CTN-0049 CRF Level Data De-Identification Notes

And

Database Descriptions

August 8, 2018

CTN-0049 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, consent date) and a positive value for days on study refers to an event that occurred after randomization (e.g, treatment-phase visit and event dates).
    - Note that patients who failed screening, or who dropped out prior to randomization, do not show any such days-on-study values, as there is no randomization event to provide the base date for such calculations.
    - Date of Birth: these dates are translated to age as of randomization, rather than a difference-in-days calculation.

- Text Fields: Most free text fields long enough to contain dates or narrative were emptied, including
    - Comment fields.
    - 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.
    - Narratives regarding interaction with patients, treatment progress or followup.
    - On dosing logs: lot numbers, who administered dose.
    - Sample identifiers.
    - Medical history narrative text.
    - All protocol violations
    - 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.

- External Dataset ID fields: Identifiers other than Patient ID that might uniquely identify a study participant, or might link to an external system's participant records, for example in a dataset derived from an external data collection instrument, were erased.

- Initials / Names: Any field designed to record a person's initials or signed name was erased.

- Exceptions: (1) AE terms were *retained* in the database, but any detailed descriptions in a separate field providing additional information about adverse events were erased, such as relevant labs, medical history, drug/alcohol use description, probable cause of death. (2) Concomitant medication names and indications were *retained* in the database.

- For a complete listing of the erased text fields by CRF, refer to the "CTN0049 Nulled Values" spreadsheet file.