Institutional Review Board Policies and Procedures Manual

SECTION 16: DATABASES, REGISTRIES, AND GENETIC TESTING IN HUMAN SUBJECTS RESEARCH

16.1 INTRODUCTION

Databases, registries, and repositories involve the collection and storage of information and/or biological specimens over time for future use for clinical and/or research purposes.

A *database* is a collection of patient data, whether identifiable or not, that is maintained for use in clinical care, research, or for future research.

A *registry* is an organized system for the collection, storage, retrieval, analysis, and dissemination of information on individual persons. Most registries focus their efforts on a particular disease or group of similar diseases. They collect data on individuals from multiple sources and/or sites. Registries are operated by many different entities, including the government, universities, groups of hospitals, non-profit organizations, or private groups. Clinical data registries are useful in tracking and evaluating health care quality over time.

A *repository,* or blood or tissue bank is a collection of biological specimens from multiple sources. Organizers maintain the specimens over time and control access to and use of the specimens. Repositories usually include demographic and/or medical information about the individuals from whom the specimens were obtained. Repositories often maintain codes that link the information and specimens to their donor's identity.

16.2 DETERMINING WHEN A DATABASE, REGISTRY, OR REPOSITORY REQUIRES IRB REVIEW

When a database, registry, or repository is created for the intended use in **clinical purposes or strictly for quality improvement projects,** then IRB Review is not required (see Section 7 for determining if a project is considered quality improvement).

When a database, registry, or repository is created for the intended use in **research that involves human subjects and activities preparatory to research** (such as hypothesis generation or the identification of subjects to determine study feasibility), review by the IRB is required.

A research database, registry, or repository may be eligible for ***Exemption*** from IRB review if the following is true:

- the personal health data collected is already existing at the time of the application,
- the data will not contain any of the 18 HIPAA identifiers, ***and***
- a coded "link" will not be maintained that will allow the investigator to link subjects back to their original medical record or specimen.

If all three criteria above are true, the submission of an IRB Exempt Application in Cayuse IRB is required (Exempt criteria: 45CFR 46.101(b)(4)).

16.3 REQUIREMENTS FOR OBTAINING INFORMED CONSENT AND HIPAA AUTHORIZATION

If direct interaction with subjects is planned for a research protocol (i.e., subjects be interviewed, tested, or otherwise contacted for the purpose of obtaining data), Informed Consent and HIPAA Authorization should be obtained for the prospective storage and future research-use of individually identified data or specimens.

In addition, if the database, registry, or repository will contain identifiable data (i.e., *protected health information* (PHI)) that will either be maintained, or shared, outside the Institution, Informed Consent and HIPAA Authorization are required.

Investigators are responsible for obtaining informed consent in accordance with 45 CFR 46.116 and 21 CFR 50.20 to ensure that no human subjects will be involved in the research prior to obtaining legally effective informed consent. Unless otherwise waived by the IRB, investigators are responsible for ensuring that legally effective informed consent will be obtained in accordance with the IRB Policy and Procedure Manual Section 11: Informed Consent.

When minors are recruited for enrollment in to the database, registry, or repository, Parent/Legally Authorized Representative Permission must be sought and Adolescent Assent from those ages 12 to 17 years. When a minor (ages 17 and younger) becomes an adult (at age 18) the investigator is required to either obtain consent from the now-adult subject for continued participation or storage of the data and/or specimens, or de-identify all current samples including the linking key to the specimens so that the identity cannot be ascertained, or destroy the current samples. Any future collections from the subject can only be obtained if the subject has been re-consented.

A database, registry, or repository maintained within the Institution for *clinical care or QI projects* do not require Informed Consent or HIPAA Authorization because employees of the Institution are already considered covered entities under the Privacy Rule.

16.4 THE USE AND STORAGE OF PROTECTED HEALTH INFORMATION

The Privacy Rule establishes a category of health information, referred to as *protected health information* (PHI), which may be used or disclosed to others only in certain circumstances or under certain conditions. PHI is any information in the medical record or designated record set that can be used to identify an individual and that was created, used, or disclosed in the course of providing a health care service such as diagnosis or treatment. PHI also includes identifiable health information about participants of clinical research collected by a researcher during the course of a research study.

The following is the list of the HIPAA identifiers:

1. Names
2. All geographical subdivisions smaller than a State, including street address, city, county, precinct, zip code, and their equivalent geocodes, except for the initial three digits of a zip code, if according to the current publicly available data from the Bureau of the Census: (1) The geographic unit formed by combining all zip codes with the

same three initial digits contains more than 20,000 people; and (2) The initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000

3. All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death, date of treatment; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older
4. Phone numbers
5. Fax numbers
6. Electronic mail addresses
7. Social Security numbers
8. Medical record numbers
9. Health plan beneficiary numbers
10. Account numbers
11. Certificate/license numbers
12. Vehicle identifiers and serial numbers, including license plate numbers
13. Device identifiers and serial numbers
14. Web Universal Resource Locators (URLs)
15. Internet Protocol (IP) address numbers
16. Biometric identifiers, including finger and voice prints
17. Full face photographic images and any comparable images
18. Any other unique identifying number, characteristic, or code (note this does not mean the unique code assigned by the investigator to code the data).

A. Coding and De-Identifying Data and Specimens

The Privacy Rule permits a covered entity to assign to, and retain with, the de-identified health information, a code, or other means of record re-identification. Any code used to replace the identifiers in datasets cannot be derived from any information related to the individual and the master codes, nor can the method to derive the codes be disclosed. For example, a subject's initials cannot be used to code their data because the initials are derived from their name. Another example is an encrypted individual identifier (e.g., a social security number) would not meet the conditions for use as a re-identification code for de-identified health information because it is derived from individually identified information.

Codes used to link data to identifiable information may also be considered identifiers unless:

- the code is unique and not used for any other purpose, and is not derived from another identifier, ***and***
- the database user will not have access to the code key and will not be permitted to re-identify any of the information.

A truly de-identified database does not capture any of the 18 PHI HIPAA identifiers above and there is no way to re-link the data to the identities of the subjects (i.e., to verify or add data

later by using a coded link). Such a database does not require informed consent or HIPAA authorization.

The Privacy Rule permits an investigator to use or disclose PHI for research databases, registries, and repositories under the following circumstances and conditions:

- If the subject has granted specific written permission through a ***HIPAA Authorization;***
- For reviews preparatory to research if certain representations are provided by the investigator;
- If the investigator receives appropriate documentation that the IRB has granted a ***Waiver of the HIPAA Authorization*** requirement;
- If the investigator obtains documentation of the IRB's approval of ***Alteration of the HIPAA Authorization*** requirement;
- If the PHI has been de-identified in accordance with the standards set by the Privacy Rule.

In addition, the IRB may waive the requirement of prospective informed consent and HIPAA Authorization if the data/specimens are prospectively entered into the database/registry or repository without any identifiable private data or information, i.e., none of the 18 HIPAA identifiers and no codes or links of any sort maintained by the investigator that would permit access to identifiable private information about the individual from whom the data/specimens were collected.

16.5 REQUIREMENTS FOR OTHER AGREEMENTS FOR SHARING OF DATA/SPECIMENS

If the data and/or specimens will be kept or shared outside of the Institution, IRB review is required. In addition other agreements, such as Business Agreements (needed when identifiable data is being shared), Data Use Agreements (needed when data shared is only a limited data set), or Materials Transfer Agreements (when specimens are shared) may be needed (contact the Office of Sponsored Programs).

A. Data Use Agreements

A data use agreement (DUA) is needed when a limited data set will be shared outside the institution. A limited data set is a data set that contains no identifiers other than the following elements of PHI, which may remain with the information:

- dates such as admission, discharge, service, DOB, DOD;
- city, state, five digit or more zip code; and
- ages in years, months or days or hours.

This agreement establishes the permitted uses and disclosures of the limited data set by the recipient, consistent with the purposes of the research, and places limits on which personnel can use or receive the data. It is important to note that this information is still protected health information or "PHI" under HIPAA. It is not de-identified information and is still subject to the requirements of the Privacy Regulations.

B. Business Associate Agreements

A business associate agreement (BAA) is needed when PHI (more than a limited data set) will be shared outside the Institution. This is a written contract between the Institution (the HIPAA covered entity) and the HIPAA Business Associate The contract protects PHI in accordance with HIPAA guidelines (e.g., it establishes permitted uses and disclosures of protected health information by the business associate).

C. Material Transfer Agreements

A Material transfer agreement (MTA) is an agreement that governs the transfer of tangible research materials (i.e., identified biologic specimens) between two organizations, when the recipient intends to use it for his or her own research purposes. The MTA defines the rights of the provider and the recipient with respect to the materials and any derivatives.

16.6 IRB SUBMISSIONS REQUIREMENTS FOR A RESEARCH DATABASE, REGISTRY, OR REPOSITORY

For databases, registries, and repositories created for research purposes, the study is to be submitted via Cayuse IRB. Refer to the IRB Policy & Procedure Manual Section 9: Types of IRB Submissions for the full review process. The protocol and/or Cayuse IRB application must include the following information regarding the database, registry, or repository:

i. **Purpose and the Justification:** Investigators are to provide the justification for developing the database, registry, or repository by explaining the purpose for storing the data and/or specimens as related to the specified research aims and hypothesis.

ii. **The Data/Specimens to be Collected and the Method of Collection:** In addition to the description of the data points and/or the types of specimens to be collected, the investigator is to outline if any HIPAA identifiers will be included and maintained. This description must also include how the data/samples will be collected and clarification on the source of such information (i.e. the data/specimens are leftover clinical samples or if data/samples will be collected prospectively from subjects for the purpose of obtaining research data).

iii. **Where the Data/Specimens will be Stored:** Investigators are to describe the location of the storage of the data/specimens (i.e., within the Institution, outside the Institution, central location, etc.).

When a database, registry or repository is created by an investigator, contains identifiers, and will be maintained outside of the Institution the IRB will not waive the requirement of informed consent or HIPAA authorization.

iv. **The Length of Time of the Storage:** Investigators must state how long the data/samples will be stored in the location(s) described.

v. **How the Data/Samples will be Stored:** Investigators must also indicate how the data/samples will be stored in the respective database, registry or repository for the specified research:

a. **Stored with Identifiers:** If data or samples will include any of the 18 HIPAA identifiers, an outline of the protections being taken to protect confidentiality of the data (e.g., by replacing identifiers with codes, storing code keys separately,

data encryption methods, and maintaining password protection on electronic files) must be included.

The consent form for subjects should outline the types of data being collected/stored, and the measures taken to protect confidentiality. IRB submissions that request a waiver of informed consent for the use of identified databases must meet the criteria for such waiver. **Note:** If there will be interaction with (e.g., interviewing or testing) subjects for the purpose of obtaining information for the database (versus drawing from existing clinical or research information), the proposal will not meet criteria for waiver, and both prospective informed consent and HIPAA Authorization will be required.

b. **Stored without Identifiers:** A de-identified database is one in which none of the 18 HIPAA identifiers are collected and it is not necessary for the investigator to be able to ever re-link information in the database to the subjects for any reason (i.e., verifying entries, adding additional information, etc.). Such follow-up is not possible with a de-identified database. A database is considered de-identified when the investigator cannot access codes that would allow identification of subjects from the dataset.

If the database, repository has already been fully de-identified (contains no identifiers, as described above) no IRB review is required. However, to de-identify data for inclusion in a research database that the investigator is creating, the investigator must first submit an Exempt Application via Cayuse IRB.

Note: Studies that are subject to FDA regulation (e.g., many In Vitro Diagnostic studies) must be conducted under an IRB-approved protocol, even if the specimens and/or information used in the study are de-identified.

vi. **Who will have Access to the Data/Specimens:** Investigators are to detail how the data/samples will be accessed and who will control the access and use of the data/specimens for the specified research. This plan should is to include the following:

a. The oversight and the process for deciding who will have access. This description should include any internal departmental/division policies and procedures that would apply. If outside investigators will have access, then a data-sharing agreement must be in place. The agreement must be arranged with the assistance of the Office of Sponsored Programs (OSP) and the type of agreement will be dependent on whether PHI or identified data will be shared.

b. It is required that the database/registry or repository met all Institutional data security policies, including the use of encryption and password-protection. Investigators are to submit to the IRB their plans for how passwords will be generated, distributed, periodically updated, and maintained, and the identification of who will be the data manager for the database/registry or repository and responsible for data protection measures.

In addition, if subsequent research projects that will utilize the data/specimens stored in the database, registry, or repository are created, a new, separate IRB submission for each proposed research project is required.

A. Appendix: Decision Chart for IRB Review of Database, Registry, Repository	
What is the purpose of the database, registry, and repository?	**IRB Approval Needed?**
Clinical care	No
Quality Improvement Activities (i.e., aim to improve systems, current programs, and/or organization performance with the intention to improve outcomes to benefit current patients)	No
Evaluation of a method, process, procedure, etc. that has not previously been proven effective; involves substantial deviations from established practice; may add additional risks or burdens, adds control of some extraneous variables; or when results are intended to be applied to a population outside of the individuals being studied. (This is considered research!)	Yes
Activities Preparatory to Research or Research that includes collected identifiable patient/participant data	Yes
Activities Preparatory to Research or Research that includes collection of patient/participant data that is de-identified (coded) and not linked in any way to identifiers	Yes (Exempt #4)

Will patient health information (PHI) be accessed or collected?	**HIPAA Waiver**	**Informed Consent & HIPAA Authorization**
Data has been previously completely de-identified (i.e. anonymized)	No	No
PHI will be accessed, but none will be collected	Yes	No
Only PHI will be accessed and collected from existing sources (i.e., medical record) (no additional information collected from interaction/intervention with human subjects)	Yes	Request a Waiver of Consent
PHI will be accessed and collected, some obtained from interaction with human subjects	No	Yes

Where will data/specimens be stored and who will have access to it?	**Agreement Needed?***
Data/specimens stored here, only Lurie Children's staff will access	None
One or more of the following elements of a limited data set are being shared outside the Institution: • dates such as admission, discharge, service, DOB, DOD; • city, state, five digit zip code; and • ages in years, months, or days or hours.	Data Use Agreement (DUA) *Contact the Office of Sponsored Programs*

PHI (more than a limited data set outlined above) is being shared outside the Institution *Note: A BAA is needed for each party who will have access to **identifiable** data*	Business Associate Agreement (BAA) *Contact Corporate Compliance/ Hospital Counsel*
Identified human biologic specimens are being shared outside the Institution.	Materials Transfer Agreement *Contact the Office of Sponsored Programs*

*Note: If there is a study specific grant or contract, the appropriate/necessary agreement language may be incorporated.

16.7 CONDUCTING GENETIC RESEARCH AND SECONDARY RESEARCH USE OF DATA AND SPECIMENS

Investigators are increasingly required to consider the ethical significance of conducting genomic research in children and families. The NIH guidelines state: "Genomics research that involves children can reveal information that might raise concerns for children and their families. Specifically, genomic analyses could reveal the presence of particular conditions, disease susceptibilities, or carrier statuses that are relevant for the enrolled children or their family members, including adult onset conditions. Some questions that may arise include whether it is appropriate to disclose the risk of adult-onset conditions to children; whether to honor future autonomy (and the right to an open future) when the child has a condition where he is unlikely to gain capacity; and whether researchers must honor a parental decision not to learn about a piece of information that has clinical significance for their child. Researchers should anticipate how such issues will be managed, including informing parents about the circumstances under which results will be disclosed to the parents and/or the child."

A. Guidance for IRB Submission of Protocols Conducting Genetic Research

When research includes genetic testing on children and/or family members. Investigators are expected to submit with their Cayuse IRB Initial Application a plan that addresses the ethical and logistical issues surrounding this testing. This includes the following:

- How consent will be obtained, including if the permission sought will be for the use in future unspecified research (broad use and/or sharing) or if the research is narrowly defined (tiered/targeted use and sharing).
- The plan for the recruitment and consent of multiple family members, if applicable.
- The plan for the recruitment of identifiable populations, which includes specific racial or ethnic groups, geographically defined communities and members of ultra-rare disease groups, and the procedures in place to protect the confidentiality and privacy of such groups. This is to also include a discussion of the known and unknown risks to the individual and the identifiable population.
- If there is a plan to deposit samples and/or genomic data in biobanks or databases the submission is to include: what samples and/or data will be shared, how the samples and/or data will be shared, how the samples and/or data might be used, if participants may be contacted in the future, what will happen to a participant's samples and/or data following a change in participation status, and what will happen to samples and/or data if the repository closes.
- The plan for when minors enrolled will reach the legal age of consent during the study period and what, if any, re-consent will take place at that time, including what will be done if research participants cannot be reached at the age of majority. This is an important consideration for genomics research in which samples, genomic data, and health information are stored for future use and there is a possibility of re-identification.

i. Risks Associated with the Broad Sharing of Genetic Data

Concerns associated with broad data sharing largely stem from the nature and extent of the genomic and phenotype data involved and the distribution of the data to approved

users for secondary research. It is important to consider any possible risks in the context of the protections put in place to minimize those risks, as well as in the context of the expected benefits of the proposed research.

1. Risk of identification

   Currently available and emerging technologies make the re-identification of specific individuals from raw genomic data increasingly feasible. Risks of re-identification of research subjects may be increased among small and easily identifiable populations; therefore, it may be appropriate to consider de-identifying research data from these populations at the community level. Although the feasibility of using genomic data to re-identify an individual through matching with other data or information is increasingly recognized, the likelihood that a subject's data will be used to re-identify them is anticipated to be very small, but it is unknown.

   All data to be submitted to a NIH-designated data repositories, such as dbGaP, are to be coded and de-identified by the submitting investigator, and the key to the code that links the data to specific individuals held by the Institution.

2. Psychosocial and other harms

   Certain data that include potentially stigmatizing genetic, phenotypic, behavioral, or social traits (e.g., mutations associated with neurological or psychological disorders) may merit particular consideration during IRB review of proposals for data submission. Harms (e.g., stress, anxiety, stigmatization, or embarrassment) to individuals, groups, or populations may potentially arise from the disclosure of such data. For example, some populations demonstrate a higher predisposition to develop certain diseases or disorders than others are generally known to do.

   Higher or lower frequencies of genetic variants that contribute to observed health patterns within these populations might be used to discriminate against or otherwise stigmatize any member of the population group, whether they possess a given genetic variant or not. Additionally, some types of research (e.g. studies of ancestry) may be considered objectionable to certain populations or groups. The IRB will consider delineating the appropriate parameters for use of the data that could minimize the potential for harm to individuals and their families, groups, or populations.

   Note that the Genetic Information Nondiscrimination Act (GINA), an Illinois state law, called the Genetic Information Privacy Act (GIPA), and the Affordable Care Act prohibit the use of genetic information in health insurance or employment decisions.

ii. Guidance for Obtaining Informed Consent for Genetic Research

The consent documents for genetic research involving children and families are to describe the research use(s) for all samples, stored genomic data, and health information collected. Investigators are to ensure that their approach to consent is appropriate for the intended use(s) and sharing with other researchers, whether this will

be for future unspecified research (i.e. "general research use"), or for narrowly defined research use(s).

When genetic testing is included in a study, the consent documents are to include:

- Description of the proposed genetic research or analysis, including any limitations of use (e.g. data may only be used for research on a specific disease);
- Description of the risks of the genetic testing, including the risks related to the broad sharing of data (if applicable);
- Description of how individual privacy and data confidentiality will be protected;
- Description of any benefits to the subject (if applicable) or to others that may reasonably be expected from the research and the broad sharing of data;
- Description of how to withdraw some or all of the data collected and shared for future research use, as applicable; and
- Statement that allows or precludes commercial use of shared data.
- If data will be shared with a NIH database, a statement allowing for submission of the coded data to a government health research database for broad sharing and a statement indicating that identifiers will not be provided to the government database(s) is to be added.
- If an investigator is planning to return results to participants, this plan is to be described in the consent documents and is to include: the type of results (research-related, incidental, and/or secondary), the method of returning results, to whom results are returned, the availability of other resources needed (e.g. counseling, etc.) and if the receipt of such results is optional.

B. Submission of Genomic Data to NIH Data Repository

The National Institutes of Health (NIH) Genomic Data Sharing (GDS) Policy sets forth expectations that ensure the broad and responsible sharing of genomic research data via a NIH data repository (e.g. dbGAP, dbSNP, etc.). The GDS Policy applies to all NIH-funded research that generates large-scale human or non-human genomic data as well as the use of these data for subsequent research. Large-scale data include genome-wide association studies (GWAS), single nucleotide polymorphisms (SNP) arrays, and genome sequence, transcriptomic, metagenomic, epigenomic, and gene expression data, irrespective of funding level and funding mechanism (e.g., grant, contract, cooperative agreement, or intramural support).

The GDS Policy applies to:

- Competing grant applications that are submitted to NIH for the January 25, 2015, due date or subsequent due dates;
- Proposals for contracts that are submitted to NIH on or after January 25, 2015; and
- NIH intramural research projects generating genomic data on or after January 25, 2015.

Effective January 25, 2015, the IRB is required to review investigators' requests to submit data to NIH data repositories and must also certify that the informed consent that was obtained from subjects was/is consistent with NIH requirements for sharing genomic data.

For studies that plan prospective enrollment of genomic data, investigators are required to submit with their Initial application in Cayuse IRB, the description of the GDS plan as provided to the NIH and informed consent documents with the required genomic elements. The IRB will review data sharing plans for consistency with the GDS Policy. For existing studies where data have been collected prior to January 25, 2015, a Modification in Cayuse IRB will be required to be submitted along with all versions of the approved IRB consent documents utilized during subject recruitment.

i. Requirements for NIH Data-Sharing Plans

For investigators, genomic data sharing plans are to be submitted as part of an application for NIH funding. The genomic data sharing plans are to include the required descriptions as outlined in the GDS policy, which includes the following:

- Data Type: Explanation of whether the research involves human data, non-human data, or both; the type of genomic data that will be shared (e.g., sequence, transcriptomic, epigenomic, and/or gene expression data) and whether it is individual-level data, aggregate-level data, or both; listing of any other information that is anticipated to be shared such as relevant associated data (e.g., phenotype or exposure data) and information necessary to interpret the data (e.g., study protocols, data collection instruments, survey tools).

- Data Repository: Identification of the data repositories to which the data will be submitted, and for human data, whether the data will be available through unrestricted or controlled-access.

- Data Submission and Release Timeline: Providing a timeline for sharing data in a timely manner. Investigators should de-identify human genomic data that they submit to NIH-designated data repositories according to the standards set forth in the HHS Regulations for the Protection of Human Subjects to ensure that the identities of research subjects cannot be readily ascertained with the data. Investigators should also strip the data of identifiers according to the HIPAA Privacy Rule. The de-identified data should be assigned random, unique codes by the investigator, and the key to other study identifiers held by the submitting institution

- IRB Assurance of the Genomic Data Sharing Plan: Statement that an IRB has reviewed the genomic data sharing aspects of the project, or provide a timeline for such review. IRB review of the investigator's proposal for data submission is an element of the Institutional Certification (see below) which assures that the proposal for data submission and sharing is appropriate.

- Appropriate Uses of the Data: The appropriate use of the data should be described. Under the GDS Policy, data is expected to be shared for broad research purposes. If such use of the data is not appropriate, as expressed in informed consent documents of the research participants whose data are included in the dataset, any limitations on the data use should be described in the Institutional Certification.

NIH provides standard language to guide the development of data use limitations. Respect for, and protection of the interests of, research participants are fundamental to NIH's stewardship of human genomic data. The informed consent under which the data or samples were collected is the basis to determine the appropriateness of data submission to NIH-designated data repositories, and whether the data should be available through unrestricted or controlled access.

- Request for an Exception to Submission: If submission of human data generated in the study would not be appropriate because the Institutional Certification criteria cannot be met, the investigator should explain why in the genomic data sharing plan and describe an alternative mechanism for data sharing.

ii. Institutional Certification

An Institutional Certification stipulating the appropriate uses of data submitted will be provided by ORIC and signed by the IO prior to award of when genomic data generation is proposed. The purpose is to assure that submission of data to a NIH data repository is consistent with the GDS Policy and with the informed consent of the original study participants.

As part of the process to develop the Institutional Certification, the IRB will review the investigator's proposal for data submission and sharing included in the funding application. This information is required to be submitted within the Initial application in Cayuse IRB for prospective studies and via a Modification application for data collected prior to January 25, 2015.