The HEALTHY Study Data Archive

The HEALTHY Archive contains the Protocol, Manual of Operations, Forms, and Study data. The files are organized into the following directories:

- Documents
- Forms
- Data
- Data Integrity Check (DSIC)

HEALTHY Documents

This directory contains:

- HEALTHY public website protocol v1_4 071408.pdf (Study Protocol)
- Repository Documentation.pdf (Data Documentation provided by the HEALTHY data coordinating center).
  - This file contains detailed descriptions of the legacy datasets, including file structure, data source, data de-identification, and statistical considerations.

HEALTHY Forms

This subdirectory contains 7 annotated data collection forms in PDF format:
- ST2 Family Information Repository.pdf
- ST3 Health Screening Repository.pdf
- ST4 Pubertal Development Scale Repository.pdf
- ST6 SAPAC Repository.pdf (Physical Activity Checklist)
- ST7 Health Utilities Index Repository.pdf
- ST8 Health Thermometer Repository.pdf
- ST9 20 Meter Shuttle Test Repository.pdf

Per instructions from the Data Coordinating Center (DCC), each form is released as a PDF for use in approved data-release analyses only. The forms provided have been modified to match what is provided in datasets. Variable names for each released data set are embedded in blue on the data form. Other notes about annotated HEALTHY forms:

- Coding and formats for all variables are found on the original data form except where described below.
- The numerical value for check-box style categorical variables is noted to the lower right of the check-boxes on the form.
- Open-ended information entered into text fields is omitted in the release datasets.

HEALTHY Data

The HEALTHY data directory contains the following raw datasets in *SAS, v.7-8-9* format:

- CBL.sas7bdat[1] -- there is no form associated with the CBL data; a list of variables is provided in <Repository Documentation.pdf>.

- Rx.sas7bdat -- includes one record for each school indicating the randomization assignment for that particular school. If *assign=0* then the school was randomized to *control*. If *assign=1* then the school was randomized to *intervention*.

- ST2.sas7bdat – corresponds with ST2 Family Information Repository.pdf

- ST3.sas7bdat-- corresponds with ST3 Health Screening Repository.pdf

- ST4.sas7bdat -- corresponds with ST4 Pubertal Development Scale Repository.pdf

- ST5.sas7bdat – corresponds with the ST5 Block Food Frequency Questionnaire; the form is not provided due to copyright restrictions. Instead, a list of variables is provided in <Repository Documentation.pdf>.

- ST6.sas7bdat -- corresponds with ST6 SAPAC Repository.pdf

- ST7.sas7bdat -- corresponds with ST7 Health Utilities Index Repository.pdf

- ST8.sas7bdat -- corresponds with ST8 Health Thermometer Repository.pdf

- ST9.sas7bdat -- corresponds with ST9 20 Meter Shuttle Test Repository.pdf

Variables Common to All Datasets

- SCHOOLID: This is used to link all records in a particular school and is unique to each school.
- STUDENTID: This identification number consists of a three-digit study ID number (985), the two-digit SCHOOLID number, and a randomly generated five-digit identification number, all separated by dashes. The final five-digit identification number will uniquely identify each participant even after all preceding numbers are removed. STUDENTID is used to link subject-specific information across datasets. Both the full identification number (3-digit number–2-digit number–5-digit number) and abbreviated identification number (the final 5-digit number) are unique to each participant.

---

[1]Datasets were originally provided as SAS CPORT transport files, as described in <Repository Documentation.pdf>, pp 2-3. We have converted each transport file to a SAS version 7-8-9 dataset. The original transport files are on file at the Repository, available upon request to approved users of the data.

- GRADE: This is a single-digit number that identifies the time at which the measures were taken. This combined with STUDENTID is used to match a participant's information across multiple forms completed for that time period. GRADE is coded as:
    o **6** if the data was collected at baseline (the first semester of 6th grade, fall 2006).
    o **7** if the data was collected as a part of the interim assessment (second semester of 7th grade, spring 2008).
    o **8** if the data was collected at the end of the study (second semester of 8th grade, spring 2009).

HEALTHY Data Integrity Check

The HEALTHY_DSIC directory contains a report of an examination of legacy datasets by statisticians and quality control specialists at the Repository. Datasets were checked for completeness, consistency, and usability. The published data from the HEALTHY study were compared to values recalculated from the HEALTHY data in the NIDDK repository.