Limited Data Sets and Data Use Agreements

For more information on sharing data with other investigators with several specific examples, download the NIH Data Sharing Workbook. This document also reviews items that should be incorporated into the protocol when data sharing with other investigators is part of the research plan.

Definitions:

Data Use Agreement - an agreement that is required under the Privacy Rule and must be entered into *before* there is any use or disclosure of a limited data set (defined below) to an outside institution or party.

HIPAA authorization – written consent obtained from a patient or health plan member that permits a covered entity or business associate to use or disclose PHI to an individual/entity for a purpose that would otherwise not be permitted by the HIPAA Privacy Rule.

Identifier - one or more data elements that renders the subjects readily identifiable. This includes a code number or study ID number that can be used to link back to the individual.

Limited data set - a set of identifiable healthcare information that the HIPAA Privacy Rule permits covered entities to share with certain entities for research purposes, public health activities, and healthcare operations without obtaining prior authorization from patients, if certain conditions are met.

Protected health information (PHI) - Protected health information includes all individually identifiable health information, including demographic data, medical histories, test results, insurance information, and other information used to identify a patient or provide healthcare services or healthcare coverage. PHI relates to information created or received in the past, present, or future. 'Protected' means the information is protected under the HIPAA Privacy Rule.

How data are obtained

Investigators may obtain data in a variety of ways. For example the data are likely to have come from one of the following:

- Research data obtained with the informed consent/HIPAA authorization of the subject. The investigator obtaining consent could have stored the data or else they could have been sent to a data coordinating center or central laboratory. In either case, one of the three following situations likely applies:
    - Consent/HIPAA authorization was obtained that included sharing the data for any type of future research (broad consent) or to limited types of future research; or

- Consent/HIPAA authorization was obtained that was silent regarding sharing the data for future research; or
- Consent/HIPAA authorization was obtained that precluded sharing the data for future research.

- Data were obtained under an IRB-approved waiver of the requirements for consent/HIPAA.

- Data were originally obtained under a Data Use Agreement between a provider and a recipient.

Sharing De-identified Data

Data that have been de-identified would not be considered human subjects research and may be used or shared under the HIPAA Privacy Rule. Most commonly, data qualify as de-identified under the Safe Harbor provision of HIPAA, where all of the 18 HIPAA identifiers have been removed.

If data is collected without recording any of the 18 HIPAA identifiers, then it is said to be anonymous. The 18 HIPAA identifiers are:

- Names
- All geographic subdivisions smaller than a State, including street address, city*, parish/county*, precinct*, zip code* and their equivalent geocode (except the initial 3 digits of a zip code)
- All elements of dates* (except Year), including date of birth, date of admission, or date of visit / service
- Telephone numbers
- Fax numbers
- Electronic mail addresses
- Social Security numbers
- Medical record numbers
- Health plan beneficiary numbers
- Account numbers
- Certificate/license numbers
- Vehicle identification numbers and serial numbers, including license plate numbers
- Device identifiers (e.g. implanted medical devices) and serial numbers
- Web URLs
- Internet protocol (IP) address numbers
- Biometric indicators, including finger and voice prints
- Full face photographic images and any comparable images
- Any other unique identifying number, characteristic, or code except as permitted above provided the investigator could not use the information alone or in combination with other information to identify an individual who is a subject of the information.

* These items may be included in a Limited Data Set.

If the 18 identifiers are removed after data collection, then the data have been **anonymized** or **de-identified**.

Data are often linked to a Master List using a unique code. Codes should not be derived from PHI, such as using initials. Codes derived from PHI are still considered PHI. Obscured or permuted dates derived from actual dates will usually still be considered to be PHI. If a data have had all PHI removed but if there is a code that links back to PHI, the data can only be considered to be de-identified under the following conditions:

1. If the data were not collected specifically for the currently proposed research; AND
2. If the investigator cannot readily ascertain the identity of the individuals whose coded data will be used because

- the key to decipher the code has been destroyed before the research begins; OR
- the investigators and the holder of key enter into an agreement prohibiting the release of the key to the investigators; OR
- there are IRB-approved written polices and operating procedures for the data registry that prohibit the release of identifiers.

Requirements for Sharing Data Stored with Identifiers

The requirements vary depending upon what type of data is shared and whether or not HIPAA applies to the data collected. One way of sharing data is to remove identifiers so that the data is no longer identifiable (#1 and #2, listed below). Another way of sharing data requires determination that the study is exempt (#3 and #4, listed below) or IRB review and approval (#5, listed below).

1. Creation of a de-identified data set (so that the data is not readily identifiable) by one of the following means:

- A knowledgeable person provides an analysis documenting that the subjects are not individually identifiable in accordance with 45 CFR 164.514(b)(1)
- All PHI elements in the data set are removed from the data in accordance with 45 CFR 164.514(b)(2).

2. Creation of a Limited Data Set (all PHI elements in the data set are removed except dates and zip codes so that the use of the data will no longer be human subjects research)

   Information about Limited Data Sets is included in this policy.

3. Recording the data to be used without inclusion of the direct identifiers and having an IRB determine that the research qualifies as exempt from the regulations (Exempt Category 4 (ii));

4. Recording the data to be used with inclusion of the direct identifiers and having an IRB determine that the research qualifies as exempt from the regulations (Exempt Category 4 (iii));

- The investigator uses a HIPAA authorization to collect data from participants; Data Use Agreement is required to share the data with outside researchers
- The investigator obtains a waiver of HIPAA authorization from the IRB if the requirements for a HIPAA authorization waiver are met; Data Use Agreement is required to share the data with outside researchers

5. Recording the data to be used with inclusion of the direct identifiers, and requesting that the IRB review and approves the research;

- The investigator obtains the subjects' consent and HIPAA authorization; OR
- The investigator obtains a waiver of the requirements for consent from the IRB.

A Data Use Agreement is required to share identifiable data with another institution that was collected for research using an informed consent form and/or HIPAA authorization. Information about Data Use Agreements is included in this policy.

Limited Data Sets

What is a limited data set?

A limited data set is a data set that is stripped of certain direct identifiers specified in the Privacy Rule. Because limited data sets may contain identifiable information, they are still PHI. A limited data set may be disclosed to an outside party without a patient's authorization only if the purpose of the disclosure is for research, public health, or health care operations purposes and the person or entity receiving the information signs a data use agreement (DUA) with the covered entity or its business associate.

Limited data sets may include only the following identifiers:

- Dates, such as admission, discharge, service, and date of birth (DOB)
- City, state, and zip code (not street address)
- Age
- Any other unique code or identifier that is not listed as a direct identifier.

This means that in order for a data set to be considered a limited data set, all of the following direct identifiers as they relate to the individual or his/her relatives, employers, or household members *must* be removed:

- Names
- Street addresses (other than town, city, state, and zip code)
- Telephone and fax numbers
- Email addresses
- Social security numbers
- Medical record numbers
- Health plan beneficiary numbers
- Account numbers
- Certificate/driver's license numbers
- Vehicle identifiers and serial numbers, including license plate numbers
- Device identifiers and serial numbers
- URLs and IP addresses
- Biometric identifiers, including finger and voice prints

- Full face photographic images and any comparable images.

How are limited data sets created?

A covered entity may use a member of its own workforce to create the "limited data set." On the other hand, the recipient may also create the "limited data set," so long as the person or entity is acting as a business associate of the covered entity.

If the intended recipient of a limited data set is also creating the limited data set as my business associate, do I need both a data use agreement and business associate agreement?

Yes, you will need both a data use agreement (DUA) and business associate agreement (BAA) because the covered entity (Woman's) is providing the recipient with PHI that includes direct identifiers.

Do I have to account for disclosures when I'm using a limited data set?

No, disclosures of "limited data sets" are not subject to the HIPAA accounting of disclosures requirements. DHHS has taken the position that the privacy of individuals with respect to PHI disclosed in a "limited data set" can be adequately protected through a single Data Use Agreement.

Sharing a Limited Data Set (LDS)

If all elements of PHI (protected health information) have been removed from the data except for dates (birth date, dates of service, etc.) and/or zip codes, then the data is considered a limited data set. Geocode data can include both 5 and 9 digit zip codes and can include census tract data (see page 53235 of the 2002 HIPAA Privacy Rule). It may not include street address.

An LDS may be used or disclosed for the purposes of research with the subjects' consent/authorization (e.g. at the time of the original consent) or without their written authorization provided that one (not both) of the following takes place:

- An IRB issues a waiver of HIPAA authorization and (if applicable) a waiver of consent; OR
- A data use agreement is duly executed that assures that the recipient of the LDS will use or disclose the PHI only for specified purposes.

If the data are not readily identifiable, an IRB can conclude that the use of the data does not constitute human subjects research (45 CFR 46 does not apply). (If the data are considered readily identifiable based on it containing dates and zip codes, then the research would be subject to 45 CFR 46 and would require IRB review.)

When the data are not readily identifiable, the research is not subject to the Common Rule (45 CFR 46). IRB review or approval is not required under the regulations. Instead of obtaining a

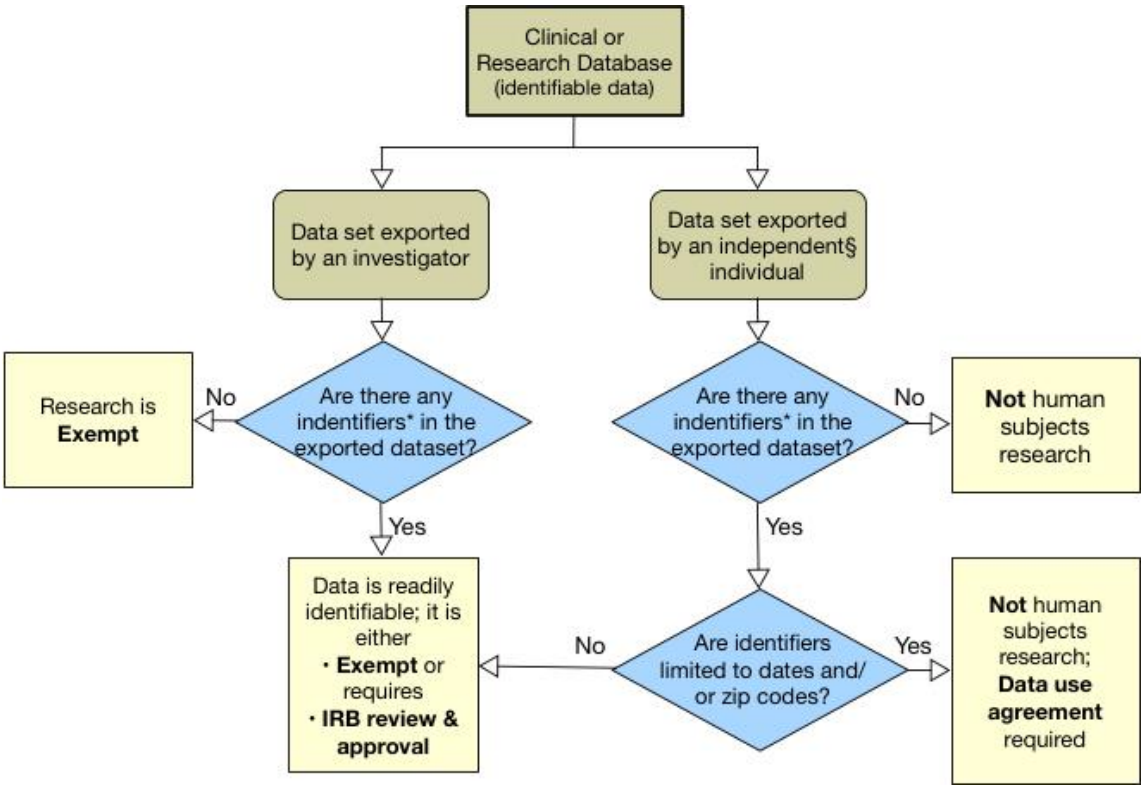
waiver of HIPAA authorization from the IRB, the data can be shared as long as a Data Use Agreement has been executed between the provider (or their institution and the recipient (or their institution). The IRB does not need to review Data Use Agreements. A Data Use Agreement can be used to permit the sharing of data in all of the following situations:

1. A provider at Woman's and a recipient investigator external to Woman's; or
2. A provider external to Woman's and a recipient investigator at Woman's

Requests for limited data sets

A Limited Data Set is information disclosed by a covered entity to a researcher who has no relationship with the individual whose information is being disclosed. WHF IRB must review and approve requests for the use or disclosure of Limited Data Sets for the purpose of research. Requests for limited data sets are applicable only to information retrievable in an electronic format.

Sharing Data from a Database Stored with Identifiers



§ Independent means not connected in any way to the original research involved in creation of the clinical or research source database or in the new research – in other words the individual is an "honest broker". The second issue is whether or not there are any identifiers in the exported (new research) data set. ******* *Identifier means one or more data elements that renders the subjects readily identifiable. This includes a code number or study ID# that can be used to link back to the individual.*

Data that include identifiers can be used or disclosed as permitted in the consent form and written authorization or with a waiver of consent and HIPAA authorization issued by the IRB.

Investigators often want to use data collected for clinical or QI purposes or to analyze research data for other purposes (secondary use). If the investigator needs to retain identifiers for a legitimate research purpose, then secondary uses of the identifiable data might be qualify as human subjects research or not as human subjects research depending on what is shared.

The diagram shows that data from a clinical or research database could qualify as "not human subjects research" – with or without a data use agreement – or might be considered exempt, or might require IRB review and approval. To qualify as "not human subjects research", the individual exporting the identifiable data has to be independent of the research.

The IRB needs to first determine whether or not the data is readily identifiable. If it is, the Common Rule applies; if it isn't, the Common Rule does not apply. Only after making this decision does the IRB consider whether or not the data contains PHI. A data set with PHI might not be considered readily identifiable but HIPAA would still apply. The data could include dates of birth and services without making the individuals *readily* identifiable. (Readily identifiable does not mean potentially identifiable or identifiable with substantial effort.)

Data Use Agreements

In addition to WHF IRB review and approval, the researcher may need to provide Woman's Hospital Research Center with a written Data Use Agreement between Woman's and the researcher. The Data Use Agreement should be signed by Woman's HIPAA Privacy Officer (or designee).

What is a data use agreement?

A data use agreement (DUA) is an agreement that is required under the Privacy Rule and must be entered into *before* there is any use or disclosure of a limited data set (defined below) to an outside institution or party. A limited data set is still protected health information (PHI), and for that reason, covered entities like Woman's must enter into a data use agreement with any recipient of a limited data set from Woman's.

At a minimum, any DUA must contain provisions that address the following:

1. Establish the permitted uses and disclosures of the limited data set;
2. Identify who may use or receive the information;
3. Prohibit the recipient from using or further disclosing the information, except as permitted by the agreement or as otherwise permitted by law;

4. Require the recipient to use appropriate safeguards to prevent an unauthorized use or disclosure not contemplated by the agreement;

5. Require the recipient to report to the covered entity any use or disclosure to which it becomes aware;

6. Require the recipients to ensure that any agents (including any subcontractors) to whom it discloses the information will agree to the same restrictions as provided in the agreement; and

7. Prohibit the recipient from identifying the information or contacting the individuals.

Additionally, covered entities must take all reasonable steps to cure a recipient's breach of the DUA. For example, if Woman's learns that data it provided to a recipient is being used in a manner not authorized under the DUA, Woman's should work with the recipient to correct this problem. If these efforts are unsuccessful, Woman's would be required to cease any further disclosures of PHI to the recipient under the DUA and report the matter to the federal Department of Health and Human Services Office for Civil Rights.

When do I need to obtain a DUA?

A DUA must be entered into *before* there is any use or disclosure of a limited data set to an outside institution or party.

Where do I obtain a DUA?

When Woman's is the *provider* of a limited data set, Woman's requires that a DUA must be signed by the HIPAA Privacy Officer (or designee) to ensure that the appropriate provisions are in place to protect the limited data set.

If a Woman's researcher is the *recipient* of a limited data set from a non-Woman's source, the Woman's researcher will most likely be asked to sign the other party's DUA.