Contents of this document:

1. Contents of the SAS Database zip file

rawdata - contains data as collected or entered directly from the data collection forms/instruments.

- crf.xpt – SAS transport file containing all the raw SAS datasets and formats. Contains data from the CRF as well as non-CRF sources.

analysis_datasets - contains analysis data derived from the raw data.

- analysis.xpt – SAS transport file containing all the analysis SAS datasets and formats.
- residual.xpt – SAS transport file containing SAS analysis datasets for RMR, TEE and AREE residuals

2. Installing SAS transport files

The data are provided in the form of SAS transport files, which can be converted to SAS datasets using the XPORT engine and PROC COPY:

1) Place the SAS transport files (crf.xpt, analysis.xpt and residual.xpt) in a location with directory path '*transport-file-path'*.
2) Designate a location for the SAS datasets, and create a folder for the raw SAS datasets with directory path '*/sasdata path/rawdata'*.
3) Create a folder for the analysis SAS datasets with directory path '*/sasdata path/anadata'*.
4) Sample code to import raw datasets:

```
libname xraw xport 'transport-file-path/crf.xpt';
libname sasraw '/sasdata path/rawdata';

proc copy in=xraw out=sasraw;
run;
```

5) Sample code to import analysis datasets:

```
libname xana xport 'transport-file-path/analysis.xpt';
libname sasana '/sasdata path/anadata';

proc copy in=xana out=sasana;
run;
```

6) Sample code to import analysis datasets for residuals, into same directory as other analysis datasets:

```
libname xrsd xport 'transport-file-path/residual.xpt';
libname sasana '/sasdata path/anadata';

proc copy in=xrsd out=sasana;
run;
```

7) The formats for raw and analysis datasets are included as a SAS dataset in each of the transport files, and will be converted to SAS datasets after running steps 4 and 5 above. To convert these datasets to formats, designate a location for the format library, with directory path *'format-library-path'* and use PROC FORMAT with the cntlin option to convert the datasets to formats in the designated format library. The formats datasets for the raw and analysis datasets are identical, so it is only necessary to do this once, and use the same format library for raw or analysis datasets.

```
libname library 'format-library-path';

proc format cntlin=sasraw.formats library=library;
run;
```

3. Contents of the ASCII Database zip file

raw_ascii_data

- Contains ASCII (.csv) files of all raw datasets

analysis_ascii_data

- Contains ASCII (.csv) files of all analysis datasets

4. Opening ASCII files

The files in the ASCII database folders are comma separated (.csv) text files. To open the files in MS EXCEL, with a separate column for each field, following these steps:

i. First, open the desired CSV.
ii. Select column A.
iii. On the Data tab, choose Text to Columns. This step opens the Convert Text to Columns Wizard.
iv. Step 1 – accept the default selection. The data is delimited. Click "Next".
v. Step 2 – choose "Comma" as the delimiter. Unselect any other delimiter choice. Click "Next".
vi. Step 3 – accept the default column data format selection, "General". Click "Finish".
vii. Voila! The data is reformatted so that each variable's data appears in its own column.

5. Raw Datasets

The database consists of raw and analysis datasets. Raw SAS datasets contain data reported on the CRF, data received from central labs (e.g., safety and outcome labs, DLW, DXA), electronic files of data collected from instruments (e.g., RMR, Core temperature, Cognitive Bias test, CANTAB), in the original database structure. Raw datasets often have stacked formats that are not convenient for analysis, and data from a single assessment may be divided into multiple raw datasets. Documentation of raw datasets, including information about dataset structure, variables, formats, etc. can be found in the CRF and non-CRF documents available in the Database Documentation Forms section of this website**.** The data dictionary for raw datasets, also found in the Forms section, has a complete listing of all raw datasets and variables.

6. Analysis Datasets

Analysis datasets are SAS datasets derived from the raw datasets, often with multiple raw datasets merged together, with analysis variables derived according to the study data handling rules and transformed to an analyzable structure, usually with 1 record per DEIDNUM / VISIT. Documentation for analysis datasets, including information about dataset structure, definitions and derivations of variables, formats, etc. is in the Analysis Dataset Details document.

It is recommended that analyses are done using the Analysis datasets instead of the raw datasets, because

the analysis datasets have a structure designed for analysis and outcome variables derived according to the study data handling rules.

7. Important Data Usage Notes

a. Data were collected for all subjects who attended at least one in clinic screening visit (n=1069). 831 subjects were screened out, and the remaining 238 signed the informed consent and started baseline evaluations. 18 dropped out during the baseline period, and the remaining 220 were randomized. 2 of the randomized subjects did not start intervention, and the remaining 218 subjects started intervention. Most datasets contain data for all subjects who had baseline evaluations, including all randomized subjects, and some who dropped out before randomization.

The variable SUBJECT1.RAND should be used to identify subjects who were randomized, and SUBJECT1.INTERVEN for subjects who started intervention, for analyses limited to these populations.

b. Most analysis datasets have one record per DEIDNUM / VISIT, where VISIT denotes the major study visits defined by the protocol: Baseline 1, Baseline 2, Month 6, Month 12, Month 18 and Month 24. Each of these major VISITs is composed of one or more SUBVISITs. Additionally, there are 'minor' VISITs for safety evaluations (Weeks 2, 6, 8, Months 1, 3, 4, 5), some of which were deleted from the protocol early in the trial. Some assessments were also performed at screening visits (e.g. clinic weight). Most assessments are done once per VISIT, but a few (e.g. clinic weight, RMR, DXA) are done more than once a VISIT, in which case the dataset will have an additional variable to distinguish records (e.g. SUBVISIT, RMRVISIT).

c. The protocol specifies which assessments are done at each visit (see schedule of assessments). Assessment schedules depend on the randomized treatment arm. For example, DLW (among many other assessments) is done once at each baseline visit, and each of Months 12 and 24 for all subjects, and at Months 6 and 18 for subjects in the CR arm only.

d. For assessments that are done at both baseline visits, or multiple times during the baseline period, the analysis dataset includes an additional record coded as VISIT=0, which represents the overall Baseline mean, and contains the mean of each variable over all baseline records, in addition to the records for each individual baseline visit.

e. When performing analyses using the analysis datasets, make sure to keep records only for the Visits of interest to the analysis, for example, keep records with VISIT = 4 and 5 if separate records are needed for each baseline evaluation. Otherwise, if a single baseline record is needed, just keep VISIT=0 for overall baseline mean.

f. The table on the following pages shows the raw and analysis datasets for each 'domain' of data that was collected in the study, along with the dataset structure, important analysis variables, and visits, coded using the following Visit codes.

g. Visit Codes:

0=Baseline mean (combines all records with VISIT=4 or 5, using the mean value of each variable over all baseline measurements)
1=SV1 (Screening Visit 1)
2=SV2 (Screening Visit 2)
3=SV3 (Screening Visit 3)
4=BL1 (Baseline submission 1 – covers up to 3 sub-visits)
5=BL2 (Baseline submission 2, covers up to 4 sub-visits)
6=Randomization
6.1=Week 2 (deleted from protocol early in the study)
7=Month 1
7.1=Week 6 (deleted from protocol early in the study)
7.2=Week 8 (deleted from protocol early in the study)
8=Month 3
8.1=Month 4 (deleted from protocol early in the study)
8.2=Month 5 (deleted from protocol early in the study)
9=Month 6 (covers 4 subvisits)
10=Month 9
11=Month 12 (covers 5 subvisits)
11.1=Month 17
12=Month 18 (covers 3 subvisits)
12.1=Month 23
13=Month 24 (covers 5 subvisits)

See detailed definitions of visits and Sub-visits in the Visits documentation file.

8. **Table of important data domains**. This table summarizes information on raw and analysis datasets and variables for some of the important 'domains' of data collected in the trial. Detailed definitions and derivations of variables are in in the Analysis Dataset Details document**.** Additional information is in the data handling rules documents.

Domain	Raw dataset(s)	Analysis Dataset(s)	Analysis dataset structure, key variables	VISITs in analysis dataset (see appendix for codes)	Important variables
Randomized Treatment	*Comes from IVRS system, not included in the raw datasets.*	**IVRSRAND**	1 record per DEIDNUM for each randomized subject	N/A	TX
Study visit details	All CRF datasets	**VISITS**	1 record per DEIDNUM / VISIT, contains information on subject attendance at each visit, visit dates, etc.	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	
Demographic variables, randomization and intervention status, Disposition	CRF/DEMOG, HEIGHT, RANDOM, CONSENT, PERMDISC, STDYCOMP	**SUBJECT1**	1 record per DEIDNUM for each subject who had at least one screening visit. RAND =1 for randomized subjects INTERVEN=1 for subjects who started intervention	N/A	RAND, INTERVEN, TXCOMP, AGE, GENDER
Clinic Weights – measured at each clinic visit (SUBVISIT), multiple times per VISIT.	CRF/WEIGHT, DATEHDR WGHTLOSS (empty dataset, not included in public database) Non-CRF/EXTRAWT	**CLWTLONG**	1 record per DEIDNUM / SUBVISIT (multiple records per DEIDNUM / VISIT).	1, 4, 5, 7, 8, 9, 10, 11, 11.1, 12, 12.1, 13 (both TX arms)	
		CLWTVIS	1 record per DEIDNUM / VISIT, mean of all clinic weights at each visit. Includes added VISIT 0, for baseline mean.	0, 1, 4, 5, 7, 8, 9, 10, 11, 11.1, 12, 12.1, 13 (both TX arms)	MCLINWT, MBMI
Daily home weights (during each DLW period)	CRF/HWGHLOG1, HWGHLOG2	**HOMEWT**	1 record per DEIDNUM/ VISIT/ HWGHTDT/ HWTKG	CR: 4, 5, 9, 11, 12, 13 AL: 4, 5, 11, 13	HWTKG
DXA (Body Composition - %body fat, FM, FFM,BMD, etc)	CRF/DXASCAN LABS/DXA	**DXAA**	1 record per DEIDNUM / VISIT. Includes added VISIT 0, for baseline mean, and VISIT 9 for mean of two Month 6 assessments (9.1 and 9.2).	CR: 0, 4, 5, 9, 9.1, 9.2, 11, 12, 13 AL: 0, 4, 5, 11, 13	BTOTPF, FM, FFM (and *PF, *FAT, *FFM, *BMD, *BMC for each region)
Food diaries (self-reported nutrition intake during each 2 week DLW period)	CRF/FOODRCD LABS/FOODNDS	**FOODWEEK**	1 record per DEIDNUM / VISIT. Includes added VISIT 0, for baseline mean.	CR: 0, 4, 5, 9, 11, 12, 13 AL: 0, 4, 5, 11, 13	KCAL, TFAT, TPROT, TCARB, PCTFATC, PCTPROTC, PCTCARBC, PCTALCC (and numerous micronutrients)
DLW (Doubly Labeled Water analysis to measure TEE)	CRF/DLWHDR, DLWCHT, LABS/DLW (not included in public database, but all records and variables are in DLWLONG)	**DLWLONG**	1 record per DEIDNUM / VISIT / DLWSMPNO (separate record for each of 8 samples during each DLW period, with raw isotope data. Not used for most TEE analyses).	CR: 4, 5, 9, 11, 12, 13 AL: 4, 5, 11, 13	ISO2H, ISO18O,
		TEERQ	1 record per DEIDNUM / VISIT. Also uses data from DXA, FOODWEEK, CLWTVIS, HOMEWT to calculate individual RQ and TEE. Use TEERQ.TEERQ for official TEE	CR: 0, 4, 5, 9, 11, 12, 13 AL: 0, 4, 5, 11, 13	ISODILNH, ISODILNO, RCO2P, RQ, TEERQ

Domain	Raw dataset(s)	Analysis Dataset(s)	Analysis dataset structure, key variables	VISITs in analysis dataset (see appendix for codes)	Important variables
			value. Includes added VISIT 0, for baseline mean.		
TEE Residuals (TEE adjusted for FM, FFM, age and sex)		**TEERESID**	1 record per DEIDNUM / VISIT	CR: 0, 9, 11, 12, 13 AL: 0, 11, 13	TEERESID
RMR (Resting Metabolic Rate)	CRF/RMR, LABS/RMRLOAD	**RMRA**	1 record per DEIDNUM / VISIT / RMRVISIT. RMR was done twice during VISIT 5 (Baseline 2) Includes added VISIT 0, for baseline mean.	CR: 0, 5, 9, 11, 12, 13 AL: 0, 5, 11, 13	RMR, RQ_RMR
RMR Residuals (RMR adjusted for FM, FFM, age and sex)		**RMRRESID**	1 record per DEIDNUM / VISIT	CR: 0, 9, 11, 12, 13 AL: 0, 11, 13	RMRRESID (Primary Endpoint)
Physical Activity Level (PAL) Also known as Activity Related Energy Expenditure (AREE)		**PAL (derived from TEERQ and RMR)**	1 record per DEIDNUM / VISIT	0, 4, 5, 9, 11, 12, 13	PAL, PA
AREE Residuals (Activity Related Energy Expenditure, adjusted for FM, FFM, age and sex)		**AREERESD**	1 record per DEIDNUM / VISIT	CR: 0, 9, 11, 12, 13 AL: 0, 11, 13	AREERESD
Core Temperature	CRF/ADMIT, LABS/CORETEMP	**CORETMPA**	1 record per DEIDNUM / VISIT.	5, 9, 11, 13 (both TX arms)	CORETEMP, CORTEMPD, CORTEMPN
PAR (7 day Physical Activity Recall, done twice during each 2 week DLW period)	CRF/PARHDR, PARCHT, PARQ	**PARVISIT**	1 record per DEIDNUM / VISIT. Includes RMR data to calculate total activity calories. Includes added VISIT 0, for baseline mean.	CR: 0, 4, 5, 9, 11, 12, 13 AL: 0, 4, 5, 11, 13	PARACTHR, PAREXHRS, TOTMETS, PARCAL1, PARCAL2, PAREXCAL, PARTDEE
Vital signs (heart rate, BP, etc)	CRF/VITALS	**VITALSA**	1 record per DEIDNUM / VISIT Includes added VISIT 0, for baseline mean.	0, 4, 5, 7, 8, 9, 10, 11, 13 (both TX arms)	MEANWST, MEANUMB, PULSE, TEMP, RESP, MEANSBP, MEANDBP, MEANBP, PULSEPRS
VO_2 max	CRF/VOMAX	**VO2MAX**	1 record per DEIDNUM / VISIT	5, 11, 13 (both TX arms)	PVOMEAS1, PVOMEAS2, VOMEAS1, VOMEAS2
Handgrip strength	CRF/HANDGRIP	**HANDGRPA**	1 record per DEIDNUM / VISIT	5, 11, 13 (both TX arms)	
Isometrics - strength	CRF/ISOMETRC	**ISOMETRA**	1 record per DEIDNUM / VISIT	5, 11, 13 (both TX arms)	
Outcome lab data	LABS/OUTCLAB CRF/OUTCMELB	**OCLABLNG**	1 record per DEIDNUM / VISIT / ASSAY (based on LABS/OUTCLAB only) (stacked with a separate record for each assay)	5, 9, 11, 11.1, 12, 13 (both TX arms)	ASSAY, RESULT
		OCLABFLT	1 record per DEIDNUM / VISIT Merges OCLABLNG and CRF/OUTCMELB	0, 9, 11, 11.1, 12, 12.1, 13 (both TX arms)	All assay results (ADIPOHMW – ODINDEX)

Domain	Raw dataset(s)	Analysis Dataset(s)	Analysis dataset structure, key variables	VISITs in analysis dataset (see appendix for codes)	Important variables
			(flat, with a separate variable for each assay)		
Safety lab data	CRF/SAFETYLB, LABS/SAFETY	**SFLABLNG**	1 record per DEIDNUM/VISIT/COLDT/TESTCODE (stacked with a separate record for each assay)	4, 6.1, 7, 7.1, 7.2, 8, 8.1, 8.2, 9, 10, 11, 12, 13 (both TX arms) (unscheduled safety labs have missing visit)	TESTCODE, RESULTN, RESULTC
		SFLABFLT	1 record per DEIDNUM / VISIT (flat, with separate variable for each assay)	0, 7, 8, 9, 10, 11, 12, 13 (both TX arms)	All assay results (AMMB – YSTUA)
BDI (Beck Depression Inventory)	CRF/BDIQ DEPRESS (empty dataset, not included in public database)	**BDI**	1 record per DEIDNUM / VISIT	4, 7, 8, 9, 10, 11, 12, 13	SUMBDI
BSQ (Body Shape Questionnaire)	CRF/BSQ	**BSQA**	1 record per DEIDNUM / VISIT	4, 9, 11, 13	BSQSCORE
Derogatis Interview for Sexual Function	CRF/DISFEM1–DISFEM4 DISMALE1 – DISMALE3	**DEROGATI**	1 record per DEIDNUM / VISIT	4, 9, 11, 13	DRGSCR1 – DRGSCR5, DRGSCORE
FCI (Food Craving Inventory)	CRF/CRAVE	**FCI**	1 record per DEIDNUM / VISIT	4, 9, 11, 13	CARBS, SWEETS, FATS, FASTFOOD
FCQ-State (Food Cravings Questionnaire –State)	CRF/FCQSTATE	**FCQSTATA**	1 record per DEIDNUM / VISIT	4, 9, 11, 13	DESIRE, ANTPOS, ANTNEG, LACKCTRL, HUNGER
FCQ – Trait (Food Cravings Questionnaire – Trait)	CRF/FCQTRAIT	**FCQTRTA**	1 record / DEIDNUM	4	INTENT, POSANTC, NEGANTC, CONTROL, THOUGHTS, HUNGER_T, EMOTIONS, CUES, GUILT
MAEDS (Multiaxial Assessment of Eating Disorder Symptoms)	CRF/ MAEDS, EDISORD	**MAEDSA**	1 record per DEIDNUM / VISIT	4, 8, 9, 11, 12, 13	TDEP, TBNG, TPRG, TFEARFAT, TRST, TAVD
POMS (Profile of Mood States)	CRF/POMS	**POMSA**	1 record per DEIDNUM / VISIT	4, 9, 11, 13	TENSION, DEPRESS, ANGER, VIGOR, FATIGUEP, CONFUSE, DISTURB
PSQI (Pittsburgh Sleep Quality Index)	CRF/PSQI1, PSQI2	**PSQI**	1 record per DEIDNUM / VISIT	4, 9, 11, 13	PSQISCR1 – PSQISCR7, PSQISCOR
PSS (Perceived Stress Scale)	CRF/PSS	**PSSA**	1 record per DEIDNUM / VISIT	4, 9, 11, 13	PERSTRESS
Rand SF-36	CRF/RANDSF1 – RANDSF3	**RANDSF36**	1 record per DEIDNUM / VISIT	4, 9, 11, 13	PFSCORE, RLPHSCOR, RLEPSCOR, EFSCORE, EWBSCORE, SFSCORE, PAINSCOR GHSCORE
TFEQ (Eating Inventory)	CRF/TFEQA, TFEQB	**TFEQ**	1 record per DEIDNUM / VISIT	4, 9, 11, 13	RESTRAIN, DISINHIB, PHUNGER, FLEXRSTR,

Domain	Raw dataset(s)	Analysis Dataset(s)	Analysis dataset structure, key variables	VISITs in analysis dataset (see appendix for codes)	Important variables
					RIGDRSTR, HABITDIS, SITUADIS, EMOTDIS, INTHUNGR, EXTHUNGR
WELQ (Weight Efficacy Lifestyle Questionnaire)	CRF/WELQ	**WELQA**	1 record per DEIDNUM / VISIT	4, 9, 11, 13	NEGEMOT, AVAILABL, SOCPRESS, PHYSDISC, POSACT, GLSCORE
CANTAB (Cambridge Neuropsychological Test Automated Battery)	Non-CRF/CANTAB	**CANTABA**	1 record per DEIDNUM / VISIT	5, 9, 11, 13	
Cognitive Bias Assessment	Non-CRF/ENCODING, STROOP, POLYSEME, WORDSTEM	**COGBIAS**	1 record per DEIDNUM / VISIT	5, 9, 11, 13	
Adverse Events	CRF/AELOG	**AE**	1 record / DEIDNUM / AESPEC / AESTRDT		AE_LLT, AE_PRFT, AE_SOC
Serious Adverse Events	CRF/SAEFORM	**SAE**	1 record / DEIDNUM / SAESTRDT		SAE_LLT, SAE_PRFT, SAE_SOC
DTH (Delayed-type Hypersensitivity)	CRF/DTHADM1, DTHADM2	**DTH**	1 record per DEIDNUM / VISIT	5, 11, 13	
ECG	CRF/ECG	**ECGA**	1 record per DEIDNUM / SUBVISIT	4, 6.1, 7, 7.1, 7.2, 8, 8.1, 8.2, 9, 10, 11, 12, 13	
Prescribed Caloric Intake	*Not officially collected on any form, but collected on CTS system*	**RXCAL**	1 record per DEIDNUM		RXCAL
Biopsy collection details	CRF/BIOPSY,CONSENTD	**BIOPSYA**	1 record per DEIDNUM / VISIT	5, 11, 13	
Hematology / Anemia details	LABS/SAFETY, CRF/ANEMIA1 – ANEMIA3	**HEMATOL**	1 record per DEIDNUM / VISIT / HCOLDT (1 record for each safety lab draw, including non-protocol draws)	4, 7, 8, 9, 10, 11, 12, 13 (non-protocol draws have missing visit)	ANEMIA
	Analysis datasets were not created for the following datasets. Use raw datasets.				
Medical History	CRF/MEDHIST	**N/A**			
Physical Exam	CRF/PEXAM	**N/A**			
Prior Medications	CRF/PRIORMED	**N/A**			
Concomitant Medications	CRF/CONMED1, CONMED2	**N/A**			
Protocol Deviations	CRF/PDEVIATE	**N/A**			