
CSP-1008A CRF Level Data De-Identification Notes

And

Database Descriptions

Oct 29, 2014

CSP-1008A 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 to produce 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 IDs were transformed to factless key values bearing no relationship to the collected study data.
- Dates: All dates except date of birth 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).
 - Date of Birth: Dates of birth were converted to age values, computed as of randomization date.
 - Note that patients who failed screening will have no computed studyday values in date fields.
 - In cases where the date was stored as 3 distinct columns (month, day, year) these component value were converted to a date value before computing days on study; the resulting variable was named following the pattern of the 3 component variable names, but substituting the suffix DT.
- 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.
 - Type of work or job description; school program name and major; narrative descriptions of living situation, illnesses, concerns, drug/alcohol use, feelings/interests.
- Text Field Exceptions: (1) AE and medical condition 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.
- Initials / Names: Any field designed to record a person’s initials or signed name was erased.
- For a complete listing of the erased text fields by CRF, refer to the “CSP1008A Nulled Values” csv file.

See the “CSP1008A Dictionary” Excel workbook file for descriptions of the data, organized into tabbed worksheets for each respective CRF/dataset. The columns that have been nulled for de-identification purposes are indicated by a ‘Y’ in the Nulled column.

CSP-1008A Data De-Identification Notes

Renaming of datasets: Datasets in the original study database were named by form number, e.g. the ECG dataset was named FORM08. For presentation in this Data Share system, the datasets have been renamed to brief descriptive dataset names:

<i>Datashare Name</i>	<i>Original Name</i>
AECON	FORM09
BLDSMPL	FORM16
CGIR	FORM12
CONSENT	FORM20
CRAV	FORM03
DEMOG	FORM01
DOSEA	FORM14A
DOSEB	FORM14B
DRUGHX	FORM02
ECG	FORM08
ENRL	FORM11
FUP	FORM19
LABS	FORM07
MEDEVAL	FORM05B
MHX	FORM04
PGIR	FORM13
PHYS	FORM05A
PREG	FORM06A
PREGFU	FORM06B
RAB	FORM10
RABA	FORM10A
SAE	FORM17
SIGN	FORM21
TERMA	FORM18A
TERMB	FORM18B
TSR	FORM15