De-identifying your study data

Identifiers

In order for study data to be considered de-identified, all personally identifiable information must be removed. This includes obvious identifiers such as name, date of birth and social security number but also includes other numbers such as school ID number, address and phone number.

Identifiers may* include:

- Names
- Address
- Dates directly related to an individual, including birth date, death date, etc…
- Telephone numbers
- Vehicle identifiers and serial numbers, including license plate numbers
- Fax numbers
- Device identifiers and serial numbers
- Email addresses
- Web Universal Resource Locators (URLs)
- Social security numbers
- Internet Protocol (IP) addresses
- Medical record numbers
- Biometric identifiers, including finger and voice prints
- Health plan beneficiary numbers
- Full-face photographs and any comparable images (tattoos, etc…)
- Account numbers
- Certificate/license numbers
- Any other unique identifying number, characteristic, or code- for example, twitter handle, social media user names, etc…

*Note: These are all HIPAA identifiers and will need to be removed for a dataset to be considered de-identified under HIPAA regulations

When dealing with smaller or more specialized datasets, please also keep in mind the potential of identifying a subject based on a set of indirect identifiers. For example, age and sex and job title- a 52-yr old male nurse may be identifiable in a dataset of survey respondents recruited from a single hospital. Because of this, in some cases datasets containing certain combinations of indirect identifiers may still be considered to be identifiable.

Coded data

In order for coded study data to be considered de-identified, any keys linking study code to personally identifying information such as name, must be destroyed. Datasets containing study ID numbers while the investigators maintain a key linking names to id numbers are still considered to contain identifiable private information and IRB approval must be maintained.

If the investigator is obtaining a dataset from an outside entity, in order for it to be considered de-identified, all of the above identifiers must be removed from the dataset. If the data includes study codes, there should be a written agreement prohibiting the investigator from accessing the link, or, the likelihood for the MU investigator to have access to the link is extremely unlikely (justification will be required).

Continuing review
Data that have been completely de-identified no longer meet the definition of "human subject" and IRB review will no longer be required. If this is the case for your study, please submit a Final Report form to the IRB so the study can be closed.