

Introduction to using the **CALERIE** Public Use Database



Duke Clinical Research Institute

FROM THOUGHT LEADERSHIP
TO CLINICAL PRACTICE

Outline

- Where to find information
- Data download instructions
- Database installation
- Raw and Analysis datasets
- Important Data usage notes
- Major data domains



Where to find information / get help

- Go to calerie.duke.edu
- In the 'Quick Navigation' panel on the left, click on Database Documentation
 - Guide to using the Database (similar to this presentation)
 - Data Contents:
 - Evaluation schedule
 - Visit codes
 - Rawdata Contents: datasets and variables in raw database
 - Analysis Data Contents: datasets and variables in analysis database
 - Analysis Dataset Details: detailed derivations and value lists of analysis dataset variables



Where to find information / get help (cont'd)

- Data Handling Rules : rules for derivation of key study endpoints
eg, RMR, Core body Temperature, QOL scoring algorithms,
Adherence to CR, etc.
- Forms: CRFs (with and without annotation), and documentation for
non-CRF data.

... or send questions to calerie@dm.duke.edu



Data Download Instructions

- Go to calerie.duke.edu
- In the 'Quick Navigation' panel on the left, click on 'Apply for Samples & Data Analysis'
- At bottom of page click on 'Click here to register and download the public database'
- Fill out database submission form: email address, name, institution, department, and submit.
- Click on 'ASCII Database' or 'SAS Database' to download `ascii_database.zip` (15.7 MB) or `sas_database.zip` (19.2 MB)



Install SAS database

- Create or designate a folder for the data.
- Click on downloaded sas_database.zip → SAS_database folder contains 2 folders:
 - Rawdata contains raw data transport file → crf.xpt
 - Analysis_data contains analysis data transport files → analysis.xpt and residual.xpt
- Highlight the desired data folder (Analysis_data, Rawdata or SAS_database for all data) and extract to the designated folder.
- Instructions and sample code for installing SAS transport files and formats using proc copy are in the guide to using the public database
- Note special steps for installing SAS data formats



Install ASCII data files

- Create or designate a folder for the data.
- Click on downloaded ASCII_database.zip → ASCII_database folder contains 2 folders:
 - Raw_ascii_data contains (84) raw data csv files
 - analysis_ascii_data contains (52) analysis data csv files
- Highlight the desired data folder (analysis_ascii_data, raw_ascii_data or ASCII_database for all data) and extract to the designated folder.
- Instructions for converting csv files to Excel spreadsheets are in [guide_to_using_the_public_database.pdf](#)



Raw Datasets

- Raw datasets correspond to data collected on CRFs, data received from central labs (eg Blood chemistry, DLW, DXA), electronic files of data collected from instruments (eg RMR, Core temperature) in its original database structure.
- Raw dataset variables can be mapped to annotated CRFs (or documentation for non-CRF data sources) in Database Documentation / Forms / CALERIE CRF Annotated Form pages 1-155 etc.
- Raw data structure is not convenient for analysis – eg data for a single assessment might be split into multiple raw datasets that need to be merged, no derived endpoints, etc.
- Raw datasets will not be useful for most users.



Analysis Datasets

- Analysis datasets are derived from raw datasets according to the study data handling rules and transformed to an analyzable structure.
- Most analysis datasets have 1 record per Subject per study Visit*.
- Analysis dataset variables are documented in Database Documentation / Data Contents / Analysis Dataset Details
- Use Analysis datasets (not raw)!



Analysis Populations

- The database includes all subjects who attended at least one in clinic screening visit (n=1,069), but only 220 were randomized and 218 started the intervention.
- Make sure to use only the relevant subject population for an analysis!
 - Randomized subjects: SUBJECT1.RAND=1
 - Subjects who started intervention: SUBJECT1.INTERVEN=1
 - Randomized treatment arm: IVRSRAND.TX='A' (CR)
'B' (AL)



Study Visits

- Major Study Visits:
 - Baseline (4-6 weeks before randomization)
 - Month 6*
 - Month 12
 - Month 18*
 - Month 24

*at months 6 and 18, subjects in AL arm had limited set of evaluations.

- Major Study Visits consist of one or more Sub-Visits. Eg, there were up to 6 Sub-visits at Baseline
- Study assessments were done one or more times at Baseline and at some or all follow-up study visits. See evaluation schedule.



Study Visits (cont'd)

- Some assessments (eg Weight, RMR, DXA, DLW) were done multiple times in a study visit.
 - Eg DLW, DXA and RMR were performed 2x at Baseline for all subjects
 - DXA was performed 2x at Month 6 for CR subjects
- For evaluations performed more than once at a Major Study Visit, the analysis dataset has a record for each individual evaluation, AND another record for the average of all evaluations at that Visit.



Study Visits (cont'd)

- eg The Analysis dataset TEERQ (TEE from DLW analysis) has:
 - 1 record for Baseline 1 DLW data (VISIT=4)
 - 1 record for Baseline 2 DLW data (VISIT=5)
 - 1 record for Baseline Mean DLW data (VISIT=0) (average over BL1 and BL2)
- Make sure to include only the relevant visits in your Analysis!
- You probably don't want to include all 3 Baseline records!



VISIT codes

- 0=Baseline (mean across all Baseline visits)
- 1-3 : Screening Visits 1-3 (weights only)
- 4=BL1 (Baseline 1 – covers up to 3 sub-visits)
- 5=BL2 (Baseline 2 – covers up to 4 sub-visits)
- 6=Randomization
- 7=Month 1
- 8=Month 3
- 9=Month 6 (covers 4 subvisits)
- 10=Month 9
- 11=Month 12 (covers 5 subvisits)
- 12=Month 18 (covers 3 subvisits)
- 13=Month 24 (covers 5 subvisits)



Key Variables

- DEIDNUM – unique subject identifier, exists in all datasets.
- VISIT – visit identifier. Most datasets have one record per DEIDNUM/VISIT, and can be merged by DEIDNUM/VISIT
- Randomized Intervention Arm: IVRSRAND.TX



Some important Analysis Datasets

- SUBJECT1: subject level data 1 record per DEIDNUM
- IVRSRAND.TX: randomized treatment arm

The following datasets have 1 record per DEIDNUM / VISIT (with separate records for Baseline 1, Baseline 2, and Baseline mean)

- CLWTVIS: Clinic Weights, (averaged over multiple Sub-Visits in each VISIT)
- DXAA: Body Composition via DXA
- TEERQ: TEE via DLW
- RMRA: Resting Metabolic Rate
- COREMPA: Core Body Temperature



Important Analysis Datasets (cont'd)

- OCLABFLT : Laboratory assays
- PCTCR : Calculated % Caloric Restriction based on TEE from DLW and changes in FM and FFM from DXA
- And many more!
- See summary of important data domains in Guide to the public database

