A note on Protected Health Information (PHI):

Protected health information is defined as information that can be linked to a particular person (person-identifiable). Although HIPAA permits the use of PHI for research purposes, the risk of re-identification of an individual is greatly decreased by removing certain elements from data during the data cleaning process. To de-identify PHI, remove the following list of 18 identifiers of the individual, and their relatives, employees, or household members when data is not needed for analysis:

1. Name
2. All geographic subdivisions smaller than a state, including street address, city, county, precinct, zip code, and other geocodes except for the first three digits of zip code.
3. All dates for elements related to the individual such as birth date, admission date, discharge date, date of death, and all ages greater than 89.
4. Telephone numbers
5. Fax numbers
6. E-mail addresses
7. Social Security numbers
8. Medical Record Numbers
9. Health Plan beneficiary numbers
10. Account numbers
11. Certificate/license numbers
12. VIN/License plate numbers
13. Device identifiers and serial numbers
14. Web URLs
15. IP Addresses
16. Biometric identifiers, including finger and voice prints
17. Full face photographs
18. Any other unique identifying number, characteristic or code

In addition to removing these eighteen identifiers, methods can be applied to de-identify PHI while preserving some of these elements when they are necessary for data analysis.

An example of an appropriate method of de-identification could be sequentially numbering patients upon admission into the study or applying codes to geographic subdivisions, however the patient IDs and other codes used to de-identify data must be stored safely in a separate file so it can be accessed at a later time.

HIPAA allows information to be communicated on a need to know and minimum necessary basis. Simply put, this information should be made available only to those who require access to the information and where it is the minimum necessary to get the job done. Certain identifiers may be necessary to conduct analysis such as geographic identifiers (city, state, zip code), date of birth, date of death, or date of service. These are generally the only PHI data which may be required for analysis and therefore all other PHI should be removed before submitting datasets.