De-identified and Limited Data Sets

De-identified Data Sets

The HIPAA Privacy Rule (45 CFR, Parts 160 and 164(A) and (E)) specifies that patient data can be to de-identified by removing 18 elements that could be used to identify an individual or an individual's relatives, employers, or household members; these elements are enumerated in the Privacy Rule. Investigators also must have no actual knowledge that the remaining information could be used alone or in combination with other information to identify the individual who is the subject of the information. Under this method, the following identifiers must be removed:

1. Names
2. All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code, and their equivalent geographical codes, except for the initial three digits of a ZIP code if, according to the current publicly available data from the Census Bureau, a) the geographical unit formed by combining all ZIP Codes with the same three initial digits contains more than 20,000 people or b) the initial three digits of a ZIP code for all such geographic units containing 20,000 or fewer people are 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 ages and elements may be aggregated into a single category of age 90 or older.
4. Telephone numbers
5. Facsimile 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 plates
13. Device identifiers and serial numbers
14. Web universal resource locators (URLs)
15. Internet protocol (IP) address numbers
16. Biometric identifiers, including fingerprints and voiceprints
17. Full-face photographic images and any comparable images
18. Any other unique identifying number, characteristic, or code, unless otherwise permitted by the Privacy Rule for re-identification.

Limited Data Sets

Data sets that include dates of service or entire ZIP codes are considered limited data sets. As defined by HIPAA, limited data sets are data sets stripped of certain direct identifiers that are specified in the Privacy Rule. They are not de-identified information under the Privacy Rule.

A limited data set is protected health information that excludes the following direct identifiers of the individual or of relatives, employers, or household members of the individual:

1. Names
2. Postal address information, other than town or city, State, and zip code
3. Telephone numbers
4. Fax numbers
5. E-mail addresses
6. Social security numbers
7. Medical record numbers
8. Health plan beneficiary numbers
9. Account numbers
10. Certificate/license plate numbers
11. Vehicle identifiers and serial numbers
12. Device identifiers and serial numbers
13. Web URLs
14. Internet Protocol (IP) address numbers
15. Biometric identifiers, including fingerprints and voiceprints
16. Full-face photographic images and any comparable images

Limited data sets may include the following: city, state and zipcodes; all elements of dates; and unique codes or identifiers not listed as direct identifiers