
**CTN-0032 CRF Level Data De-Identification Notes
and
Database Descriptions**

November 1, 2011

CTN-0032 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 randomization (BASEDATE) = 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, 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.

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.

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.

DROPPFILE 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.

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

AE: Removed AEID(Staff id)

AEMEDRA: Removed subjno(patientnumber), entry_id(staff id) and coder(person who coded)

COMP: Removed comp01 (staff id), compsp (PI decision) and compsig (Investigator’s signature)

EE: Removed ee16a (unstable house condition), ee03a(specify field), ee03b(specify field), ee06a(specify field), ee06b(specify field), ee09a(specify field), ee09b(specify field), ee12a(specify field), ee12b(specify field), ee15a(specify field), ee15b(specify field) and ee20(general comments).

CTN-0032 CRF Level Data De-Identification Notes

EEU: Removed eeu16a (unstable house condition), eeu03a(specify field), eeu03b(specify field), eeu06a(specify field), eeu06b(specify field), eeu09a(specify field), eeu09b(specify field), eeu12a(specify field), eeu12b(specify field), eeu15a(specify field), eeu15b(specify field) and eeu20(general comments).

HIVG1: Removed h101(staff id), h102sp(Specify judge CTP), h102rs(Specify other reason), h103sp(Specify other reason), h104sp(Specify other reason), h105sp(Specify judge CTP) and h105rs(Specify other reason)

HIVG2: Removed h201(staff id), h202sp(Specify other), h202rsp(Specify other), h203sp(Specify other), h204sp(Specify other), hg205sp(Specify other) and h205rsp(Specify other)

HIVG3: Removed h301(staff id), h302sp(Specify other), h302rsp(Specify other) and h303sp(Specify other)

HIVRER: Removed rer02(staff id), rer06sp(Specify other reason), rer06ps(Specify other plan) and rer07sp(Specify other)

HIVRERF: Removed rfr02(staff id), rfr06sp(Specify other reason), rfr06ps(Specify other plan) and rfr07sp(Specify other)

HIVFR: Removed hrf01(staff id), hrf02sp(specify other), hrf03sp(specify other), hrf04sp(specify other), hrf05sp(specify other), hrf06sp(specify other) and hrf07sp(specify other)

HIVREP: Removed rep02(staff id), rep06sp(Specify other), rep06psp(Specify other) and rep07sp(Specify other)

HIVCC: Removed icc01 (staff id), icc03asp, icc03bsp, icc04asp, icc04esp, icc05sp, icc06dsp, icc06esp and icc03e

HIVCO: Removed ico01(staff id), ico03sp(Specify reason), ico04asp(Specify plan), ico04fsp(Specify other contact), ico05sp(Specify reason), ico06bsp(Specify reason), ico07asp(Specify reason), ico07bsp(Specify plan) and ico08sp(Specify reason)

HIVPOS: Removed pos01(staff id), pos02sp(Specify other), pos03sp(Specify other), pos03rs(Specify other), tcpos05ac(Specify other), pos06sp(Specify other), pos07sp(Specify other) and pos08sp(Specify other)

HIVEXP: Removed exp01(staff id) and exp03sp(Other travel mode)

HTVR: Removed tvr01(staff id), tvr02sp(No med. rec. review specify), tvr03sp(Specify no HIV test), tvr04sp(Specify test type) and tvr05sp(Specify no HIV test)

MRVPC: Removed mrv01(Staff id)

VIS: Removed vis01(staff id), vis03sp(specify other), vis03rs(specify other) and vis07sp(specify other)

WIC: Removed wic01(staff id), wic02sp(CTP specify), wic02rs(specify other), wic03sp(specify other), wic03rs(specify other), wic04sp(specify other) and wic04rs(specify other)

HIVRCO, HIVREPF and HIVRC: These datasets are not included for de-identification as there was no data in these 3 datasets.

CTN-0032 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\CTN0032] ";  
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. Renaming of SAS dataset to conform to 8 character limits:
 - a. WEEKLYFTP was renamed to WEEKLY34
2. Variables dropped because they are greater than 200 characters long
 - a. CTS_REASON from AEMEDRA
 - b. ELG209_COMMENT from ELG2
 - c. CTS_REASON from SAEMEDRA
3. Renaming of variables to conform to naming of variables to be no longer than 8 characters.

Data Set	Old Variable Name	New Variable Name
AE	PATIENTNUMBER	PAT0001
AEMEDRA	ENTRY_DATETIME	ENTR0004
AEMEDRA	MERGE_DATETIME	MERG0001
AEMEDRA	PATIENTNUMBER	PAT0001
AEMEDRA	SUBJECT_ID	SUBJ0007
COMP	PATIENTNUMBER	PAT0001
DEM	PATIENTNUMBER	PAT0001
DEM2	PATIENTNUMBER	PAT0001
EE	PATIENTNUMBER	PAT0001
EEU	PATIENTNUMBER	PAT0001
ELG	PATIENTNUMBER	PAT0001
ELG2	PATIENTNUMBER	PAT0001
ENR	PATIENTNUMBER	PAT0001
HHE	PATIENTNUMBER	PAT0001
HIVCC	PATIENTNUMBER	PAT0001
HIVCO	PATIENTNUMBER	PAT0001
HIVEXP	PATIENTNUMBER	PAT0001
HIVFR	PATIENTNUMBER	PAT0001
HIVG1	H102EDTMDT	H1020018
HIVG1	H102EDTMHH	H1020019
HIVG1	H102EDTMMN	H1020020

Data Set	Old Variable Name	New Variable Name
HIVG1	H102EDTMSS	H1020021
HIVG1	H102SDTMDT	H1020022
HIVG1	H102SDTMHH	H1020023
HIVG1	H102SDTMMN	H1020024
HIVG1	H102SDTMSS	H1020025
HIVG1	H105EDTMDT	H1050027
HIVG1	H105EDTMHH	H1050028
HIVG1	H105EDTMMN	H1050029
HIVG1	H105EDTMSS	H1050030
HIVG1	H105SDTMDT	H1050031
HIVG1	H105SDTMHH	H1050032
HIVG1	H105SDTMMN	H1050033
HIVG1	H105SDTMSS	H1050034
HIVG1	PATIENTNUMBER	PAT0001
HIVG2	H202EDTMDT	H2020018
HIVG2	H202EDTMHH	H2020019
HIVG2	H202EDTMMN	H2020020
HIVG2	H202EDTMSS	H2020021
HIVG2	H202SDTMDT	H2020022
HIVG2	H202SDTMHH	H2020023
HIVG2	H202SDTMMN	H2020024

CTN-0032 Database Descriptions

Data Set	Old Variable Name	New Variable Name
HIVG2	H202SDTMSS	H2020025
HIVG2	H205EDTMDT	H2050027
HIVG2	H205EDTMHH	H2050028
HIVG2	H205EDTMMN	H2050029
HIVG2	H205EDTMSS	H2050030
HIVG2	H205SDTMDT	H2050031
HIVG2	H205SDTMHH	H2050032
HIVG2	H205SDTMMN	H2050033
HIVG2	H205SDTMSS	H2050034
HIVG2	PATIENTNUMBER	PATI0001
HIVG3	H302EDTMDT	H3020011
HIVG3	H302EDTMHH	H3020012
HIVG3	H302EDTMMN	H3020013
HIVG3	H302EDTMSS	H3020014
HIVG3	H302SDTMDT	H3020015
HIVG3	H302SDTMHH	H3020016
HIVG3	H302SDTMMN	H3020017
HIVG3	H302SDTMSS	H3020018
HIVG3	PATIENTNUMBER	PATI0001
HIVPOS	PATIENTNUMBER	PATI0001
HIVPOS	TCPOS05AC	TCPO0003
HIVREP	PATIENTNUMBER	PATI0001
HIVREP	REP4EDTMDT	REP40021
HIVREP	REP4EDTMHH	REP40022
HIVREP	REP4EDTMMN	REP40023
HIVREP	REP4EDTMSS	REP40024
HIVREP	REP4SDTMDT	REP40025
HIVREP	REP4SDTMHH	REP40026
HIVREP	REP4SDTMMN	REP40027
HIVREP	REP4SDTMSS	REP40028
HIVRER	PATIENTNUMBER	PATI0001
HIVRER	RER4EDTMDT	RER40021
HIVRER	RER4EDTMHH	RER40022
HIVRER	RER4EDTMMN	RER40023
HIVRER	RER4EDTMSS	RER40024
HIVRER	RER4SDTMDT	RER40025
HIVRER	RER4SDTMHH	RER40026
HIVRER	RER4SDTMMN	RER40027
HIVRER	RER4SDTMSS	RER40028
HIVRERF	PATIENTNUMBER	PATI0001
HIVRERF	RFR4EDTMDT	RFR40021
HIVRERF	RFR4EDTMHH	RFR40022
HIVRERF	RFR4EDTMMN	RFR40023

Data Set	Old Variable Name	New Variable Name
HIVRERF	RFR4EDTMSS	RFR40024
HIVRERF	RFR4SDTMDT	RFR40025
HIVRERF	RFR4SDTMHH	RFR40026
HIVRERF	RFR4SDTMMN	RFR40027
HIVRERF	RFR4SDTMSS	RFR40028
HTVR	PATIENTNUMBER	PATI0001
MRVPC	PATIENTNUMBER	PATI0001
MRVPC2	PATIENTNUMBER	PATI0001
RAN	PATIENTNUMBER	PATI0001
RANDOM	NUMBERED_INV_ID	NUMB0009
RANDOM	PATIENTNUMBER	PATI0001
RANDOM	RANDOM_ID	RAND0012
RANDOM	SLOT_STATUS	SLOT0008
RANDOM	TRIALNAME	TRIA0001
SAE	PATIENTNUMBER	PATI0001
SAE2	PATIENTNUMBER	PATI0001
SAEMEDRA	ENTRY_DATETIME	ENTR0004
SAEMEDRA	MERGE_DATETIME	MERG0001
SAEMEDRA	PATIENTNUMBER	PATI0001
SAEMEDRA	SUBJECT_ID	SUBJ0007
SCR	PATIENTNUMBER	PATI0001
SRST	PATIENTNUMBER	PATI0001
VIS	PATIENTNUMBER	PATI0001
WEEKLYFTP	CAR03MULTI0	CAR00534
WEEKLYFTP	CAR03MULTI1	CAR00535
WEEKLYFTP	CAR03MULTI2	CAR00536
WEEKLYFTP	CAR03MULTI3	CAR00537
WEEKLYFTP	CAR03MULTI4	CAR00538
WEEKLYFTP	CAR03MULTI5	CAR00539
WEEKLYFTP	CAR06MULTI0	CAR00543
WEEKLYFTP	CAR06MULTI1	CAR00544
WEEKLYFTP	CAR07MULTI0	CAR00546
WEEKLYFTP	CAR07MULTI1	CAR00547
WEEKLYFTP	CAR07MULTI2	CAR00548
WEEKLYFTP	CAR07MULTI3	CAR00549
WEEKLYFTP	CAR07MULTI4	CAR00550
WEEKLYFTP	CAR07MULTI5	CAR00551
WEEKLYFTP	CAR11MULTI0	CAR10556
WEEKLYFTP	CAR11MULTI1	CAR10557
WEEKLYFTP	CAR11MULTI2	CAR10558
WEEKLYFTP	CAR11MULTI3	CAR10559
WEEKLYFTP	CAR11MULTI4	CAR10560
WEEKLYFTP	CAR11MULTI5	CAR10561

CTN-0032 Database Descriptions

Data Set	Old Variable Name	New Variable Name
WEEKLYFTP	CAR11MULTI6	CAR10562
WEEKLYFTP	CAR20MULTI0	CAR20572
WEEKLYFTP	CAR20MULTI1	CAR20573
WEEKLYFTP	CAR20MULTI10	CAR20574
WEEKLYFTP	CAR20MULTI11	CAR20575
WEEKLYFTP	CAR20MULTI12	CAR20576
WEEKLYFTP	CAR20MULTI13	CAR20577
WEEKLYFTP	CAR20MULTI14	CAR20578
WEEKLYFTP	CAR20MULTI15	CAR20579
WEEKLYFTP	CAR20MULTI16	CAR20580
WEEKLYFTP	CAR20MULTI17	CAR20581
WEEKLYFTP	CAR20MULTI18	CAR20582
WEEKLYFTP	CAR20MULTI19	CAR20583
WEEKLYFTP	CAR20MULTI2	CAR20584
WEEKLYFTP	CAR20MULTI3	CAR20585
WEEKLYFTP	CAR20MULTI4	CAR20586
WEEKLYFTP	CAR20MULTI5	CAR20587
WEEKLYFTP	CAR20MULTI6	CAR20588
WEEKLYFTP	CAR20MULTI7	CAR20589
WEEKLYFTP	CAR20MULTI8	CAR20590
WEEKLYFTP	CAR20MULTI9	CAR20591
WEEKLYFTP	DEM11MULTI0	DEM10014
WEEKLYFTP	DEM11MULTI1	DEM10015
WEEKLYFTP	DEM11MULTI2	DEM10016
WEEKLYFTP	DEM11MULTI3	DEM10017
WEEKLYFTP	DEM11MULTI4	DEM10018
WEEKLYFTP	DEM11MULTI5	DEM10019
WEEKLYFTP	DRG01MULTI0	DRG00058
WEEKLYFTP	DRG01MULTI1	DRG00059
WEEKLYFTP	DRG01MULTI10	DRG00060
WEEKLYFTP	DRG01MULTI11	DRG00061
WEEKLYFTP	DRG01MULTI12	DRG00062
WEEKLYFTP	DRG01MULTI13	DRG00063
WEEKLYFTP	DRG01MULTI14	DRG00064
WEEKLYFTP	DRG01MULTI15	DRG00631
WEEKLYFTP	DRG01MULTI2	DRG00065
WEEKLYFTP	DRG01MULTI3	DRG00066
WEEKLYFTP	DRG01MULTI4	DRG00067
WEEKLYFTP	DRG01MULTI5	DRG00068
WEEKLYFTP	DRG01MULTI6	DRG00069
WEEKLYFTP	DRG01MULTI7	DRG00070
WEEKLYFTP	DRG01MULTI8	DRG00071
WEEKLYFTP	DRG01MULTI9	DRG00072

Data Set	Old Variable Name	New Variable Name
WEEKLYFTP	DRG18MULTI0	DRG10090
WEEKLYFTP	DRG18MULTI1	DRG10091
WEEKLYFTP	DRG18MULTI10	DRG10092
WEEKLYFTP	DRG18MULTI11	DRG10093
WEEKLYFTP	DRG18MULTI2	DRG10094
WEEKLYFTP	DRG18MULTI3	DRG10095
WEEKLYFTP	DRG18MULTI4	DRG10096
WEEKLYFTP	DRG18MULTI5	DRG10097
WEEKLYFTP	DRG18MULTI6	DRG10098
WEEKLYFTP	DRG18MULTI7	DRG10099
WEEKLYFTP	DRG18MULTI8	DRG10100
WEEKLYFTP	DRG18MULTI9	DRG10101
WEEKLYFTP	DRG20MULTI0	DRG20104
WEEKLYFTP	DRG20MULTI1	DRG20105
WEEKLYFTP	DRG20MULTI2	DRG20106
WEEKLYFTP	EFM07XEDIT	EFM00387
WEEKLYFTP	END04MULTI0	END00458
WEEKLYFTP	END04MULTI1	END00459
WEEKLYFTP	END04MULTI2	END00460
WEEKLYFTP	ENP05MULTI0	ENP00420
WEEKLYFTP	ENP05MULTI1	ENP00421
WEEKLYFTP	ENP05MULTI10	ENP00422
WEEKLYFTP	ENP05MULTI11	ENP00423
WEEKLYFTP	ENP05MULTI12	ENP00424
WEEKLYFTP	ENP05MULTI13	ENP00425
WEEKLYFTP	ENP05MULTI14	ENP00426
WEEKLYFTP	ENP05MULTI2	ENP00427
WEEKLYFTP	ENP05MULTI3	ENP00428
WEEKLYFTP	ENP05MULTI4	ENP00429
WEEKLYFTP	ENP05MULTI5	ENP00430
WEEKLYFTP	ENP05MULTI6	ENP00431
WEEKLYFTP	ENP05MULTI7	ENP00432
WEEKLYFTP	ENP05MULTI8	ENP00433
WEEKLYFTP	ENP05MULTI9	ENP00434
WEEKLYFTP	ENP07MULTI0	ENP00438
WEEKLYFTP	ENP07MULTI1	ENP00439
WEEKLYFTP	ENP07MULTI10	ENP00440
WEEKLYFTP	ENP07MULTI11	ENP00441
WEEKLYFTP	ENP07MULTI12	ENP00442
WEEKLYFTP	ENP07MULTI13	ENP00443
WEEKLYFTP	ENP07MULTI14	ENP00444
WEEKLYFTP	ENP07MULTI2	ENP00445
WEEKLYFTP	ENP07MULTI3	ENP00446

CTN-0032 Database Descriptions

Data Set	Old Variable Name	New Variable Name
WEEKLYFTP	ENP07MULTI4	ENP00447
WEEKLYFTP	ENP07MULTI5	ENP00448
WEEKLYFTP	ENP07MULTI6	ENP00449
WEEKLYFTP	ENP07MULTI7	ENP00450
WEEKLYFTP	ENP07MULTI8	ENP00451
WEEKLYFTP	ENP07MULTI9	ENP00452
WEEKLYFTP	EXP01MULTI0	EXP00596
WEEKLYFTP	EXP01MULTI1	EXP00597
WEEKLYFTP	EXP01MULTI10	EXP00598
WEEKLYFTP	EXP01MULTI2	EXP00599
WEEKLYFTP	EXP01MULTI3	EXP00600
WEEKLYFTP	EXP01MULTI4	EXP00601
WEEKLYFTP	EXP01MULTI5	EXP00602
WEEKLYFTP	EXP01MULTI6	EXP00603
WEEKLYFTP	EXP01MULTI7	EXP00604
WEEKLYFTP	EXP01MULTI8	EXP00605
WEEKLYFTP	EXP01MULTI9	EXP00606
WEEKLYFTP	EXP02MULTI0	EXP00608
WEEKLYFTP	EXP02MULTI1	EXP00609
WEEKLYFTP	EXP02MULTI10	EXP00610
WEEKLYFTP	EXP02MULTI2	EXP00611
WEEKLYFTP	EXP02MULTI3	EXP00612
WEEKLYFTP	EXP02MULTI4	EXP00613
WEEKLYFTP	EXP02MULTI5	EXP00614
WEEKLYFTP	EXP02MULTI6	EXP00615
WEEKLYFTP	EXP02MULTI7	EXP00616
WEEKLYFTP	EXP02MULTI8	EXP00617
WEEKLYFTP	EXP02MULTI9	EXP00618
WEEKLYFTP	EXP03MULTI0	EXP00620
WEEKLYFTP	EXP03MULTI1	EXP00621
WEEKLYFTP	EXP03MULTI10	EXP00622
WEEKLYFTP	EXP03MULTI2	EXP00623
WEEKLYFTP	EXP03MULTI3	EXP00624
WEEKLYFTP	EXP03MULTI4	EXP00625
WEEKLYFTP	EXP03MULTI5	EXP00626
WEEKLYFTP	EXP03MULTI6	EXP00627
WEEKLYFTP	EXP03MULTI7	EXP00628
WEEKLYFTP	EXP03MULTI8	EXP00629
WEEKLYFTP	EXP03MULTI9	EXP00630
WEEKLYFTP	HXT02MULTI0	HXT00134

Data Set	Old Variable Name	New Variable Name
WEEKLYFTP	HXT02MULTI1	HXT00135
WEEKLYFTP	HXT02MULTI10	HXT00136
WEEKLYFTP	HXT02MULTI11	HXT00137
WEEKLYFTP	HXT02MULTI12	HXT00138
WEEKLYFTP	HXT02MULTI13	HXT00139
WEEKLYFTP	HXT02MULTI14	HXT00140
WEEKLYFTP	HXT02MULTI15	HXT00141
WEEKLYFTP	HXT02MULTI16	HXT00142
WEEKLYFTP	HXT02MULTI2	HXT00143
WEEKLYFTP	HXT02MULTI3	HXT00144
WEEKLYFTP	HXT02MULTI4	HXT00145
WEEKLYFTP	HXT02MULTI5	HXT00146
WEEKLYFTP	HXT02MULTI6	HXT00147
WEEKLYFTP	HXT02MULTI7	HXT00148
WEEKLYFTP	HXT02MULTI8	HXT00149
WEEKLYFTP	HXT02MULTI9	HXT00150
WEEKLYFTP	INJ05MULTI0	INJ00187
WEEKLYFTP	INJ05MULTI1	INJ00188
WEEKLYFTP	INJ05MULTI2	INJ00189
WEEKLYFTP	INJ05MULTI3	INJ00190
WEEKLYFTP	INJ05MULTI4	INJ00191
WEEKLYFTP	INJ05MULTI5	INJ00192
WEEKLYFTP	INJ05MULTI6	INJ00193
WEEKLYFTP	INJ07MULTI0	INJ00197
WEEKLYFTP	INJ07MULTI1	INJ00198
WEEKLYFTP	INJ07MULTI2	INJ00199
WEEKLYFTP	INJ07MULTI3	INJ00200
WEEKLYFTP	INJ07MULTI4	INJ00201
WEEKLYFTP	INJ07MULTI5	INJ00202
WEEKLYFTP	OPTOUTFIELD	OPTO0527
WEEKLYFTP	PATIENTNUMBER	PATI0001
WEEKLYFTP	TRT02MULTI0	TRT00632
WEEKLYFTP	TRT02MULTI1	TRT00633
WEEKLYFTP	TRT02MULTI2	TRT00634
WEEKLYFTP	TRT02MULTI3	TRT00635
WEEKLYFTP	TRT02MULTI4	TRT00636
WEEKLYFTP	TRT02MULTI5	TRT00637
WIC	PATIENTNUMBER	PATI0001