

## Protected Health Information (PHI)

### What is PHI?

Protected health information (PHI) is individually identifiable health information transmitted or maintained in any form or medium by a Covered Entity or its Business Associate. Individually identifiable health information is information, including demographic data that relates to an individual's physical or mental health or the provision of or payment of health care, which identifies the individual.

PHI is used in research studies involving review of existing medical records for research information, such as retrospective chart review. Also, studies that create new medical information because health care is being performed as part of research, such as diagnosing a health condition or a new drug or device for treating a health condition, create PHI that will be entered into the medical record. For example, sponsored clinical trials that submit data to the U.S. Food and Drug Administration involve PHI and are therefore subject to HIPAA regulations.

### What is not PHI?

De-identified health information neither identifies nor provides a reasonable base to identify an individual. Health information by itself without the 18 identifiers is not considered to be PHI. For example, a dataset of vital signs by themselves do not constitute protected health information. However, if the vital signs dataset includes medical record numbers, then the entire dataset must be protected since it contains an identifier. PHI is anything that can be used to identify an individual such as private information, facial images, fingerprints, and voiceprints. These can be associated with medical records, biological specimens, biometrics, data sets, as well as direct identifiers of the research subjects in clinical trials.

**List of 18 Identifiers:** Under HIPAA Privacy Rule "identifiers" include the following:

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; 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)

There are also additional standards and criteria to protect individual's privacy from 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.