

---

**CTN-0031A CRF Level Data De-Identification Notes**

**And**

**Database Descriptions**

**October 15, 2013**

---

---

## CTN-0031A CRF Level 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 a list of variables that contained, or could potentially contain, identifying information, such as dates and free text fields. The variables were programmatically removed or modified in the raw dataset, resulting in de-identified datasets. All dates were converted to “days on study,” where the date of consent (BASEDATE) = 0. A negative value for days on study refers to an event that occurred prior to consent (e.g., date of birth) and a positive value for days on study refers to an event that occurred after consent (e.g., 12-month visit date). The following free text fields were emptied:

- 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.
- Comment fields
- AE/SAE terms were retained in the database, but any detailed descriptions in a separate field providing additional information the AE or SAE were erased. This includes relevant labs, medical history and medications, drug/alcohol use description, probable cause of death.
- On dosing logs: lot numbers, who administered dose.
- All medical history information.
- Concomitant medication names, purposes, and start year were *retained* in the database, but start dates and months were dropped.
- 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.

**Note:** **ELIG31 (STUDY ELIGIBILITY FORM FOR CTN-0031 PARTICIPANTS)** dataset has 195 eligible participants from CTN-0031 study that participated in CTN-0031A and **ELIGNC (STUDY ELIGIBILITY FORM FOR NORMAL COMPARISONS)** dataset has 49 additional study participants who were not a part of CTN-0031 study. We have merged both of these dataset so that ELIG31 dataset contains all 244 CTN-0031A study participants. Consent date in ELIG31 is used as the base date for all the subjects.

**DEM** dataset contains data of birth for all subjects participated in CTN 0031A and the field name is DEM01DT. Date of birth of all CTN 0031 subjects who participated in CTN 0031A was added to DEM dataset from ELIG31 dataset (field name for date of birth in ELIG31 is EGI04DT).

Following is a list of all variables in the dataset that were modified or emptied in order to de-identify the data, based on the following commands:

PATIDDEID assigns to each patient a unique value in a variable called PATIDDEID. This PATIDDEID is same for all the subjects which are in both 0031 and 0031A studies, this was done so that merging of data from both studies can be done for a similar PATIDDEID.

BASEDATE names the variable that is used as the base date to calculate days on study. In this study, BASEDATE is date of consent.

DOS names a variable that is to be converted from a date into days on study.

EMPTY defines a variable to fill with missing values.

DROPFIELD defines a data set to drop.

The following documentation presents information on the variable based on the name within the database. To get the corresponding label of the variable, one must look within the data dictionary in Excel that is also provided on the CTN Data Share web site.

---

## CTN-0031A CRF Level Data De-Identification Notes

---

**For all the datasets dates were changed to days of study and site was emptied.**

**scr (System Screening):** Removed scrpid (Participant ID)

**wurs (Wender Utah Rating Scale):** Removed WURS60\_SP (Specify) and WURS61\_SP (Specify)

**PI (PATIENT INFORMATION):** Removed pinum and pinode

**Cbd (COMPLICATIONS FROM BLOOD DRAW):** Removed CBD1FSP, CBD2A, CBD3B and CBD4

**Dem (Demographics):** Removed DEM04SP (Specify)

**dsm1 (DSM-IV Criteria- Substance Related Disorders):** Removed DSM02HSP (Specify other)

**elig31 (STUDY ELIGIBILITY FORM FOR CTN-0031 PARTICIPANTS):** Removed EGIDOB and EGIDTR

**OSM (OXIDATIVE STRESS/ DAMAGE MEASURES):** Removed OSM3B, OSM4B, OSM5B, OSM6B, OSM7B, OSM8B, OSM9B and OSM1DTM

**term (Study Termination):** Removed TERM02sp (Specify) and TERMSIG

**uds (Urine Drug Screen):** Removed UDS01SP (Specify) and UDS10

## CTN-0031A Database Descriptions

Databases for this protocol are available for download in both SAS and ASCII/comma separated value (CSV) format. When you retrieve the datasets from the data download page on the CTN Data Share website, a ZIP file will be provided containing the datasets.

Note that when you download the SAS datasets, the files will be SAS transport files. The transport.xpt contains all the SAS datasets available. The following code can be executed in SAS to retrieve the individual SAS databases.

```
libname newlib "[Name of Path where you want to store the SAS datasets. E.g.: c:\temp\CTN0031A] ";
filename tranfile "[Name of path where you store transport.xpt. E.g.: C:\temp\ transport.xpt]";
proc cimport library=newlib infile=tranfile; run;
```

Also in the SAS ZIP file are SAS transport files corresponding to each individual database. These files are provided to allow import into SPSS and differ slightly from the SAS datasets in the Transport.xpt and from the data dictionaries provided. We describe below specific modifications to the SAS datasets to accommodate SPSS import.

**1. String variables that are truncated because > 200 characters long:**

- a. DSM02A in dataset DSM1
- b. TTF1 in dataset TTF

**2. Renaming of variables to conform to naming of variables to be no longer than 8 characters.**

Variable	NewVar	Member
PATIENTNUMBER	PATI0462	SAS.BISA
PATIENTID	PATI0461	SAS.BLOOD
PATIENTNUMBER	PATI0462	SAS.BLOOD
PATIENTID	PATI0461	SAS.CBD
PATIENTNUMBER	PATI0462	SAS.CBD
PATIENTID	PATI0461	SAS.DEM
PATIENTNUMBER	PATI0462	SAS.DEM
PATIENTID	PATI0461	SAS.DSM1
PATIENTNUMBER	PATI0462	SAS.DSM1
PATIENTID	PATI0461	SAS.DSM2
PATIENTNUMBER	PATI0462	SAS.DSM2
PATIENTID	PATI0461	SAS.DSM3
PATIENTNUMBER	PATI0462	SAS.DSM3
PATIENTID	PATI0461	SAS.DSM4
PATIENTNUMBER	PATI0462	SAS.DSM4
PATIENTID	PATI0461	SAS.DSM5
PATIENTNUMBER	PATI0462	SAS.DSM5
PATIENTID	PATI0461	SAS.ELIG31
PATIENTNUMBER	PATI0462	SAS.ELIG31
PATIENTID	PATI0461	SAS.ELIGNC
PATIENTNUMBER	PATI0462	SAS.ELIGNC
FRSB10BEF	FRSB0245	SAS.FRSB
FRSB10NOW	FRSB0246	SAS.FRSB
FRSB11BEF	FRSB0247	SAS.FRSB

Variable	NewVar	Member
FRSB11NOW	FRSB0248	SAS.FRSB
FRSB12BEF	FRSB0249	SAS.FRSB
FRSB12NOW	FRSB0250	SAS.FRSB
FRSB13BEF	FRSB0251	SAS.FRSB
FRSB13NOW	FRSB0252	SAS.FRSB
FRSB14BEF	FRSB0253	SAS.FRSB
FRSB14NOW	FRSB0254	SAS.FRSB
FRSB15BEF	FRSB0255	SAS.FRSB
FRSB15NOW	FRSB0256	SAS.FRSB
FRSB16BEF	FRSB0257	SAS.FRSB
FRSB16NOW	FRSB0258	SAS.FRSB
FRSB17BEF	FRSB0259	SAS.FRSB
FRSB17NOW	FRSB0260	SAS.FRSB
FRSB18BEF	FRSB0261	SAS.FRSB
FRSB18NOW	FRSB0262	SAS.FRSB
FRSB19BEF	FRSB0263	SAS.FRSB
FRSB19NOW	FRSB0264	SAS.FRSB
FRSB20BEF	FRSB0267	SAS.FRSB
FRSB20NOW	FRSB0268	SAS.FRSB
FRSB21BEF	FRSB0269	SAS.FRSB
FRSB21NOW	FRSB0270	SAS.FRSB
FRSB22BEF	FRSB0271	SAS.FRSB
FRSB22NOW	FRSB0272	SAS.FRSB
FRSB23BEF	FRSB0273	SAS.FRSB

## CTN-0031A Database Descriptions

Variable	NewVar	Member
FRSB23NOW	FRSB0274	SAS.FRFB
FRSB24BEF	FRSB0275	SAS.FRFB
FRSB24NOW	FRSB0276	SAS.FRFB
FRSB25BEF	FRSB0277	SAS.FRFB
FRSB25NOW	FRSB0278	SAS.FRFB
FRSB26BEF	FRSB0279	SAS.FRFB
FRSB26NOW	FRSB0280	SAS.FRFB
FRSB27BEF	FRSB0281	SAS.FRFB
FRSB27NOW	FRSB0282	SAS.FRFB
FRSB28BEF	FRSB0283	SAS.FRFB
FRSB28NOW	FRSB0284	SAS.FRFB
FRSB29BEF	FRSB0285	SAS.FRFB
FRSB29NOW	FRSB0286	SAS.FRFB
FRSB30BEF	FRSB0289	SAS.FRFB
FRSB30NOW	FRSB0290	SAS.FRFB
FRSB31BEF	FRSB0291	SAS.FRFB
FRSB31NOW	FRSB0292	SAS.FRFB
FRSB32BEF	FRSB0293	SAS.FRFB
FRSB32NOW	FRSB0294	SAS.FRFB
FRSB33BEF	FRSB0295	SAS.FRFB
FRSB33NOW	FRSB0296	SAS.FRFB
FRSB34BEF	FRSB0297	SAS.FRFB
FRSB34NOW	FRSB0298	SAS.FRFB
FRSB35BEF	FRSB0299	SAS.FRFB
FRSB35NOW	FRSB0300	SAS.FRFB
FRSB36BEF	FRSB0301	SAS.FRFB
FRSB36NOW	FRSB0302	SAS.FRFB
FRSB37BEF	FRSB0303	SAS.FRFB
FRSB37NOW	FRSB0304	SAS.FRFB
FRSB38BEF	FRSB0305	SAS.FRFB
FRSB38NOW	FRSB0306	SAS.FRFB
FRSB39BEF	FRSB0307	SAS.FRFB
FRSB39NOW	FRSB0308	SAS.FRFB
FRSB40BEF	FRSB0311	SAS.FRFB
FRSB40NOW	FRSB0312	SAS.FRFB
FRSB41BEF	FRSB0313	SAS.FRFB
FRSB41NOW	FRSB0314	SAS.FRFB
FRSB42BEF	FRSB0315	SAS.FRFB
FRSB42NOW	FRSB0316	SAS.FRFB
FRSB43BEF	FRSB0317	SAS.FRFB
FRSB43NOW	FRSB0318	SAS.FRFB
FRSB44BEF	FRSB0319	SAS.FRFB

Variable	NewVar	Member
FRSB44NOW	FRSB0320	SAS.FRFB
FRSB45BEF	FRSB0321	SAS.FRFB
FRSB45NOW	FRSB0322	SAS.FRFB
FRSB46BEF	FRSB0323	SAS.FRFB
FRSB46NOW	FRSB0324	SAS.FRFB
PATIENTNUMBER	PATI0462	SAS.FRFB
frsb_date	frsb0668	SAS.FRFB
PATIENTID	PATI0461	SAS.GHF1A
PATIENTNUMBER	PATI0462	SAS.GHF1A
PATIENTID	PATI0461	SAS.GHF1B
PATIENTNUMBER	PATI0462	SAS.GHF1B
PATIENTID	PATI0461	SAS.GHF2
PATIENTNUMBER	PATI0462	SAS.GHF2
PATIENTID	PATI0461	SAS.GT
PATIENTNUMBER	PATI0462	SAS.GT
PATIENTID	PATI0461	SAS.OSM
PATIENTNUMBER	PATI0462	SAS.OSM
PATIENTNUMBER	PATI0462	SAS.PHQ
PATIENTID	PATI0461	SAS.PI
PATIENTNUMBER	PATI0462	SAS.PI
PATIENTID	PATI0461	SAS.RAVLT
PATIENTNUMBER	PATI0462	SAS.RAVLT
PATIENTID	PATI0461	SAS.SCR
PATIENTNUMBER	PATI0462	SAS.SCR
PATIENTID	PATI0461	SAS.STROOP
PATIENTNUMBER	PATI0462	SAS.STROOP
PATIENTID	PATI0461	SAS.TERM
PATIENTNUMBER	PATI0462	SAS.TERM
PATIENTID	PATI0461	SAS.TTF
PATIENTNUMBER	PATI0462	SAS.TTF
PATIENTID	PATI0461	SAS.UDS
PATIENTNUMBER	PATI0462	SAS.UDS
PATIENTID	PATI0461	SAS.VIS
PATIENTNUMBER	PATI0462	SAS.VIS
PATIENTID	PATI0461	SAS.WCST
PATIENTNUMBER	PATI0462	SAS.WCST
PATIENTNUMBER	PATI0462	SAS.WURS
WURS60_SP	WURS0660	SAS.WURS
WURS60_YN	WURS0661	SAS.WURS
WURS61_SP	WURS0663	SAS.WURS
WURS61_YN	WURS0664	SAS.WURS

---

## **CTN-0031A Database Descriptions**

---

THIS PAGE INTENTIONALLY LEFT BLANK.