
CTN-0044 CRF Level Data De-Identification Notes

And

Database Descriptions

April 2nd, 2014

CTN-0044 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 with the exception of date of birth 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., Informed consent signed date) and a positive value for days on study refers to an event that occurred after consent (e.g, 12-month visit date). Note that the values for date of birth in the DEM and Microcog data sets were modified to age at randomization. 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.
- All medical history information.
- 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.

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 PATDEID

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

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.

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.

Note: BFSID on BFS dataset was masked with a random number.

For all the datasets dates were changed to days of study, site and randomization date (RANDDT) was emptied.

enra (Screening enrollment): Removed sccnstdt and sccomm (comments)

enrb (Randomization enrollment): Removed RAASMTDT, RAOTHRSP (Specify) and RAJUDGSP (Specify)

enrz (Screen Failures): Removed RAASMTDT, RAOTHRSP (Specify) and RAJUDGSP (Specify)

ad1 (Adverse Event): Removed (Specify), a1aebps, a1comm, a1hospad, a1hospd

ad2 (Serious Adverse Event Summary): Removed a2summ, a2medhx, a2follup and a2addinf

ad3 (Serious Adverse Event Medical Reviewer): Removed a3addinf and a3comm

Dem (Demographics): Removed deothrsp (Specify) and demcomm (Comments)

alb (Alcohol Breathalyze): Removed ab2othsp (Specify), ab1othsp (Specify) and albcomm (Comments)

CTN-0044 CRF Level Data De-Identification Notes

bfs (Brief Screen): Removed BFPATID, BFSID, BFSUABSP (Specify), bfiotfsp (Specify) , bfscomm, bfrothsp (Specify) and bfothrsp (Specify)

css (Coping Strategies Scale-Brief Version): Removed csscomm (Comments)

dsm (DSM-IV Criteria - Substance Related Disorder): Removed dsotprdr

ftn (Fagerstrom Test for Nicotine Dependence): Removed ftcesssp (Specify) and ftncomm (Comments)

mca (MicroCog Assessment): Removed mcacomm (Comments)

nms (Non-CTP/Study Medical and Other Services): Removed nmothspe, nmotcrsp (Specify) and nmscomm

phq (Patient Health Questionnaire): Removed phevntsp (Specify) and phqcomm (Comments)

qol (EuroQoL Questionnaire): Removed qolcomm (Comments)

rbs (Risk Behaviors Survey): Removed rbscomm (Comments)

tfb (Time Line Follow Back): Removed tfbcomm (Comments)

uds (Urine Drug Screen): Removed ud1nocsp (Specify), ud1coltm, ud2nocsp (Specify), ud2coltm and udscomm (Comments)

alo (Assessment Location Form): Removed alvisasp, alvisbsp, alviscsp, alvisdsp, alvisesp, alvisfsp, alvisgsp, alvishsp, alvisisp, alvisjsp, alvisksp, alvislsp, alvismsp, alvisnsp, alvisosp, alvispsp, alvisqsp, alvisrsp, alvisssp, alvistsp, alvisusp, alvisvsp, alviswsp, alvisxsp, alvisyvsp, alviszsp, alvi1asp, alcomasp, alcombsp, alcomcsp, alcomdsp, alcomesp, alcomfsp, alcomgsp, alcomhsp, alcomjsp, alcomisp, alcomksp, alcomlsp, alcommsp, alcomosp, alcompsp, alcomqsp, alcomrsp, alcomssp, alcomtsp, alcomusp, alcomvsp, alcomwsp, alcomxsp, alcomysp, alcomzsp and alco1asp (All above fields are for specify purpose)

ccl (Counselor Checklist): Removed ccstffid and cclcomm (Comments)

mvf (Missed Visit Form): Removed mvothrsp (Specify)

pfb (Participant Feedback): Removed pfcomprg, pfsugprg, pfcomtes, pfsugtes and pfbcomm (Comments)

sit (Study Intervention Termination): Removed sitermsp (Specify) and sitcom (Comments)

stt (Study Termination): Removed trtrmosp, sttcomm, trtrmrsp (Specify) and TRPISIGN

tff (Treatment-as-Usual Tracking Form): Removed tffcomm (Comments)

tes_modaccess: Removed activity_time, user_name, site_id, user_id

tes_locations: Removed user_name, site_id, user_id, data_id

tes_acct: Removed user_name, hist_time, user_id, data_id, hist_id, site_id

tes_usefulness: Removed user_name, site_id, user_id, data_id

tes_prizes_redeemed: Removed USER_NAME, redeem_datetime, hist_time, redeem_note, user_id, data_id, hist_id, site_id

tes_modcomplete: Removed user_name, user_id, site_id

CTN-0044 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\CTN0044]";
```

```
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. Dataset names that were changed because of > 8 character long:

- a. TES_LOCATIONS to TES_LO29
- b. TES_MODACCESS to TES_MO30
- c. TES_MODCOMPLETE to TES_MO31
- d. TES_PRIZES_REDEEMED to TES_PR32
- e. TES_USEFULNESS to TES_US33

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

Variable	Len	Member
A1COMM	2000	SAS.AD1
A2ADDINF	233	SAS.AD2
A2FOLLUP	2000	SAS.AD2
A2MEDHX	526	SAS.AD2
A2SUMM	1982	SAS.AD2
A3ADDINF	208	SAS.AD3
A3COMM	236	SAS.AD3
ALBCOMM	587	SAS.ALB
BFSCOMM	610	SAS.BFS
CCLCOMM	625	SAS.CCL
CSSCOMM	320	SAS.CSS
RACOMM	510	SAS.ENRB
RACOMM	414	SAS.ENRZ
FTNCOMM	281	SAS.FTN
NMSCOMM	319	SAS.NMS

CTN-0044 Database Descriptions

Variable	Len	Member
PFBCOMM	500	SAS.PFB
PFCOMPRG	424	SAS.PFB
PFCOMTES	341	SAS.PFB
PFSUGPRG	351	SAS.PFB
PFSUGTES	479	SAS.PFB
PHEVNTSP	363	SAS.PHQ
PHQCOMM	236	SAS.PHQ
QOLCOMM	287	SAS.QOL
RBSCOMM	281	SAS.RBS
SITCOMM	385	SAS.SIT
STTCOMM	281	SAS.STT
TRTMRSSP	1312	SAS.STT
TFBCOMM	291	SAS.TFB
TTFCOMM	370	SAS.TTF
UDSCOMM	475	SAS.UDS

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

Variable	NewVar	Member
Years_of_Education	Year0858	SAS.MICROCOG
assessment_date	asse0864	SAS.MICROCOG
date_of_birth	date0869	SAS.MICROCOG
prof_scaled_subtest_Analogies	prof0879	SAS.MICROCOG
prof_scaled_subtest_Clocks	prof0880	SAS.MICROCOG
prof_scaled_subtest_NumbersForwa	prof0881	SAS.MICROCOG
prof_scaled_subtest_NumbersRever	prof0882	SAS.MICROCOG
prof_scaled_subtest_ObjectMatchA	prof0883	SAS.MICROCOG
prof_scaled_subtest_ObjectMatchB	prof0884	SAS.MICROCOG
raw_score_Analogies_TotalScore	raw_0885	SAS.MICROCOG
raw_score_Analogies_avgRespTime	raw_0886	SAS.MICROCOG
raw_score_Clocks_TotalScore	raw_0887	SAS.MICROCOG
raw_score_Clocks_avgRespTime	raw_0888	SAS.MICROCOG
raw_score_Numbers_Forward_Highes	raw_0889	SAS.MICROCOG
raw_score_Numbers_Forward_TotalS	raw_0890	SAS.MICROCOG
raw_score_Numbers_Forward_avgRes	raw_0891	SAS.MICROCOG

CTN-0044 Database Descriptions

Variable	NewVar	Member
raw_score_Numbers_Reversed_Highe	raw_0892	SAS.MICROCOG
raw_score_Numbers_Reversed_Total	raw_0893	SAS.MICROCOG
raw_score_Numbers_Reversed_avgRe	raw_0894	SAS.MICROCOG
raw_score_ObjectMatch_TrialA_Tot	raw_0895	SAS.MICROCOG
raw_score_ObjectMatch_TrialA_avg	raw_0896	SAS.MICROCOG
raw_score_ObjectMatch_TrialB_Tot	raw_0897	SAS.MICROCOG
raw_score_ObjectMatch_TrialB_avg	raw_0898	SAS.MICROCOG
raw_score_Wordlist1_Discriminabl	raw_0899	SAS.MICROCOG
raw_score_Wordlist1_FalsePositiv	raw_0900	SAS.MICROCOG
raw_score_Wordlist1_TotalScore	raw_0901	SAS.MICROCOG
raw_score_Wordlist1_respBias	raw_0902	SAS.MICROCOG
raw_score_Wordlist2_Discriminabl	raw_0903	SAS.MICROCOG
raw_score_Wordlist2_FalsePositiv	raw_0904	SAS.MICROCOG
raw_score_Wordlist2_TotalScore	raw_0905	SAS.MICROCOG
raw_score_Wordlist2_respBias	raw_0906	SAS.MICROCOG
subtest_scaled_scores_Analogies1	subt0915	SAS.MICROCOG
subtest_scaled_scores_Analogies_	subt0916	SAS.MICROCOG
subtest_scaled_scores_Clocks_Tot	subt0917	SAS.MICROCOG
subtest_scaled_scores_Clocks_avg	subt0918	SAS.MICROCOG
subtest_scaled_scores_Numbers_F1	subt0919	SAS.MICROCOG
subtest_scaled_scores_Numbers_F2	subt0920	SAS.MICROCOG
subtest_scaled_scores_Numbers_Fo	subt0921	SAS.MICROCOG
subtest_scaled_scores_Numbers_R1	subt0922	SAS.MICROCOG
subtest_scaled_scores_Numbers_R2	subt0923	SAS.MICROCOG
subtest_scaled_scores_Numbers_Re	subt0924	SAS.MICROCOG
subtest_scaled_scores_Object_Ma1	subt0925	SAS.MICROCOG
subtest_scaled_scores_Object_Ma2	subt0926	SAS.MICROCOG
subtest_scaled_scores_Object_Ma3	subt0927	SAS.MICROCOG
subtest_scaled_scores_Object_Mat	subt0928	SAS.MICROCOG
subtest_scaled_scores_Wordlist_1	subt0929	SAS.MICROCOG
subtest_scaled_scores_Wordlist_2	subt0930	SAS.MICROCOG
subtest_scaled_scores_Wordlist_3	subt0931	SAS.MICROCOG
subtest_scaled_scores_Wordlist_4	subt0932	SAS.MICROCOG
subtest_scaled_scores_Wordlist_5	subt0933	SAS.MICROCOG
subtest_scaled_scores_Wordlist_6	subt0934	SAS.MICROCOG

CTN-0044 Database Descriptions

Variable	NewVar	Member
subtest_scaled_scores_Wordlist_7	subt0935	SAS.MICROCOG
subtest_scaled_scores_Wordlist_8	subt0936	SAS.MICROCOG
description	desc0870	SAS.TES_ACCT
hist_date	hist0871	SAS.TES_ACCT
hist_time	hist0873	SAS.TES_ACCT
source_code	sour0913	SAS.TES_ACCT
source_subcode	sour0914	SAS.TES_ACCT
user_name	user0940	SAS.TES_ACCT
voucher_prog_id	vouc0941	SAS.TES_ACCT
data_date	data0867	SAS.TES_LOCATIONS
user_name	user0940	SAS.TES_LOCATIONS
Activity_date	Acti0208	SAS.TES_MODACCESS
activity_code	acti0860	SAS.TES_MODACCESS
activity_map	acti0861	SAS.TES_MODACCESS
activity_time	acti0862	SAS.TES_MODACCESS
module_id	modu0875	SAS.TES_MODACCESS
module_name	modu0876	SAS.TES_MODACCESS
user_name	user0940	SAS.TES_MODACCESS
completed_date	comp0865	SAS.TES_MODCOMPLETE
completed_time	comp0866	SAS.TES_MODCOMPLETE
module_id	modu0875	SAS.TES_MODCOMPLETE
module_name	modu0876	SAS.TES_MODCOMPLETE
user_name	user0940	SAS.TES_MODCOMPLETE
USER_NAME	USER0854	SAS.TES_PRIZES_REDEEMED
description	desc0870	SAS.TES_PRIZES_REDEEMED
hist_date	hist0871	SAS.TES_PRIZES_REDEEMED
hist_time	hist0873	SAS.TES_PRIZES_REDEEMED
redeem_datetime	rede0907	SAS.TES_PRIZES_REDEEMED
redeem_note	rede0908	SAS.TES_PRIZES_REDEEMED
source_code	sour0913	SAS.TES_PRIZES_REDEEMED
source_subcode	sour0914	SAS.TES_PRIZES_REDEEMED
voucher_prog_id	vouc0941	SAS.TES_PRIZES_REDEEMED
data_date	data0867	SAS.TES_USEFULNESS
module_id	modu0875	SAS.TES_USEFULNESS

CTN-0044 Database Descriptions

Variable	NewVar	Member
usefulness	usef0938	SAS.TES_USEFULNESS
user_name	user0940	SAS.TES_USEFULNESS

CTN-0044 Database Descriptions

THIS PAGE INTENTIONALLY LEFT BLANK.