
2. What data in particular will be shared?
3. What other documents will be available?
4. When will data be available (start and end dates)?
5. With whom will data be shared?
6. For what types of analyses will data be shared?
7. By what mechanism will the data be made available?"

From: International Council of Medical Journal Editors' Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals (updated December 2018).

Table 1 summarizes data sharing requirements of select academic journals and publishers to give researchers an idea of what may be required for publication.

Table 1. Data Sharing Requirements of Select Academic Journals and Publishers		
Journal/Publisher	**Requirements**	**Recommended Repository**
BMJ	Requires data from clinical trials to be made available upon request and requires a data sharing statement.	For clinical data, BMJ recommends controlled access repositories, such as clinicalstudydatarequest.com, the YODA project, or Vivli.
Elsevier	Encourages submission of a data paper, uploading data to a repository, or a data sharing statement stating why data can't be shared.	
Nature	Authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications. Restrictions on the availability of data must be disclosed upon submission.	Unstructured repositories like figshare and Dryad if no structured public repositories exist.
NEJM	Data sharing statement	Aligned with ICJME
PLOS	Data sharing statement	Dryad
Wiley	Data sharing statement	Mendeley Data

Examples from NIH Pragmatic Trials Collaboratory Demonstration Projects

NIH Collaboratory Demonstration Project investigators explored the risks to providers and health systems of sharing data. In Table 2 we describe the risks, the steps taken to mitigate the risks, and the data sharing structure that will be used for each of these pragmatic trials.

Table 2. NIH Pragmatic Trials Collaboratory Data Sharing Plans*			
Study name	**Risks to providers or health systems**	**Data sharing structure**	**Steps to mitigate risks to providers or health systems**
ABATE Active Bathing to Eliminate Infection	Data regarding infection rates could be used for inappropriate comparisons of facilities or with public reports. Detailed	Private enclave managed by study team	Potential users may propose specific queries. Only query results (not individual data) will be shared.

Table 2. NIH Pragmatic Trials Collaboratory Data Sharing Plans*			
	information regarding facilities and utilization patterns could reveal proprietary business information.		
ICD-Pieces **Improving Chronic Disease management with Pieces**	Data regarding patterns of care could be used for biased or inappropriate comparisons across facilities or health systems. Given different specifications, comparison to publicly reported quality measures would be misleading.	Private archive managed by NIDDK	Patient-level data will be de-identified and stored in aggregate database. Identifiers for healthcare system, primary practice and patients will be removed. Use of aggregate dataset will be governed by authorized agreements with NIDDK.
LIRE **Lumbar Image Reporting with Epidemiology**	Data regarding treatment patterns and resource use could be used for inappropriate or biased comparisons across health systems and could reveal proprietary health system business information.	Private archive managed by study team	Patient-level datasets will de-identified by health systems, clinics, providers, and patients. Investigators will authorize release to specific users for specific purposes.
PPACT **Pain Program for Active Coping and Training**	Data on opioid prescribing patterns could be misused for inappropriate comparisons of providers or facilities.	Public archive of a modified dataset	Public-use dataset will not include facility or health system identifiers, characteristics or prescribing/referral practices of individual providers, or patient-level data on race or ethnicity.
SPOT **Suicide Prevention Outreach Trial**	Data on suicide attempt rates could be used for biased or inappropriate comparisons of suicide attempts or suicide mortality across health systems.	Public archive of a modified dataset	Public-use dataset will not include indicator for health system.

Table 2. NIH Pragmatic Trials Collaboratory Data Sharing Plans*			
STOP CRC Strategies and Opportunities to Stop Colon Cancer in Priority Populations	Data on screening rates could be misused for inappropriate or biased comparisons of performance across clinics or inaccurate comparisons with public quality measures.	Private archive managed by study team	De-identified patient-level data will be available, with permissions and data use agreements in place. Data use agreements will limit to specific research uses and require destruction after authorized analyses are completed.
TiME Time to Reduce Mortality in End-Stage Renal Disease	Data regarding mortality could be misused for inappropriate or biased comparisons of facilities or healthcare systems. Detailed data regarding patterns of care could reveal proprietary business information.	Private archive managed by NIDDK	De-identified patient-level data that are aggregated across provider organizations will be stored at the NIDDK Central Repository. Facility identifiers, dialysis provider organization identifiers, and data elements that are unique to one of the dialysis providers will be removed. Data will be made available through formal request and a data use agreement between the requestor and the NIDDK.
TSOS Trauma Survivors Outcomes and Support	Data regarding baseline patient characteristics and study outcomes could be used for biased or inappropriate comparisons of care in participating facilities.	Private archive managed by study team	De-identified patient level data will be provided, with priority given to research that will effect trauma care systems nationwide and Collaboratory investigators.

*Assumes HIPAA-compliant patient de-identification for all patients and a data use agreement where appropriate.

Table from: Simon G, et al. Data Sharing and Embedded Research: Data Sharing Solutions for Embedded Research. In: *Rethinking Clinical Trials: A Living Textbook of Pragmatic Clinical Trials*. Bethesda, MD: NIH Pragmatic Trials Collaboratory. Available at: https://rethinkingclinicaltrials.org/chapters/dissemination/data-share-top/data-sharing-solutions-for-embedded-research/. Updated December 20, 2021. DOI: 10.28929/070.

Data Sharing Mechanisms

In Table 3, we describe different technical structures for data sharing and considerations that may assist researchers in selecting the appropriate mechanism for their trial. For more details, see the Living Textbook Chapter on Data Sharing.

Table 3. Technical Structures for Data Sharing From Least Restrictive (and Least Expensive) to Most Restrictive (and Most Expensive)

Structure	Description	Additional elements	Resource needs	Example
Public archive	Analyzable data can be obtained by any user for any use No restriction on the kinds of research questions new users can address	May impose restrictions like prohibitions against re-identification or access to small cell counts May de-identify certain elements, such as study site or demographics, or present sensitive data as an aggregate summary variable	Initial development and annotation Maintenance and access costs	Agency for Healthcare Research and Quality (AHRQ) Healthcare Cost and Utilization Project (HCUP)
Private archive	Analyzable data can be obtained by authorized users Honest broker or the original owner of the data decides which uses to authorize Requires binding agreement by recipient regarding protection and use of transferred data	As noted for public archive	As noted for public archive Evaluation of requests Execution of data sharing, data use, data transfer, and other agreements, including agree-ments covering data with full identifiers Monitoring of compliance with agreements, and response to breach of agreements	Yale University Open Data Access (YODA) Project Centers for Medicaid and Medicare (CMS) Limited Data Sets National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Central Repository

Table 3. Technical Structures for Data Sharing From Least Restrictive (and Least Expensive) to Most Restrictive (and Most Expensive)				
Public enclave	Any user may query the data, but not take possession of it. Only aggregate results may be removed from the enclave No restriction on the kinds of questions users can address	May impose restrictions like prohibitions against re-identification, passing the data to other users, or access to small cell counts May de-identify certain elements, such as study site or demographics	Initial development and annotation Ongoing curation and governance Creation and maintenance of informatics support for analyses, including software licenses and computational capabilities, and file storage Personnel needed to ensure data quality, etc.	Centers for Medicare and Medicaid Services (CMS) Virtual Research Data Center (VRDC)
Private enclave	Similar to public enclave with regard to provisions for analyzing data without taking possession of it Honest broker or the original owner of the data decides which uses to authorize	Moderated by an honest broker or by representatives of the study and/ or site (either queries or results)	As noted for public enclave Additional resources to evaluate requests and supervise the conduct of approved studies	Food and Drug Administration (FDA) Sentinel Distributed Data Set

Table from: Simon G, et al. Data Sharing and Embedded Research: Data Sharing Solutions for Embedded Research. In: *Rethinking Clinical Trials: A Living Textbook of Pragmatic Clinical Trials*. Bethesda, MD: NIH Pragmatic Trials Collaboratory. Available at: https://rethinkingclinicaltrials.org/chapters/dissemination/data-share-top/data-sharing-solutions-for-embedded-research/. Updated December 20, 2021. DOI: 10.28929/070.

Examples of Data Sharing Platforms

There are many public and private data sharing platforms to choose from, and some will fit some projects more than others. In Table 4, we list and briefly describe some of them for informational purposes. Note that this list is not comprehensive nor is the Collaboratory mandating use of one of these platforms. This list represents possible platforms for consideration.

Table 4. Data Sharing Platforms

Platform	Description
clinicalstudydatarequest.com	Platform for sharing patient-level data
Dryad	A curated resource that makes the data underlying scientific publications discoverable, freely usable, and citable; provides a general purpose home for different data types
FAIRsharing	General data repository
figshare	Allows uploading of files up to 5GB in any file format and previewing of them in browser.
GitHub	Large code hosting platform; private, public, open source
HCUP	Agency for Healthcare Research and Quality (AHRQ) Healthcare Cost and Utilization Project
Mendeley Data	Certified, free-to-use repository that hosts open data from all disciplines, whatever its format (e.g., raw and processed data, tables, codes and software)
NIH Data Sharing Repositories	NIH supported data repositories that make data accessible for re- use. Most accept submissions of appropriate data from NIH- funded investigators (and others), but some restrict data submission to only those researchers involved in a specific network.
OSF	General data repository
re3data.org	Catalogues of registered and certified data repositories
Sentinel Distributed Data Set	Food and Drug Administration (FDA) Sentinel initiative (claims data)
Vivli	Global Clinical Research Data Sharing Platform
VRDC	Centers for Medicare and Medicaid Services (CMS) Virtual Research Data Center
YODA Project	A controlled access repository
Zenodo	General data repository

Examples of Data Sharing Statements

As previously described, the International Council of Medical Journal Editors (ICMJE) requires that 7 key elements be addressed in the data sharing statement. Below are example statements that that have been used to fulfill these requirements.

Suicide Prevention Outreach Trial (SPOT) Data Sharing Statement

"A deidentified version of the analytic dataset will be made available at the time of the initial publication of primary study findings. Consistent with policies of the NIH Collaboratory, all resources (intervention materials, specifications, computer code, etc.) will be shared at or before the publication of study results."

From: Simon GE, Beck A, Rossom R, Richards J, Kirlin B, King D, Shulman L, Ludman EJ, Penfold R, Shortreed SM, et al. 2016. Population-based outreach versus care as usual to prevent suicide attempt: study protocol for a randomized controlled trial. Trials. 17(1):452. doi:10.1186/s13063-016-1566-z.

NIH Pragmatic Trials Collaboratory Data Sharing Statement

Links to the de-identified data set as well as resources, such as the study protocol, consent documents, phenotypes and the data dictionary can be found at https://rethinkingclinicaltrials.org/data-and-resource-sharing/.