CTN-0052 CRF Level Data De-Identification Notes

And

Database Descriptions

March 21, 2014

CTN-0052 Data De-Identification Notes

The Case Report Form (CRF)-level de-identified data sets were prepared by reviewing the CRFs and the data dictionary to identify variables that contained, or could potentially contain, identifying information, such as identifiers, dates and free text fields. These variables were programmatically removed or modified in the raw dataset, resulting in de-identified datasets.

- Patient ID: All patient IDs were transformed to factless key values bearing no relationship to the collected study data.

- Site ID: All site ID fields were erased.

- Dates: All dates were converted to "days on study," where the date of randomization = day 0. A negative value for days on study refers to an event that occurred prior to randomization (e.g., date of birth) and a positive value for days on study refers to an event that occurred after randomization (e.g, treatment-phase visit and event dates).

- Text Fields: Most free text fields long enough to contain dates or narrative were emptied, including

    o Any responses derived from "Other. Please specify" or "Specify other reason". For example, other races, marital status, therapy sessions, SAE categories, reasons study drug not given, abnormalities on ECG and Urine Analysis forms.

    o Comment fields

    o On dosing logs: lot numbers, who administered dose.

    o All medical history information.

    o All protocol violations

    o Type of work; job description/parents' type of work; school program name and major; descriptions of living situation, illnesses, concerns, drug/alcohol use, feelings/interests.

- Exceptions: AE terms were *retained* in the database, but any detailed descriptions in a separate field providing additional information about adverse events were erased. This includes relevant labs, medical history and medications, drug/alcohol use description, probable cause of death. Concomitant medication names, purposes, and start year were *retained* in the database.

- Initials: Any field designed to record a person's initials was erased.

- For a complete listing of the erased text fields, refer to the "CTN0052 Dropped Columns.xls" file.

Renaming of datasets: Datasets with names longer than 8 characters were renamed for consistency with SAS Transport requirements.

Original Name	Short Name
CIDI_CTN0052_CC_SUDS_SCORES	CID_ACSD
CIDI_CTN0052_CIDI_AJ_VIEW	CID_AJVW
CIDI_CTN0052_CIDI_L_VIEW	CID_LVEW
CIDI_CTN0052_QUES_COM	CID_QCOM

Renaming of variables: In order to meet the SAS Transport format's 8-characters limit on field names, several datasets required some field names to be shortened for the SAS format files. Note that field names were not shortened for the delimited-text format files. See the "Original Name" column in Data Dictionary tabs for the following datasets:

- ALL1PP

- CIDI52

- CID_ACSD
- CID_AJVW
- CID_LVEW
- CID_QCOM
- ENROLL