Database

A research database is any collection of patient-level data, whether identifiable or not, that is maintained for use in future research. Federal regulations and UMC policies that protect the privacy of patients and research subjects apply to both the creation and use of research databases, as described below.

Getting Started

First, determine whether the database will contain identifiers. Identifiers include not only information (e.g., name, address, SSN, or medical record number) that can be used to identify someone directly, but also any of the other 18 data elements listed in the HIPAA privacy rule (see Attachment B), such as dates of birth or treatment.

A code number that is linked to an identifier is itself an identifier, unless

- The code is unique and not used for any other purpose (and is not derived from another identifier, such as SSN); and
- The database user will not have access to the code key and will not be permitted to re-identify any of the information.

Second, determine how information will be obtained for the database. Will the source be existing clinical or research information, or will patients/subjects be interviewed, tested, or otherwise contacted for the purpose of obtaining research data? A researcher who will interact with subjects for the purpose of collecting identifiable data for a database should ask for the subjects' consent and HIPAA authorization, as described below.

A physician-investigator may also ask his or her patients to give their consent/HIPAA authorization to permit clinical data (and associated specimens) to be included in a research database. Procedures for obtaining consent and authorization are described below.

Creating and Using the "De-Identified" Database

To determine whether a de-identified database is suitable for your research, consider whether it may ever be necessary to re-link information in the database to the identities of the data subjects for the purpose of verifying entries, adding additional information, etc. Such follow-up is generally not possible with a de-identified database that a researcher creates for his or her own use, because the researcher is not permitted to treat the database as "de-identified" if he or she has access to the keys to re-identification codes.
To use a research database that has already been fully de-identified (contains no identifiers, as described above) contact the IRB office for a determination of whether the research constitutes human subjects research, however, to de-identify data for inclusion in a research database that the PI is creating, the PI must first submit an Exempt Research Application and Application for Waiver of HIPAA Privacy Authorization to the IRB.

Creating a Research Database of Identifiable Information

If you will maintain identifiable data or information in your research database, then you must submit a New Research Application to the IRB. You will be the "Database PI" for the database protocol and will be responsible for maintaining the database, controlling access to it, and making required submissions to the IRB.

In the application you should describe the data and identifiers to be included in the database, explain the scope of intended research uses for the data, and indicate how you will protect the privacy of subjects' data and security of the information (e.g., by replacing identifiers with codes, storing code keys separately, and maintaining password protection on electronic files). The application will request the information necessary for the IRB to grant a waiver of informed consent (unless you intend to ask each subject for his/her consent to include information in the database). Note that waiver of informed consent is not available for databases that are used in FDA-regulated research.

You must also comply with the HIPAA Privacy Rule in one of two ways: have each subject sign an authorization for the inclusion of his/her data (the authorization may be combined with a consent form), or submit an Application for IRB Waiver of HIPAA Privacy Authorization to the IRB.

Note that if you are interacting with (e.g., interviewing or testing) subjects/patients for the purpose of obtaining new information for the database (versus drawing from existing clinical or research information), your project will not meet criteria for waiver and you must obtain both informed consent and HIPAA authorization.

NOTE: If you are submitting an application to collect data only for use in a single study (i.e., the purpose of your research application is not to create or add to a research database to be used in future studies), the collection of identifiable information for your research is not the creation of a "research database" and is not covered by this guidance.

Using a Research Database of Identifiable Information

Each time that you query or use the identifiable research database to answer a research question you must submit to the IRB a New Research Application and Application for IRB Waiver of HIPAA Authorization.
De-identified Data

De-identified data are not considered to be Protected Health Information (PHI). The Safe Harbor under HIPAA permits a covered entity to consider data "de-identified" if all of the following identifiers removed:

- Names
- Geographic subdivisions smaller than a state except first three digits of the zip code;
- All elements of dates (except year) for individuals under 90 years old; all elements of dates (including year) for those 90 years old or older;
- Telephone numbers;
- Fax numbers;
- E-mail addresses;
- Social security numbers;
- Medical record numbers;
- Health plan beneficiary numbers;
- Account numbers;
- Certificate/license numbers;
- Vehicle identifiers and serial numbers, including license plate numbers;
- Device identifiers and serial numbers;
- Web Universal Resource Locators (URLs);
- Internet protocol address numbers;
- Biometric identifiers, including voice and finger prints;
- Full face photographic images and any comparable images;
- Any other unique, identifying number characteristic, or code, except for unique codes, provided that the persons who receive or use the data do not have access to the code keys or any means of re-identifying data subjects.

Question 1: I understand that certain types of databases require IRB approval and a HIPAA waiver under the Privacy Rule. I'm not sure I understand what databases are covered by this requirement.

Maybe the easiest way to approach this is to talk about what databases are not covered by this requirement. First, UMC's entire clinical database is not a "research database" even though it may be used for research. When patients come to UMC, clinicians collect data and specimens that identify each patient ("health information"). This information is used for a myriad of treatment, payment and operations purposes. For example, this health information may be used to follow patients, perform look-backs when clinical problems are discovered, and to schedule follow-up visits, or for billing purposes, plan eligibility verification, quality assurance activities, teaching activities, preparing clinical protocols, etc.

Identifiable health information in the general clinical database may also be used for research under certain circumstances. Although the general clinical database is not itself a "research database" that requires IRB approval and a HIPAA waiver, the clinical database may not be used
for a research purpose (e.g., queried to answer a research question) without meeting IRB and HIPAA requirements.

**Question 2: But what about a database that was created by a clinician/researcher by extracting health information from the general clinical database?**

A database that was created from the general clinical database may or may not be a "research database" requiring a protocol and an IRB waiver. The answer depends upon the intended use of the secondary database. If the clinician/researcher created the secondary database as the "shadow record" for his or her patients and uses the database to follow patients, perform look-backs when clinical problems are discovered, schedule follow-up visits, perform QA or QI activities, prepare teaching materials, or prepare clinical protocols, then, like the general clinical database, this secondary or "abstracted" database would not be a "research database."

On the other hand, if this abstracted database was/is to be created principally for research purposes (sole or predominant purpose is to analyze the data to answer a research question or to use the data for future studies), whether by the clinician/researcher for his/her own use or for the use of others in the department or elsewhere, this would be a "research database." The investigator who has/is going to create the database should submit an application for IRB approval of the database, along with a HIPAA waiver of Authorization, using eIRB.

**Question 3: What about adding health information to these research databases**

Researchers must have Authorization and informed consent from the patient or a waiver of both from the IRB to add health information to research databases