Aware Study Deidentified and Anonymized Data Documentation

Data Extraction Instructions:

The deidentified and anonymized Aware database can be found in the file aware_database.cpt. This is a SAS transport catalog created with SAS version 9.2. It can be extracted into SAS with the following SAS code:

```
filename cportf 'aware_database.cpt';
proc cimport library=work file=cportf memtype=all;
run;
```

The "library=work" can be replaced with a permanent SAS libref as well to store the information on your computer or network.

General Notes:

The following document describes the Aware study database that has been de-identified per HIPAA Safe Harbor rules and subsequently anonymized. Specific details about each type of data are discussed below, but there are a couple of general comments that apply to all data. There is a variable in all datasets called DEIDNUM that is a random number that uniquely identifies a patient in this study. That takes the place of any subject identifier that you might see on the CRF annotation or study documentation. Also, any date field in the database has been converted to a study day offset from the randomization date (RAN.RAN02DT) in the main study. Datetime variables are floating point numbers where the fractional portion of the value represents the time of day.

Inform Datasets:

The Inform electronic data capture form variable annotation can be found in the file named "projectaware_cdd_20100805.pdf." Site identification and all date variable parts, including month, day, and year components, were dropped from all Inform datasets. The following Inform datasets had the following customized deletions as some variables were dropped because they contained significant identifying information.

COMP dataset:
Dropped variables comp01 = staff id, compsig = investigator name, and compsp = investigator decision.

FRRR dataset:
Dropped variable hrf01 = staff id, hrf07 = FRRR Comments.

HIVCC dataset:
Dropped variables icc01 = staff id, icc04esp = No check in call specify, icc07 = HIVCC Comments, icc03bsp = Specify plan test 1, icc03e = test kit lot number.

HIVCO dataset:
Dropped variables ic001 = staff id, ic008 = HIVCO Comments, ico06asp = specify other reason,

HIVG1 dataset:
Dropped variables h101 = staff id, h106 = HIVG1 Comments, h105rs = RESPECT-2 counseling not done other specification.

HIVG12 dataset:
Dropped variables dh101 = staff id, dh106 = HIVG12 Comments, h105rs = RESPECT-2 counseling not done other specification, dh105rs = RESPECT-2 counseling not done other specification.

HIVG2 dataset:
Dropped variables h201 = staff id, h206 = HIVG2 Comments.

HIVPOS dataset:
Dropped variables pos01 = staff id, pos08 = HIVPOS Comments, pos02sp = assessed for potential suicidality other specify, pos03rs = referral provided other specify, pos03sp = no referral provided other specify, pos04sp = HIV info sheet not provided other specify, pos05sp = no check in call scheduled reason other specify, pos06sp = no risk counseling provided other specify, pos07sp = no partner notification services provided other specify.

HIVREP dataset:
Dropped variables rep02 = staff id, rep03b = rapid test kit lot number, rep08 = HIVREP Comments.

HIVRER dataset:
Dropped variables rer02 = staff id, rer08 = HIVRER Comments, rer03b = init test kit number, rer06sp = no initial rapid hiv test result other specify.

HTVR dataset:
Dropped variables tvr01 = staff id, tvr07 = HTVR Comments, tvr02sp = medical records not reviewed other specify, tvr03sp = insufficient data explanation, trv05sp = hiv test results insufficient data explanation.

MRVPC dataset:
Dropped variables mrv01 = staff id, mrv13 = MRVPC Comments.

MRVPC2 dataset:
Dropped variables column10 = MRVPC2 Comments.

PVL dataset:
Dropped variables pvlid = staff id, pvl03 = specify other violation, pvl04 = PV description, pvl06 = PVL Comments, pvl05sp = protocol violation resolved specified.

RAN dataset:
Dropped variable ran03 = IVRS randomization number.

SAE dataset:
Dropped variables sae02 = SAE narrative, sae003 = SAE history specify.

SCQF1 dataset:
Dropped variables scfsta = staff id, scf05asp = Hispanic other specify, scf06asc = Asian other specify, scf06bsc = Black other specify, scf06esp = Racial background other specify, scf06c = Native Hawaiian/Pacific Islander., scf04 and scf04a – dropped transgender information due to small sample, scf06bsp = Black ethnic background.

SCQF2 dataset:
Dropped variables scf15ksp = Reason for clinic visit other specify, scf18ksp = Other drug used past six months specify.

SCQF3 dataset:
Dropped variables scf21asp = Inform consent other specify judgment of staff, scf21bsp = Informed consent not obtained other specify, scf22osp = Did patient complete ACASI other specify, scf22rsp = Judgment of clinical staff other specify, scf23asp = Failed to randomize judgment of clinical staff other specify, scf23bsp = Failed to randomize other specify.

SCQR1 dataset:
Dropped variables scrsta = staff id, scr05asp = Hispanic other specify, scr06asc = Asian other specify, scr06bsc = Black other specify, scr06esp = Racial background other specify, scr04 and scr04a – dropped transgender information, scr06c = Native Hawaiian/Pacific Islander., scr06bsp = Black ethnic background.

SCQR2 dataset:
Dropped variables scr15ksp = Reason for clinic visit other specify, scr18ksp = Other drug used past six months specify.

SCQR3 dataset:
Dropped variables scr21asp = Inform consent other specify judgment of staff, scr21bsp = Informed consent not obtained other specify, scr22osp = Did patient complete ACASI other specify, scr22rsp = Judgment of clinical staff other specify, scr23asp = Failed to randomize judgment of clinical staff other specify, scr23bsp = Failed to randomize other specify.

SCR and SFI datasets:
Dropped the entire dataset as it just contained patient identifier. SFI was empty anyway.

SMR2 dataset:
Dropped variables sm201 = staff id, sm209 = SMR2 Comments, sm203asp = interim visit specify text, sm205gsp = other visit specify 1, sm205hsp = other visit specify 2, sm205isp = other visit specify 3, sm206hsp = clinical diagnosis other specify 1, sm206isp = clinical diagnosis other specify 2, sm206jsp = clinical diagnosis other specify 3, sm207gsp = in house test other specify 1, sm207hsp = in house test other specify 2, sm207isp = in house test other specify 3.

SMR22 dataset:
Dropped variable sm208c = treatment given other, specify.

SMRF dataset:
Dropped variables smf01 = staff id, smf10 = SMRF Comments, smf03asp = interim visit other specify.

SMRM dataset:
Dropped variables smm01 = staff id, smm11 = SMRM Comments, smm03asp = interim visit other specify.

VIS dataset:
Dropped variables vis01 = staff id, vis06 = Six month visit comment, vis03rs = other primary reason 6 month visit not completed, vis03sp = STD clinic staff reason 6 month visit not completed

ACASI Data:

The ACASI dataset is called "weeklyftp." It is described in the two spreadsheets named "ACASI_DD_v7.2 STI extension 04.23.10_clean.xls" and "ACASI_DD_6mth v5.2 STI extension 04.22.10_clean.xls." Site identification and all date variable parts, including month, day, and year components, were dropped from

the dataset. Also, variables dem11sp = specify other insurance, adm03 = clinic, adm04 = reviewer's initials, hxt02sp = other reason never been tested for HIV, dem01a/dem02/dem03 = transgender identifier variables were dropped. Variable dem01 = gender was modified so that transgender information was removed.

IVRS Data:

The IVRS data containing the treatment assignment can be found in a dataset called "random" which has the fields described in the document called "random_dataset.doc."

RATER Data:

The Rater datasets are described in the electronic data capture annotation found in the file, "aware_ratercdd_2010_05_11.pdf." Rater datasets CI, SCR, and ENR were dropped as they contained only the counselor id. All of the participant ID variables have been replaced with the DEIDNUM variable. The Rater datasets therefore consist of these files: frpr2, frri1, frri2, frrn1, frrn2, frrr1, frsr2, and prs1. The counselor id is found in a variable called COUNSELOR.

WESTBLOT Data:

The Westblot datasets are described in the file, "westblot_datasets.docx." Only the Westblot assessment comments had to be dropped from the westblot dataset.