

Analytic Notes for the SHOW COVID-19 Impact Survey data (Waves I, II, III) Public Use Datasets

This document is focused on the steps taken to create the public use datasets for waves 1-3 of the SHOW Covid-19 Impact Survey.

Are sampling weights available for the COVID-19 data?

Sampling weights have not been calculated and will not be available for this study because a subset of participants was enrolled in such a way (non-probability sampling of underserved populations in the City of Milwaukee in 2018-2019 were included) that it was not possible to calculate sampling weights for them. Therefore, these datasets are more well suited to address research questions focused on the associations between outcomes and determinants of health, as opposed to generating representative estimates of the adult population of Wisconsin.

Final sample sizes:

The final sample size for your analysis will likely be less than the total for each wave's dataset. For example, wave I of the survey had n=1,403 total participants, but because some participants did not complete the entire survey, you will likely have some missing responses for variables of interest and end up with a smaller n. Also, in addition to the 1,403 participants who took the Wave I survey online, there are 55 participants who completed a shortened telephone version of the study. Non-public use datasets for the telephone-collected data should be available upon request soon.

Codebooks:

Within the "4_Data" folder, each wave of COVID-19 data collection contains an HTML codebook.

In order to open and view the codebook:

- (1) Download the codebook html file
- (2) Open it in a new website browser

The codebooks contain the survey variable names, labels, brief descriptions and frequencies or means of the variables that are in the public use datasets provided. For Waves II and III, they also list "REMOVED_VARIABLES" that were dropped from the public use datasets and "MODIFIED_VARIABLES" that were altered to protect participant privacy. The original unmodified form of these variables may be available upon request. For Wave I, see the COVID19_Wave1_deleted_modified_variables_list.docx.

For each wave, there is also an Excel-based data dictionary that lists the variable names, labels, and format values.

Variable notation:

Survey questions that remained identical in wording throughout all waves have the exact same variable name in each wave.

If a survey question's wording, response options or skip pattern changed from one wave to another, an "_R2" was added to the variable name to notate this change (if another change was made in subsequent wave, an "_R3" was added to the variable name to notate this). Review the annotated PDFs for the wording changes between original and "_R2" and "_R3" versions of a variable. Some "_R2" extensions can easily be dropped from the variable name before merging. See this example below:

Survey Wave	Survey question	Variable name
Wave I:	Have you ever been tested for COVID-19?	CV19B010
Wave II:	Have you ever been tested for COVID-19 using a nasal swab or saliva sample?	CV19B010_R2
Wave III:	Have you ever been tested for COVID-19 using a nasal swab or saliva sample?	CV19B010_R2

In other cases, original and “_R2” versions may not be combined across waves without more consideration. For example:

Survey Wave	Survey question	Variable name
Wave I:	Do you have the same number of individuals living in your house as before COVID-19 began? Valid responses: Yes, No	CV19N001
Wave II:	Compared to six months ago, do more, less, or the same number of individuals live in your house? Valid responses: More, Less, The same	CV19N001_R2
Wave III:	Compared to six months ago, do more, less, or the same number of individuals live in your house? Valid responses: More, Less, The same	CV19N001_R2

For survey questions that have identical wording across waves, but refer to timepoints or time durations unique that survey wave, variable names are notated with “_W2” to indicated Wave II timeframe and “_W3” to indicate Wave III timeframe. See an example of this below:

Survey Wave	Survey question	Variable name
Wave I:	Have you had to use public transportation to get to work, get groceries or other travel since COVID-19 began?	CV19F005
Wave II:	Have you had to use public transportation to get to work, get groceries or other travel since July 1, 2020?	CV19F005_W2
Wave III:	Have you had to use public transportation to get to work, get groceries or other travel since February 1, 2021?	CV19F005_W3

Changes made across survey waves:

Not all survey questions were asked in all survey waves. In the BOX folder “3_Annotated Surveys” you can determine what waves a survey item was asked by:

- (1) viewing the annotated survey instrument – the variable name is listed with each survey question
- (2) viewing the excel spreadsheet “Across Wave Question Comparison” – this spreadsheet lists all survey question (and uses Wave III survey item working when possible), to notate what waves a survey item was asked in. When a survey question was asked in more than one wave, only the most recent wave is

listed for the purpose of this overview. It is not meant to be a comprehensive list of variables (ie. in the example above, only CV19F005_W3 is listed in the spreadsheet to notate that this item was asked in all three waves, it does not list CV19F005_W2 & CV19F005).

Public use data changes to protect privacy:

To help protect privacy of SHOW participants while still allowing maximum flexibility in allowing researchers to access these data, the following changes were made to create limited datasets. If your needs require a different approach, please contact SHOW to see what options might be available.

Changes required were in three categories:

1. Open-Ended Text Responses were removed and dropped from public use
2. Sensitive Demographic of Health-Related Responses with fewer than 5 responses
 - a. When possible, responses were moved and included in a catch-all other category
 - b. Waves 1 & 2: When no catch-all other category existed data and responses were fewer than 5 then the variable was removed. In Wave 3, variables were dropped when cell sizes were < 20.
 - c. When no catch-all other category existed and responses were fewer than 5 (or 20 for Wave 3) for specific categories, those responses were set to missing.
3. Household Structure Data and Demographic Categories were Collapsed or Removed to Decrease the Likelihood of Identifying Specific Families

We ask all researchers using this limited COVID-19 Web Survey dataset to not publish or report out on survey responses with a $N \leq 5$. If you notice a set of features that should be restricted in the dataset, please contact data@show.wisc.edu so we can make sure we can continue to offer limited datasets that allow for quicker analyses by research partners while still respecting the privacy of the SHOW Participants.