NCATS National Clinical Cohort Collaborative (N3C) Clinical Data Transfer Protocol

Contributor-side transfer of existing clinical data to the N3C Enclave; Advarra single IRB protocol

Reviewing IRB	Advarra (single IRB / reliance version)
Related NIH platform protocol	NCATS National Clinical Cohort Collaborative (N3C) Data Enclave Repository; NIH IRB IRB002511
Related NIH review status	Initial study approval effective March 3, 2026; no greater than minimal risk; waiver of the consent process
Overall Principal Investigator	Christopher M. Hartshorn, PhD
Protocol scope	Contributor-side identification, preparation, and transfer of existing clinical data to NCATS; no downstream researcher access or analysis is authorized by this protocol
Study duration	120 months
Version date	April 2, 2026

This protocol governs only the activities performed by participating data contributors to identify eligible records, create a HIPAA Limited Data Set, generate privacy-preserving linkage tokens, and transfer data to NCATS. Storage, harmonization, researcher access, dynamic workspace provisioning, and analysis inside the N3C platform are governed separately by NIH IRB protocol IRB002511.

1. Protocol Summary

This protocol describes the multisite activity by which participating United States data contributors identify eligible records in their own systems, create a HIPAA Limited Data Set (LDS), generate privacy-preserving record linkage (PPRL) tokens at the source, and transfer the approved data to the National Center for Advancing Translational Sciences (NCATS) for inclusion in the National Clinical Cohort Collaborative (N3C). The later use of those data for secondary research inside the enclave is governed separately by NIH IRB protocol IRB002511.

All source data already exist. They were originally collected for clinical care, billing, administrative operations, regulatory purposes, or public health reporting. There is no recruitment. There is no intervention. There is no direct contact with individuals. The main foreseeable risk is loss of confidentiality.

Critical scope point: this protocol authorizes contributor-side transfer activity only. It does not authorize data download or research analysis by sites or outside investigators.

2. Scope and Relationship to NIH IRB Protocol IRB002511

This Advarra single IRB protocol is limited to contributor-side transfer activities. It covers local record identification, LDS creation, generation of PPRL tokens, secure transmission to NCATS, response to initial data-quality questions, and planned refreshes of contributed data.

It does not cover repository stewardship, harmonization into the common data model, review of downstream Data Access Requests (DARs), provisioning of dynamic workspaces, disclosure review of outputs, or investigator analysis inside the N3C enclave. Those activities fall under the separate NIH IRB protocol titled NCATS National Clinical Cohort Collaborative (N3C) Data Enclave Repository, IRB002511.

Project-specific requests by investigators to use deidentified data or LDS data within N3C remain outside this protocol. Those requests are controlled through the Data Use Agreement (DUA) and DAR process. Where required by local policy, the requesting institution must also provide its own IRB determination for that project.

Activity	Primary oversight and governing documents
Local extraction, LDS creation, PPRL token generation, secure transfer, and refreshes	This Advarra sIRB protocol; executed Data Transfer Agreement (DTA) and Linkage Honest Broker Agreement (LHBA); the participating institution's local privacy and compliance processes
Repository receipt, harmonization, storage, DAR review, dynamic workspace provisioning, and output controls	NIH IRB protocol IRB002511; N3C repository Operating Procedures; DUA and DAR governance
Downstream investigator projects requesting enclave access	Executed DUA, project-specific DAR, Data Access Committee review, and any requesting-institution IRB determination required by local policy

1. The Data Transfer Agreement (DTA) for this protocol can be found as Addendum A attached to this protocol.
2. The Data Use Agreement (DUA) for this protocol can be found as Addendum B attached to this protocol.
3. The Linkage Honest Broker Agreement (LHBA) can be found as Addendum C attached to this protocol.
4. NIH IRB Protocol IRB002511 that this protocol references can be found as Addendum D attached to this protocol.
5. NIH IRB IRB002511 approval can be found as Addendum E attached to this protocol.
6. N3C repository Operating Procedures can be found as Addendum F attached to this protocol.

3. Background and Rationale

The original N3C program showed that large-scale, privacy-conscious data sharing across United States health systems can support timely secondary research while protecting the people represented in the data. The current N3C Clinical effort expands that approach beyond the original pandemic-focused environment and supports research across many diseases and conditions.

That broader effort depends on a lawful and standardized transfer process at the contributing institution. This protocol provides that contributor-side framework. Participating institutions transfer

existing clinical data to NCATS under an executed DTA. NCATS then conducts harmonization and repository operations within the separately approved NIH platform protocol.

The transferred file is structured as a HIPAA LDS because date elements, and geographic elements permitted under the DTA, may be necessary for longitudinal and population-level research. Linkage of data across permitted clinical data sources is accomplished through PPRL and an independent Linkage Honest Broker (LHB), which allows matching across data sources without sending direct identifiers to NCATS.

4. Objectives

The primary objective is to establish a single, consistent, IRB-reviewed pathway by which participating institutions may contribute existing clinical data to NCATS for inclusion in N3C.

A second objective is to support recurring data refreshes so that the contributed data remain longitudinal and useful over time, including historical information dating back to January 1, 2016 when available in the contributor's records. For data contributed by the Centers for Medicare and Medicaid Services (CMS) historical information dating back to January 1, 1994 can be transferred.

A third objective is to reduce unnecessary multisite variation by offering a central reliance model for sites that choose Advarra review, while preserving each institution's responsibility for lawful disclosure, local privacy compliance, and local operational controls.

5. Study Design and Site Procedures

5.1 Overview

This is a multisite secondary research data-transfer protocol. At each participating site, trained institutional personnel will identify eligible records from pre-existing clinical data sources, apply the current NCATS-provided phenotype, prepare the approved transfer file as an LDS, generate required PPRL tokens within the institutional environment, and transmit the files under the executed agreements.

The term screening in this protocol refers only to algorithmic selection and quality review of records. No person is contacted, recruited, consented, or prospectively evaluated for participation.

5.2 Source record eligibility and the NCATS phenotype

Under the final DTA, data eligible for inclusion must include:

1. Patients of all ages
2. Having a record since Jan 1, 2016 (or Jan 1, 1994 for CMS data).
3. Patients having at least 2 encounters (vitals [e.g. BP, weight, etc.] OR lab measurements OR diagnoses) on different dates at least 30 days apart.

The operative extract is defined by the NCATS-provided phenotype in effect under the executed DTA.

Transferred records will include historical information dating back to January 1, 2016 when available. Participating institutions may provide recurring refreshes to extend follow-up and add newly eligible records.

Minor technical updates to a site's implementation may be made to preserve compatibility with local source systems or the common data model. Any material change that substantially alters the nature of

the records transferred under this protocol must be handled through appropriate protocol governance and amendment, as required.

5.3 Data elements transferred

Within the bounds of the DTA, the contributor may transfer the following categories of data. Availability will vary by institution and source system.

Data domain	Scope within the approved transfer
Demographic elements	Demographic variables permitted in the LDS, including date elements and geographic elements permitted under the DTA
Encounter and care-setting information	Encounter or visit history and related dates; admission, discharge, site location, and transfer information as available
Clinical observations	Vital signs, diagnoses, procedures, laboratory tests and results, and medications with start and stop dates where available
Linkage and study identifiers	Locally generated study identifiers and required hashed linkage fields used for PPRL; direct identifiers remain at the site
Optional ancillary linked data	Patient registries, monitoring-device data, wearables, or other supplemental datasets, but only when they fall within executed agreement scope and are approved for transfer

The contributor institution decides which data within this approved scope are available and lawfully disclosable from its own systems. Data outside the executed agreement are not transferred under this protocol.

5.4 Local preparation, privacy-preserving record linkage, and secure transmission

1. Local institutional personnel identify eligible records inside the contributor's existing systems and prepare the extraction using the current NCATS-provided phenotype.
2. Before any transfer occurs, the site removes direct identifiers from the research file and creates the LDS authorized under the DTA.
3. PPRL tokens are generated at the source from identifiers held only by the contributor. Those tokens are transmitted only to the independent LHB under the LHBA. NCATS does not receive direct identifiers or raw tokens.
4. The approved LDS is transmitted to NCATS by a secure NCATS-approved method under the DTA. The LHB returns only anonymized linkage keys needed to associate corresponding records across approved data sources.
5. No direct identifiers or protected health information beyond what is permitted in a HIPAA Limited Data Set may be transferred. If periodic scans detect impermissible PHI, the contributor and NCATS will work together to remove it.
6. The data extracted at each contributing site is transferred by secure file transfer protocol (sFTP) by default to a FedRAMP[7] certified staging (operational) area on an NCATS server. The sFTP must use credentials from the contributing organization for secure login. The sFTP

protocol is used by default, however, if sFTP is not possible any comparably secure transfer protocol can be used in coordination with NCATS (e.g. HTTPS, SSL/TLS, SCP, etc.).

5.5 NCATS receipt, data-quality review, and refresh cycles

After receipt, NCATS conducts validation, harmonization, and quality review within the separately approved NIH-governed platform environment. Data are harmonized into the Observational Medical Outcomes Partnership (OMOP) Common Data Model. If data-quality issues are identified, NCATS may request correction and retransmission by the contributor institution.

Refresh cycles may continue throughout the approved protocol period. The contributor institution determines the timing and frequency of those updates, subject to the executed agreements and its operational capacity, ideally once a month can be less frequent as site capacity allows. Each refresh follows the same basic sequence as the initial transfer.

6. Study Population

The study population is composed of existing records, not actively enrolled participants. Records may represent individuals of all ages, sexes, races, ethnicities, and health conditions whose data reside in participating contributor systems and meet the NCATS phenotype.

6.1 Inclusion

Records may be included when they are already present in the contributor's systems, meet the NCATS-provided phenotype: a patient of any age with a record after 01/01/2016, (or 01/01/1994 for CMS data) contain at least two encounters at least 30 days apart with clinically relevant data, and can be lawfully disclosed under the executed agreements and applicable institutional policies.

6.2 Exclusion

No records are excluded because of age, sex, race, ethnicity, pregnancy, socioeconomic status, or diagnosis. Records are excluded only when they do not meet the phenotype, cannot be lawfully disclosed, fail required data-quality checks, or are subject to local opt-out or withdrawal handling that prevents future transfer.

6.3 Vulnerable populations

Because the source data arise from routine care and administrative systems, the dataset may include children, pregnant individuals, older adults, and other populations considered vulnerable under 45 CFR 46. Their inclusion is incidental to existing care records. It does not involve targeted recruitment or participant interaction. The privacy protections in this protocol apply equally to all represented groups.

7. Human Subjects and HIPAA Considerations

7.1 No recruitment, intervention, or biospecimen collection

This protocol does not involve recruitment, advertising, consent visits, intervention, biospecimen collection, or follow-up with individuals. Site personnel may access identifiable source records within their own institutional systems in order to run the phenotype, create the LDS, and generate PPRL tokens. Those direct identifiers remain within the institutional environment.

7.2 Request for waiver of informed consent

This protocol requests a waiver of informed consent for the contributor-side transfer activity under 45 CFR 46.116(f). The activity presents no more than minimal risk because the main foreseeable harm is unauthorized disclosure of health information. The waiver will not adversely affect subjects' rights or welfare because the data are pre-existing, the transfer is limited by executed agreements, direct identifiers are not transferred to NCATS, and each institution remains responsible for lawful disclosure under applicable law. The activity could not practicably be carried out with individual consent because it involves large volumes of existing records across many institutions and refresh cycles over time. Providing additional information after participation is not appropriate because there is no direct interaction with individuals.

7.3 HIPAA and local privacy authority

This protocol involves disclosure of HIPAA Limited Data Sets as defined at 45 CFR 164.514(e)(2). Each participating institution remains responsible for determining and documenting its own legal basis for disclosure under HIPAA and any applicable state or local law. That determination may involve the institutional privacy office, a Privacy Board, an IRB, or another process required by the institution. NCATS operates as a federal research entity and is not acting as a HIPAA Covered Entity or Privacy Board for contributor disclosures. The separate NIH platform IRB does not serve as the HIPAA Privacy Board for contributing institutions. This Advarra sIRB protocol does not displace site-specific privacy responsibilities unless a separate arrangement says otherwise. For CMS data, the legal basis for disclosure is established by CMS as the covered entity under 45 CFR 164.512(i) and applicable CMS data sharing policies.

8. Risks, Protections, and Potential Benefits

8.1 Foreseeable risks

The main risk is loss of confidentiality. Because the transferred file is an LDS, it may contain dates and geographic elements permitted under the DTA. There is also a low risk that linkage across approved data sources could increase identifiability if technical or administrative controls failed. There are no physical risks, no treatment-related risks, and no risks from participant interaction because there is none.

8.2 Privacy and confidentiality protections

Several controls reduce that risk. Direct identifiers stay at the contributing institution. PPRL tokens are generated locally and sent only to the independent LHB. NCATS receives only the approved LDS and anonymized linkage keys. The platform environment uses federal security controls for encryption, role-based access, audit logging, and incident response. Data use is bounded by the DTA, the later DUA and DAR process, and the N3C user code of conduct. Row-level data cannot be downloaded by any users.

Contributors remain responsible for excluding records that cannot be lawfully disclosed and for honoring local opt-out requests in future refreshes. If impermissible PHI is detected during periodic scans, the contributor and NCATS will work together to remove it.

8.3 Potential benefits and risk-benefit assessment

There is no direct benefit to individual subjects. The expected benefit is creation of a secure national data resource that supports secondary research on health outcomes, healthcare use, disease patterns, and public health questions. With the protections described above, the overall risk is no greater than minimal.

9. Confidentiality, Data Security, and Incident Response

9.1 Data handling and access restrictions

At the site, identifiable source data remain under local institutional security controls and are used only to prepare the approved transfer. After transfer, repository data are stored within the NCATS FedRAMP-authorized, FISMA Moderate-compliant environment. Access is limited to authorized NIH staff and contractors for repository operations and, later, to approved users operating under DUA and DAR controls. Individual-level data are not downloadable. Aggregate outputs may be released only through formal disclosure review.

Repository records, metadata, DTAs, DUAs, DARs, and associated governance documents are maintained within the secure NCATS environment in accordance with NIH records-management requirements. If the LDS is coded, the contributor does not release the code key and NCATS does not request it.

9.2 Certificate of Confidentiality

The DTA states that transferred data are protected by a Certificate of Confidentiality under section 301(d) of the Public Health Service Act and the NIH Certificate of Confidentiality policy. NCATS may not redisclose identifiable, sensitive research information except as allowed by law and NIH policy.

9.3 Security incident handling

NCATS maintains a formal incident-response process for suspected or confirmed security events. Affected accounts may be suspended immediately, and events are escalated to the NCATS Information Systems Security Officer. Reportable incidents are handled under NIH Human Research Protection Program procedures. Any contributing institution whose data are implicated is notified under the DTA. The N3C DUA also requires downstream approved users and their institutions to notify NIH promptly of data-management incidents and to provide written follow-up.

10. Single IRB Reliance Plan

10.1 Reviewing IRB and reliance mechanism

Advarra will serve as the reviewing IRB for participating institutions that choose reliance for this contributor-side protocol. Reliance may be documented through the SMART IRB Agreement when applicable, or through another written reliance arrangement acceptable to Advarra and the institution. A site may elect not to rely and may instead use its own IRB. This protocol is written so that sites can use a common submission if they choose Advarra review.

Important limitation: Advarra review under this protocol does not replace the separate NIH IRB oversight for the repository platform, and it does not replace site-specific privacy office duties.

10.2 Responsibilities of Advarra, participating institutions, and NCATS

Role	Core responsibilities under this protocol
Advarra	Review the contributor-side protocol and required amendments for relying sites; make Common Rule determinations for the transfer activity; receive site reports required by the reliance arrangement
Participating institution	Confirm lawful local disclosure authority; execute the DTA and LHBA before transfer; provide local context and institutional requirements when needed; ensure trained local personnel conduct extraction and transmission; report material local changes, privacy issues, or noncompliance that affect the transfer activity; implement local patient opt-out handling in future refreshes
NCATS	Receive data under the DTA; maintain secure repository operations under the separate NIH IRB protocol; conduct data-quality review and harmonization; enforce downstream DUA and DAR controls; notify sites when their data are implicated in an incident or when correction and retransmission are required

11. Withdrawal, Site Discontinuation, and Termination

11.1 Individual withdrawal and local patient opt-out requests

Because this protocol involves only secondary use of existing records and no participant interaction, conventional participant withdrawal from research procedures does not apply. Participating institutions remain responsible for honoring any patient opt-out or withdrawal requests required by their local privacy or research-compliance policies. When such a request applies, the affected record must be excluded from future refreshes transmitted to NCATS. After identifiers have been removed at the source and the record has been harmonized in the enclave, prior versions cannot be selectively re-identified for deletion. Instead, the record is administratively flagged as inactive and omitted from subsequent refreshes.

11.2 Site withdrawal and DTA termination

Either NCATS or the contributing institution may terminate the DTA with 30 days written notice. Once contributed data have been ingested into the N3C enclave, they are retained. NCATS will generate future dataset versions that do not include the withdrawn contributor's data. Researchers who already received approval to use those data may continue to use them for the remainder of the approved term and any authorized renewal, consistent with the DTA, DUA, and DAR.

12. Records Retention

NCATS retains repository data, metadata, DTAs, DUAs, DARs, and associated governance records within the secure environment for the project duration and for the retention period required by NIH policy. Participating institutions should retain local regulatory and transfer records, including reliance

documentation, executed agreements, and key transfer correspondence, in accordance with institutional policy, applicable law, and any reliance arrangement.

13. Required Agreements and Governance Documents

No institution may begin transfer activity under this protocol until all required agreements for its role are fully executed. The following documents control the transfer and any later platform access.

Document	Purpose	When required
DTA	Authorizes transfer of contributor data to NCATS for inclusion in the N3C enclave and states the permitted transfer conditions	Required for every participating data-contributing institution before any transfer
LHBA	Enables the contributor's use of the Linkage Honest Broker for required PPRL services	Required because PPRL is required for data submitted to N3C
DUA	Governs downstream institutional or investigator access to N3C data inside the enclave	Required before any approved user may access data for research
DAR	Project-specific request reviewed by the Data Access Committee for downstream research access	Required for each later research project that seeks enclave access

A site that contributes data but does not seek investigator access ordinarily needs the DTA and LHBA for transfer. The DUA and DAR become relevant when an institution or its investigators later request access to repository data for a separate research project.

14. References

1. IRB002511 Initial IRB Outcome - Approval 03MAR2026. National Institutes of Health Institutional Review Board. Dated March 4, 2026.
2. IRB002511_Protocol_03MAR2026. NCATS National Clinical Cohort Collaborative (N3C) Data Enclave Repository. Version date March 3, 2026.
3. N3C Data Transfer Agreement Final Clean 12-11-2025. NCATS National Clinical Cohort Collaborative (N3C) Enclave Data Transfer Agreement.
4. NCATS N3C Clinical Master DUA Dynamic CADR version Clean. NCATS National Clinical Cohort Collaborative Data Use Agreement. Version dated May 8, 2025.
5. N3C Clinical Cohort. National Center for Advancing Translational Sciences website. Accessed April 2, 2026.
6. 45 CFR 46.116(f). Federal Policy for the Protection of Human Subjects; waiver or alteration of informed consent.
7. 45 CFR 164.514(e)(2). HIPAA Privacy Rule; Limited Data Set.
8. 42 USC 241(d). Certificates of Confidentiality.

Appendix A. Abbreviations

Abbreviation	Meaning
DAR	Data Access Request
DAC	Data Access Committee
DTA	Data Transfer Agreement
DUA	Data Use Agreement
EHR	Electronic health record
FedRAMP	Federal Risk and Authorization Management Program
FISMA	Federal Information Security Management Act
HIPAA	Health Insurance Portability and Accountability Act
LDS	Limited Data Set
LHB	Linkage Honest Broker
LHBA	Linkage Honest Broker Agreement
NCATS	National Center for Advancing Translational Sciences
N3C	National Clinical Cohort Collaborative
NIH	National Institutes of Health
OMOP	Observational Medical Outcomes Partnership
PHI	Protected health information
PII	Personally identifiable information
PPRL	Privacy-preserving record linkage
sIRB	Single Institutional Review Board