
# De-identification of PHI in Accordance with the HIPAA Privacy Rule

The De-identification Standard: Section 164.514(a) of the HIPAA Privacy Rule provides the standard for de-identification of protected health information. Under this standard, health information is not individually identifiable if it does not identify an individual and if the covered entity has no reasonable basis to believe it can be used to identify an individual.

The Privacy Rule provides two methods by which health information can be designated as de-identified:

- Expert Determination Method
- Safe Harbor Method

**Expert Determination Method**

A covered entity may determine that health information is not individually identifiable health information only if:
(1) A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable:
(i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and
(ii) Documents the methods and results of the analysis that justify such determination.

**Safe Harbor Method**

The following identifiers of the individual or of relatives, employers, or household members of the individual, are removed:

(A) Names
(B) All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code, and their equivalent geocodes, except for the initial three digits of the 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
(C) All elements of dates (except year) for dates that are directly related to an individual, including birth date, admission date, discharge date, death date, 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
(D) Telephone numbers
(E) Fax numbers
(F) Email addresses
(G) Social security numbers
(H) Medical record numbers
(I) Health plan beneficiary numbers
(J) Account numbers
(K) Certificate/license numbers
(L) Vehicle identifiers and serial numbers, including license plate numbers
(M) Device identifiers and serial numbers
(N) Web Universal Resource Locators (URLs)
(O) Internet Protocol (IP) addresses
(P) Biometric identifiers, including finger and voice prints
(Q) Full-face photographs and any comparable images
(R) Any other unique identifying number, characteristic, or code, except as permitted by paragraph (c) of this section [Paragraph (c) is presented below in the section "Re-identification"]

Satisfying either method would demonstrate that a covered entity has met the standard in §164.514(a) above. De-identified health information created following these methods is no longer protected by the Privacy Rule because it does not fall within the definition of PHI.

**Re-identification**

The implementation specifications further provide direction with respect to re-identification, specifically the assignment of a unique code to the set of de-identified health information to permit re-identification.

If a covered entity or business associate successfully undertook an effort to identify the subject of de-identified information it maintained, the health information now related to a specific individual would again be protected by the Privacy Rule, as it would meet the definition of PHI. Disclosure of a code or other means of record identification designed to enable coded or otherwise de-identified information to be re-identified is also considered a disclosure of PHI.

(c) Implementation specifications: re-identification. A covered entity may assign a code or other means of record identification to allow information de-identified under this section to be re-identified, provided that:
(1) Derivation. The code or other means of record identification is not derived from or related to information about the individual and is not otherwise capable of being translated so as to identify the individual; and
(2) Security. The covered entity does not use or disclose the code or other means of record identification for any other purpose, and does not disclose the mechanism for re-identification.

**Guidance on Satisfying the Safe Harbor Method**

*May parts or derivatives of any of the listed identifiers be disclosed consistent with the Safe Harbor Method?*
No. For example, a data set that contained patient initials, or the last four digits of a Social Security number, would not meet the requirement of the Safe Harbor method for de-identification.

*Can dates associated with test measures for a patient be reported in accordance with Safe Harbor?*
No. Dates associated with test measures, such as those derived from a laboratory report, are directly related to a specific individual and relate to the provision of health care. Such dates are protected health information. As a result, no element of a date may be reported to adhere to Safe Harbor.

*What constitutes "any other unique identifying number, characteristic, or code" with respect to the Safe Harbor method of the Privacy Rule?*
This category corresponds to any unique features that are not explicitly enumerated in the Safe Harbor list (A-Q), but could be used to identify a particular individual. Thus, a covered entity must ensure that a data set stripped of the explicitly enumerated identifiers also does not contain any of these unique features. The following are examples:
Identifying Number: There are many potential identifying numbers. For example, the preamble to the Privacy Rule at 65 FR 82462, 82712 (Dec. 28, 2000) noted that "Clinical trial record numbers are included in the general category of 'any other unique identifying number, characteristic, or code.'
Identifying Code: A code corresponds to a value that is derived from a non-secure encoding mechanism. For instance, a code derived from a secure hash function without a secret key (e.g., "salt") would be considered an identifying element. This is because the resulting value would be susceptible to compromise by the recipient of such data. As another example, an increasing quantity of electronic medical record and electronic prescribing systems assign and embed barcodes into patient records and their medications. These barcodes are often designed to be unique for each patient, or event in a patient's record, and thus can be easily applied for tracking purposes.
Identifying Characteristic: A characteristic may be anything that distinguishes an individual and allows for identification. For example, a unique identifying characteristic could be the occupation of a patient, if it was listed in a record as "current President of State University."

Source: Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule; https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html