

Patient Case Records and CoC-Accredited Cancer Programs

Regulatory Requirements

The Privacy Rule is part of a suite of regulations promulgated pursuant to the administrative simplification provisions of The Health Insurance Portability and Accountability Act (HIPAA), among which are requirements for the de-identification of protected health information under 45 CFR § 164.514 (b).

The Privacy Rule protects individually identifiable health information that is held or transmitted by covered entities and their business associates; such information is called protected health information (PHI). The data in the NCDB are designated as a limited data set and have been stripped of 14 direct identifiers as defined by §164.514(e). HIPAA also requires patients over the age of 89 to be grouped as 90 and older. A further constraint identified in the production of the Participant User Files (PUF) is that The American College of Surgeons (“The College”), through a Business Associate Agreement (BAA) with each Commission on Cancer (CoC) accredited cancer program, ensures the protection of the identity of accredited programs and any information that might identify individual physicians.

The National Cancer Data Base (NCDB) PUFs have been developed with these requisites in mind:

1. Data files are de-identified beyond the requirements stipulated in 45 CFR § 164.514.
2. Cancer program identity is masked.
3. No physician-specific information is provided.

De-Identification and Confidentiality Considerations

Geographically isolated facilities can easily be identified by a combination of hospital characteristics and information describing their location (e.g., university hospitals in southern states; community hospitals in the northeast, upper mid-west, Alaska or Hawaii; or NCI-designated cancer centers in the Rocky Mountain region).

Information about patient residence at the state-level can be used to identify specific facilities as clusters or concentrations of patients residing in particular states and can be used to differentiate facilities. For example, two NCI-designated cancer centers in New York State, Roswell Park and Memorial Sloan Kettering, draw patients from noticeably different geographic areas and have distinguishable patient case-mix characteristics.

De-Identification and Confidentiality Actions

In the release of the NCDB Participant Use Files (PUFs), the following steps are taken so that the risk that information in these files could be used alone or in combination with other information, to identify a subject or a CoC accredited cancer program, is very small. These steps, critical to privacy and the obligations incurred through the College’s BAAs, have imposed minimal limitations on potential analyses that might be undertaken by interested investigators.

1) Only year portions of date items reported to the NCDB are included in the PUFs, where appropriate. In place of full eight digit dates (MMDDYYYY), measures of elapsed time from a common reference dates are provided.

Rationale: In order to comply with HIPAA privacy rule regulations related to de-identified data sets (45 CFR § 164.514 (b)), full dates related to dates of service (diagnosis, provision of medical services, or death) may not be made available in distributed PUFs.

Impact: None. In lieu of providing exact date values the following date-related information are present in the PUFs: calendar year of diagnosis; the number of days between the date of the patient's diagnosis and the initiation of various treatment modalities (i.e., surgery, radiation therapy, and systemic therapies) where applicable and calculable; and the number of months from diagnosis to the date of death of the patient or the last date of contact the reporting facility had with the patient. Analyses assessing either the passage of time between clinical events or understanding the sequenced order of clinical events should not be affected.

2) Per HIPAA guidelines, any patient 90 years of age or older will have age aggregated into a single category of age 90 or older in the PUF.

Rationale: In order to comply with HIPAA privacy rule regulations related to de-identified data sets (45 CFR § 164.514 (b)), patients aged over 89 must be aggregated into a single age group.

Impact: Approximately 1.5% of reported adult patients have their age accordingly adjusted in the PUFs.

3) Limit geographic data describing location of facilities and patients to the level of the US census regions (New England, Middle Atlantic, South Atlantic, East North Central, East South Central, West North Central, West South Central, Mountain, and Pacific). State, county and zip-code specific information will not be included in distributed PUFs.

Rationale: This action exceeds the HIPAA privacy rule regulations related to de-identified data sets (45 CFR § 164.514 (b)), where data may be geo-coded to the level of the first three digits of a zip code, or when populations with less than 20,000 inhabitants patients are aggregated into a single group. This solution offers the highest reasonable level of anonymity and protection to both patients and facilities while still allowing 1) aggregated patient characteristics such as education, income, travel distance, etc., and 2) facility characteristics to be included in distributed PUFs.

Rationale: In only a few instances have published works using NCDB data reported geographic specificity beyond the level of census region. State specific comparisons generally presume population-based coverage, an inquiry that may not be appropriate to a non-population based data source as is the NCDB. NCDB coverage varies from state to state with regard to both types of hospitals and proportion of incident diagnoses reported to the database. Data describing hospital location at the state level can lead to program identification.

4) Removal of all case records reported from Federal (Veteran Affairs [VA] and Department of Defense [DoD]) facilities.

Rationale: The American College of Surgeons has a unique BAA with the VA that restricts release of individual patient information, de-identified or otherwise.

5) Limiting adult files to patients who were 18 years old or older at diagnosis (18-90+).

AND

6) Added precautions have been taken to ensure de-identification of hospitals and patients for any PUF file that contains pediatric and young adult diagnoses (≤ 39). Identifying data items of facility location and facility type are unavailable.

Rationale: Cancer is rare among children, and these actions will protect individual identification information.

7) Removal from distributed PUFs all case records reported from *Freestanding* and *Hospital Associate* accredited cancer programs.

Rationale: These are a small set of programs that report fewer than 100 cases annually to the NCDB and are typically reported as “unknown” or “other” type of facility and are frequently excluded from analyses.

Impact: Collectively, these facilities report approximately 3,000 cases annually to the NCDB.

8) Removal from distributed PUFs all case records reported from accredited programs located in Puerto Rico.

Rationale: With the removal of Federal facilities from the PUF, the remaining accredited program located in Puerto Rico is uniquely exposed due to patient case mix considerations.

9) Collapse NCI centers into the same category of CoC-accredited programs as Teaching/Research hospitals in distributed PUFs.

Rationale: The potential to re-identify any a patient among NCD designated comprehensive cancer centers is high. Combining these centers with other Teaching/Research hospitals represents a reasonable precaution.

10) Provide unique facility and case identifiers in the PUFs.

Rationale: The NCDB collects and maintains reported case records using a combination of facility identification numbers and unique administrative codes maintained at the local reporting registry. The facility identification numbers and the cancer programs to which they correspond are in the public domain and can be accessed on the CoC web site. In order to allow investigators to identify separate reporting facilities and patients, unique identification numbers will be randomly assigned to each reporting facility and to each case record included in the PUF. A cross-walk to the NCDB analytic files for both the facility and patient level identification numbers will be retained by the CoC in order to facilitate any technical support or data review/reconciliation actions that may be required.

11) Require that a data-use agreement be signed by the principle investigator prior to downloading the PUF. The PUF is the property of the American College of Surgeons. Principle Investigators may not copy or distribute provided data files, or profit from the sale or use of the data.

Rationale: The responsibility for protecting the data rests with the principle investigator. Therefore, assurances which specify that no attempt will be made, through direct or indirect means, to identify patients, hospitals, or providers using the data provided through the NCDB PUFs rests with the principle investigator, who holds the accountability for protection.

Removal of specific, patient-identifiable data or case records reported from specific types of cancer programs from distributed PUFs will not deleteriously affect analyses by investigators using the PUF data files derived from the NCDB.