SDTM De-identification for CTN0046   4/29/2013

The CTN0046 SDTM database has been de-identified according to the HIPAA Safe Harbor rules. Those rules are:

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 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 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;

(D) Telephone numbers;

(E) Fax numbers;

(F) Electronic mail 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) address numbers;

(P) Biometric identifiers, including finger and voice prints;

(Q) Full face photographic images and any comparable images; and

(R) Any other unique identifying number, characteristic, or code, except as permitted by paragraph (c) of this section; and

The covered entity does not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information.

The following rules were applied to the CTN0046 SDTM database to de-identify them:

1. All (**DTC) dates converted to year only.
2. All USUBJIDs converted to a random id number called DEIDNUM.
3. Age verified to be less that 89.
4. Specific file by file de-identification changes follow by domain.

AE: Clean.

SUPPAE:
- Removed month and day from dates.
- Removed proper names and patient ids from responses.

CM:
- Removed month and day from dates in CMINDC.

SUPPCM: Clean.

DA: Clean.

SUPPDA: Clean.

DM:
- Recode several low count RACEs to the single group of "OTHER".

DS: Clean.

SUPPDS:
- Day and month removed from signature dates.
- PI name removed.

EG: Clean.

SUPPEG:
- Day and month removed from signature dates.
- PI name removed.

EX: Clean.

SUPPEX: Clean.

IE: Clean.

SUPPIE:
- Day and month removed from signature dates.
- PI name removed.

LB: Clean.

SUPPLB: Clean.

MH:
- Removed month and day from MHTERM.

SUPPMH:
- Removed month and day from QVAL.
- Removed treatment location from QVAL.
- PI name removed.

PE:
- Removed month and day from physical exam results.

SUPPPE: Clean.

QS:
- Date results were truncated to year only.
- Proper names and patient identifiers removed

SUPPQS:
- Date results were truncated to year only.
- Proper names and patient identifiers removed

SC:

- Dropped all RACE based SCTESTs that could identify someone based on a specific race entry. RACE should be gotten from DM.RACE.

SU: Clean.

TA, TE, TI, TS, and TV are all trial metadata files and have no identifying information

VS: Clean.