SHOW 2008-2013 Public Use Dataset Documentation
Updated May 2022

Table for Contents for this document:

1. BOX folders and their contents

There are 2 primary folders you have access to:

1_Documentation

2_Data

An overview of what is within each folder:

1_Documentation

- **2011_Annotated_Questionnaires:** We shared just one year (2011) of surveys that was most representative of all years 2008-2011. It includes all surveys as asked during the survey year, including items not included in the public use data set. It also includes the formats and variable names associated with each survey item. Use CTRL + F to search for keywords or use as a reference to help with understanding the data set and survey questions and responses.

- **Paper 1. Nieto et al. 2010 & Paper 2. Malecki et al. 2022:** For detailed information on the sampling design, recruitment methods and participation of the 2008-2013 study sample, please see the Nieto et al. 2010 and the Malecki et al. 2022 methods paper.

- **PUBUSE08_13_DER_PUBUSE:** This is an HTML codebook that contains the survey variable names, labels, brief descriptions and frequencies or means of the variables that are in the public use datasets provided,
  In order to open and view the codebook:
  (1) Download the codebook html file
  (2) Open it in a new website browser

- **SHOW Instruments Time Table_2008_2013:** This indicates when variables were collected and the total sample size for each part of the survey. To minimize participant burden, data collection was spread out over three separate time periods. Dropoff in response was observed; be aware that missing responses may be due to the timing of the administration of the instrument and not necessarily due to participants

intentionally skipping questions. Refer to the Nieto et al., 2010 methods paper for more information on how the survey is conducted.

SHOW Visit Label	Description	Maximum N
Time 1	In-home visit, using computer-assisted interviews	3380
Time 2	Self-administered questionnaire left with the participant to mail back to SHOW	2963
Time 3	Body Measurements and Blood Sample Collection, Additional Questionnaires (approximately 2-4 weeks after the in-home visit?)	2940

Instruments not included (to be included in a later release):
Health Questionnaire (HHQ),
Diet (DIQ),
Discrimination (QG)

Instruments not included:
Oral Health (OHQ),
Foreclosure (FOR)
Cardiovascular Health Index (CVH)

- **SHOW_Public_Use_Dataset_2008_2013_Data_Dictionary_Final:** This is a searchable spreadsheet of all variables in the dataset, variable label, response options and their SAS labels, and their SAS formats. Use CTRL + F to search for keywords or use as a reference to help with understanding the data set and survey responses.

- **SHOW_Public_Use_Dataset_2008_2013_Deleted_Variables:** Several variables were deleted – see masking process and analytic notes below as to why this is the case.

- **SHOW_Public_Use_Dataset_2008_2013_Modified_Variables:** Several variables were modified for protection of participants and to keep the data set non-identifiable – see masking process and analytic notes below as to why this is the case.

2_Data

- Three different versions of the same data set are provided:
  - SAS
  - .csv with formats
  - .csv without formats

- Please use the SAS code "How to read in formats" in order to apply the show_formats sas data file to the SAS data set.

2. Analytic notes for 2008-2013 public use data

All data in SHOW 2008-2013 public use datasets were masked and deidentified. The following outlines the steps SHOW used to create the public use dataset.

Data Masking Process:

- All data are deidentified. Study participants were randomly issued a participant ID number (NEW_ID).
- Geographical data were limited (either grouped or not included).
- Data were grouped categorically (i.e. response options were collapsed).
    - Data were grouped to include >=20 participants.
    - In some cases of non-identifiable data, cell sizes are <20.
    - When data with cell sizes <20 were not able to be grouped, variables were not included.
- Intermediate variables were dropped when derived summary variables were available.
- Questions with small sample sizes were dropped.
- Questions from instruments that were collected at only one or two time periods were not included.
- Continuous data that may be identifiable were categorized. Categories could be based on useful cut points (i.e. age by 5-year increments), distribution (i.e. number of health conditions), quintiles (i.e. how much did you weight a year ago?). Anthropometry variables (BMI, weight, waist) were winsorized at 1% and 99%.
- PROC RANK was used to assign quintiles to some continuous variables.
- Free text responses were dropped.
- Data that was modified from the original for the public use dataset have an "_MOD" extension added to the variable name.

- Variables with an "_R#" extension indicate that something changed in the collection of the original data, either: question wording changes, response options changes, targeted population who answered the question changes, etc. We combined as many of the questions that made sense to combine. Be aware that it may not be appropriate to combine variables with an "_R#" with the non-"_R#" version.

- All responses that were originally coded as refused (.R) or don't know (.D) were recoded to missing (.).

Sampling Weights

The SHOW core data was collected using a sampling frame across the state of Wisconsin. Due to the deidentified nature of the data, no sampling weights were provided. Data is still usable; the difference is that statements indicating that results are weighted to be representative of the residents of state of Wisconsin cannot be made. What can be said is that SHOW surveyed the state of Wisconsin; responses represent WI residents who participated in SHOW.

Clustering

All eligible adults within a selected household were invited to participate in SHOW. There are participants in SHOW that who have one more additional household members who also participated in SHOW and in the data set. For some analyses, correlations may exist among household members, and this may violate the assumption of regression analyses that all events are independent of one another. The variable Household Identification (HHID) is available upon request and will require a data request and IRB approval.

3. Data use terms and agreement

You are agreeing to the following same terms and conditions when accessing and using this public use data. Should you have any questions or concerns regarding these terms, please contact us at researchers@show.wisc.edu .

1. I agree to allow SHOW staff to contact me via the provided email address for any and all of the following reasons: the status of my project, routine program updates and any data or metadata updates.

2. I agree to cite the "Survey of the Health of Wisconsin" as a source of the data in all presentations and publications proceeding from the proposed study, and will include the following acknowledgement:

   > Funding for the Survey of the Health of Wisconsin (SHOW) was provided by the Wisconsin Partnership Program PERC Award (233 AAG9971). The authors would also like to thank the University of Wisconsin Survey Center, SHOW administrative, field, and scientific staff, as well as all the SHOW participants for their contributions to this study.

3. I agree to not disclose or publish data whereby a sample unit or survey respondent could be identified or related to any particular individual, family or household. This includes not publishing or disclosing data on survey responses with a N less than or equal to 5.

4. I understand that failure to adhere to these terms by me or anyone on my research team will be deemed non-compliant with SHOW policies as well as UW-Madison rules and regulations and at minimum could result in the loss of the opportunity to use SHOW data in the future.

4. Accessing 2008-2013 non-public use data

Non-public, restricted use variables from 2008-2013 include those that were modified (categorized or collapsed responses) from being continuous or were removed from the public use data set. You may request restricted use 2008-2013 data by submitting a data request via our consultation form: **https://show.wisc.edu/data/** **All requests for non-public, restricted use variables require IRB approval letter and application.**

These additional variables are FREE:

- Household Identification number (HHID)
- Sampling weights, strata and cluster variables

These data are NOT subject to the $3000+ restructured use data request fee if, but subject to the $103/hr analytic rate:

If a specific research aim has been sufficiently analyzed with the public use data and additional variables or unmodified variables are needed. For example, final regression models and results have been run and presented, but you need continuous variable, not collapsed responses, for publication. Another example may be that you need a specific health outcomes variable in from health history questionnaire HHQ in order to run your analyses.

These data are subject to the $3000+ restricted use data request fee:

If a specific research aim has not been sufficiently identified and analyzed with the public use data provided first.