

UnderStandingAmericaStudy

UAS COVID-19 LONGITUDINAL FILES DATA DESCRIPTION



USC Dornsife Center for Economic and Social Research
Last Updated July 7, 2020

Contents

1. Introduction	2
2. Overview of the UAS Covid-19 Longitudinal Files.....	3
2.1 Obtaining UAS Covid-19 Related Data.....	3
2.2 Input Files.....	3
2.3 Sampling Frame and Fielding Process of the UAS Covid-19 Surveys	6
2.3.1 National Sample.....	6
2.3.2 Los Angeles County Sample.....	6
2.4 UAS Covid-19 Longitudinal Files Data Formats	6
2.5 Helpful Links.....	7
2.6 Merging with other UAS Data Sets	7
3. UAS Covid-19 Longitudinal Files Description.....	8
3.1 Variable Renaming.....	8
3.1.1 Specific Variable Renaming Changes	9
3.2 Longitudinal Files Content	9
3.2.1 UAS Covid Survey Questionnaire Data	10
3.2.2 My Household Survey Data	10
3.2.3 Derived Variables	10
3.2.4 Survey Information Variables.....	11
3.3 Sample Weights.....	11
References.....	12
Appendix A. Variables with Crosswave Content Differences.....	13
Table A.1 Variable Content Differences in the Longitudinal Files	13

1. INTRODUCTION

This document describes the UAS Covid-19 National Sample Longitudinal File and the UAS Covid-19 Los Angeles County Sample Longitudinal File of the Understanding America Study (UAS). These files were created at the University of Southern California's Center for Economic and Social Research (CESR). Main support for this data set has been provided by the University of Southern California and the Bill & Melinda Gates Foundation.

The UAS Covid-19 National Sample Longitudinal File and the UAS Covid-19 Los Angeles County Sample Longitudinal File (known collectively as the Covid longitudinal files) are comprised of online survey data collected by CESR through its UAS internet panel. The UAS is a nationally representative panel of American households randomly recruited from United States Postal Service delivery sequence files. UAS respondents are individuals age 18 and older who answer surveys via an online interface that is technologically powerful yet friendly for the participants and quick to deliver results. More information about the methodology of the UAS is available from the [UAS data pages](#)

The data in the Covid longitudinal files consist of appended data sets from the UAS *Understanding Coronavirus in America* ("Covid") tracking surveys of attitudes and behaviors around the Novel Coronavirus pandemic in the United States. The first survey was launched on March 10, 2020, and includes a national bi-weekly long-form questionnaire and a weekly Los Angeles County short-form questionnaire administered in each bi-weekly wave. The Covid longitudinal files are updated bi-weekly, as each wave becomes final¹.

The Covid longitudinal files have a consistent variable naming structure so that the content associated with each variable is consistent from wave to wave in the data. Appending the individual UAS Covid survey data sets required renaming variables from their original names if they were used in different surveys to ask different questions. To assist users of the Covid longitudinal files in mapping original UAS Covid survey variables to their counterparts in the longitudinal files, the [UAS Covid-19 Survey Longitudinal Files Crosswalk](#) was created. This document connects variable names from each UAS Covid survey to their names in both the national sample longitudinal file and the Los Angeles County sample longitudinal file.

The details contained in this document pertain to both the UAS Covid-19 National Longitudinal Sample File and the UAS Covid-19 Los Angeles County Sample Longitudinal File, except where differences are explicitly stated.

The Covid-19 longitudinal files can be linked with any other UAS survey. How to perform that linkage is described later in this document. An updated listing of all available UAS data sets is available on the [UAS All Surveys page](#).

¹ The structure of the first wave differs from later waves; see Section 2.3.1 for details

Please send all questions about UAS data sets or this data description to uas-l@mymaillists.usc.edu.

2. OVERVIEW OF THE UAS COVID-19 LONGITUDINAL FILES

2.1 Obtaining UAS Covid-19 Related Data

The data described in this document are based on UAS public data release files. These files are accessible through the [UAS Covid-19 National Survey Data](#) and the [UAS Covid-19 Los Angeles County Survey Data](#) webpages.

Before accessing any UAS data, permission must first be obtained by [registering on the UAS site](#). Registering with UAS is considered to be agreeing to the “conditions of use” governing access and usage of the data. Completion and submission of an additional form is required to access UAS Covid-19 related data. These steps are described in detail on the [UAS Covid-19 Data Access](#) page.

2.2 Input Files

The current Covid longitudinal files are updated with data from the individual UAS Covid surveys as they are concluded, a new wave every two weeks. Table 1 lists the concluded waves through July 7, 2020.

Table 1. Fielding Dates for UAS Covid Surveys in the Longitudinal Files

Wave	UAS Surveys	Launch Date	Closing Date
1	230	3/10/2020	3/31/2020
2	235, 236	4/1/2020	4/28/2020
3	240, 241	4/15/2020	5/12/2020
4	242, 243	4/29/2020	5/26/2020
5	244, 245	5/13/2020	6/9/2020
6	246, 247	5/27/2020	6/23/2020
7	248, 249	6/10/2020	7/7/2020

Each UAS Covid-19 survey is classified as either a long-form or a short-form survey. Long-form surveys are administered to the national sample of respondents (including Los Angeles County residents) on a bi-weekly basis. Long-form surveys consist of more questions and more topics

than short-form surveys, thus generally take longer for respondents to complete. The data from the long-form surveys are used to create the UAS Covid-19 National Sample Longitudinal File.

Short-form surveys are administered on a weekly basis to Los Angeles County residents only, and contain a subset of questions from the long-form surveys. In addition, the short-form surveys may contain questions not found in the long-form surveys, mainly pertaining to topics related specifically to the greater Los Angeles area. The data from the short-form surveys, combined with the data from long-form surveys solely pertaining to Los Angeles County residents, are used to create the UAS Covid-19 Los Angeles County Sample Longitudinal File.

Included in each longitudinal file are variables that indicate the survey number that corresponds to the UAS survey source for each data observation (variable name: *survey_source*), as well as the wave number associated with all data from each specific UAS Covid survey (variable name: *wave*).

The national sample longitudinal file consists of data from the long-form survey administered during each wave of data collection.

Los Angeles County sample respondents are administered two surveys during each wave of data collection (1 long-form and 1 short-form survey), therefore the structure of the wave variable in the Los Angeles County longitudinal file differs from the structure of the wave variable in the national sample longitudinal file. Tables 2A and 2B illustrate this difference. [Section 2.3](#) of this document explains the differences in the two samples in greater detail.

Table 2A. National Sample Longitudinal File Information (Example Only²)

Wave Number	UAS Survey	Survey Type
1	230	Long-form
2	235	Long-form
3	240	Long-form
4	242	Long-form

Table 2B. Los Angeles County Sample Longitudinal File Information (Example Only)

Wave Number	UAS Survey	Survey Type
1	230	Long-form
2.1	235	Long-form
2.2	236	Short-form
3.1	240	Long-form
3.2	241	Short-form
4.1	242	Long-form
4.2	243	Short-form

Only 1 survey was administered in Wave 1

Table 2A shows the numbering of the wave variable in the national sample longitudinal file as an integer that represents the one long-form survey collected during each wave. Table 2B shows that the numbering of the wave variable in the Los Angeles County longitudinal file is broken down into 2 parts:

- The number to the left of the decimal point represents the overall wave of data collection.
- The number to the right of the decimal point, in which a value of 1 indicates the long-form survey data from the wave, and a value of 2 indicates the short-form survey data from the wave.

² The current UAS Covid-19 longitudinal files contain more than 4 waves of data. Tables 2A and 2B demonstrate the wave numbering and corresponding survey information for descriptive purposes only.

2.3 Sampling Frame and Fielding Process of the UAS Covid-19 Surveys

2.3.1 National Sample

The UAS Covid-19 National Sample Longitudinal File consists of data from respondents of the Covid long-form surveys.

The first survey was fielded on March 10, 2020. The survey was in the field until April 1, 2020. In contrast to later waves, all respondents were invited on March 10. The distribution of responses over the survey period is therefore not random and concentrated in the first part of the survey period (see the Survey Methods tab on the [Covid-19 Pulse site](#)). As of April 1, 2020, a new survey is fielded every two weeks. Each day one fourteenth of the respondents are invited to take the survey. Since respondents have two weeks to answer the survey, the total field period is 4 weeks, so that responses during the last two weeks of a field period of one survey overlap with responses in the first two weeks of the subsequent survey. Respondents are paid an additional incentive to complete their surveys on their assigned day, and as a result, more than 90% of responses to each wave are completed in the first 14 days.

2.3.2 Los Angeles County Sample

The UAS Covid-19 Los Angeles County Sample Longitudinal File consists of data from Los Angeles County residents in the March 2020 survey, from respondents of the UAS Covid short-form surveys, and from long-form survey respondents who reside in Los Angeles County.

The March 2020 wave of data for Los Angeles County is simply a subset of the national data, only including Los Angeles County residents. As of April 1, 2020, the data collection for Los Angeles County consists of two components. The first component is the national survey administered to Los Angeles County residents once every fourteen days. The second component consists of a shorter survey (short-form) with the same 14 days periodicity, but administered in the alternative weeks. To give an example, if a Los Angeles County respondent is invited to a survey on Monday, April 6, for the national survey then the same respondent is invited for a shorter survey on Monday, April 13. Conversely, a respondent who is invited for a national survey on Monday, April 13, will be invited for a short survey on Monday, April 6, etc. Since the content of the short survey is mainly a subset of the questions in the national survey, the result is that for the short survey, we collect data at a weekly frequency in Los Angeles County, rather than every other week. This set-up implies that a respondent in Los Angeles County has only one week to answer a survey, rather than two weeks as in the rest of the country.

2.4 UAS Covid-19 Longitudinal Files Data Formats

The longitudinal files are available for download in STATA, CSV, and SAS format from the [National Survey Data page](#) and [Los Angeles County Survey Data page](#). If you would like the data in a different format, please email your request to uas-l@mymailists.usc.edu.

2.5 Helpful Links

For easy reference, the following are links to pages and files that are mentioned in this document:

[UASDATA Covid-19 home page](#)

[Data Access page](#): Lists the steps for how to gain access to UAS survey data, and also specifically, UAS Covid-19 survey data.

[National Survey Data page](#): Contains links to the national sample longitudinal file download and documentation, each individual Covid survey questionnaire, codebook, and file download, as well as links to reports that use national survey data.

[Los Angeles County Survey Data page](#): Contains links to the Los Angeles County sample longitudinal file download and documentation, each individual Covid Los Angeles County survey questionnaire, codebook, and file download, as well as links to reports that use Los Angeles County survey data.

[UAS Covid-19 Survey Longitudinal Files Crosswalk](#)

[UAS All Surveys page](#): Contains the complete list of all UAS surveys that have been administered, and provides the codebook, data download link, and other information about each UAS survey.

[My Household Survey Data page](#): Contains information about the My Household survey, which is demographic and socioeconomic information updated quarterly by each UAS participant. More information about My Household is included in [Section 3.2.2](#).

[UAS Standard Variables page](#): Lists the full set of descriptive variables that are included in each data set.

[UAS Sample Weights page](#): Provides detailed information about UAS weighting procedures

2.6 Merging with other UAS Data Sets

The Covid longitudinal files can be merged with any other UAS data set by using the unique person identifier variable, *uasid*. This variable, assigned to a respondent at recruitment into the UAS panel, stays fixed for each survey taken. The longitudinal files are panel data sets, meaning that multiple records are likely for each *uasid*; one for each instance a respondent completed a Covid UAS survey.

For more information about default identification variables included in each survey, such as household identifier, *uashhid*, please visit the [UAS standard variables page](#).

3. UAS COVID-19 LONGITUDINAL FILES DESCRIPTION

The UAS Covid-19 National Sample Longitudinal File and the UAS Covid-19 Los Angeles County Sample Longitudinal File both contain data from survey questions covering many aspects of how people are being affected by the coronavirus pandemic. In addition, derived variables, survey information variables, and sample weights are included in the files. This section describes the situations that require original UAS Covid survey variables to be renamed in the longitudinal files, and also provides some detail about the content included in the longitudinal files.

3.1 Variable Renaming

Efforts are made when creating each longitudinal file to preserve the original variable names from each UAS Covid data set. However, there are circumstances that require original variable names to be changed to assure consistent naming of identical questions across the individual surveys. These circumstances are:

1. The same question and response options have different original variable names in two or more UAS Covid surveys.

For example, original variable *cr005* from UAS 230 and original variable *cr023* from UAS 235 both ask:

“On a scale of 0 to 100 percent, what is the chance that you will get the coronavirus in the next three months? If you’re not sure, please give your best guess.”

Therefore, both variables have been renamed to *prisk_infection* in the longitudinal files.

2. Different question text and/or response options have the same original variable name in two or more UAS Covid surveys.

For example, *cr002* from UAS 230 asks:

“Has a doctor or another healthcare professional diagnosed you with the coronavirus (COVID-19)?”

While *cr002* in UAS 235 asks:

“Have you been tested for coronavirus? If so, what was the result?”

In this case, *cr002* from UAS 230 has been renamed to *covid_diagnosed* and *cr002* from UAS 235 has been renamed to *cor_tested* in the longitudinal files.

To quickly identify the longitudinal file variable name equivalent to an original Covid survey variable name, please use the [UAS Covid-19 Survey Longitudinal Files Crosswalk](#). The crosswalk has separate sections for the national sample file and the Los Angeles County-only sample file. This document will be updated as each new wave of data is added to the longitudinal files.

When question text or response options differ only slightly, the original variable name is preserved in the longitudinal files. Each instance is reviewed by the UAS team manually and a decision is made on whether to preserve the original variable name or to rename the original variable into two new variable names that reflect the different content from each survey. For cases where the decision is made to preserve the original variable name despite a slight difference, a note is attached to the variable in the longitudinal data sets that explains the difference and identifies in which UAS survey it first occurred.

A complete list of variables in the longitudinal files with slight differences in crosswave question text or response options (but maintaining the same variable name) is given in Table A.1 of [Appendix A](#). Table A.1 identifies the wave of data collection for UAS Covid surveys in which the change occurred. This includes situations where the sole difference of a variable between two waves is an added answer option. For example, the variable *beh_facemask* in the longitudinal files asks whether you have worn a facemask in the past seven days to keep yourself safe from the coronavirus. In UAS 230, the answer options are Yes or No. In UAS 235, the answer options are Yes, No, Unsure.

3.1.1 Specific Variable Renaming Changes

- For original survey questions cr031a through cr031d, in UAS 235 there were six answer options. For these same variables, starting in UAS 240 the number of answer options was reduced to four. As a result, the data in the longitudinal files for these questions from UAS 235 is stored in the variables named *cor_percp_uas235_1* through *cor_percp_uas235_4*, and for UAS 240 and forward the variables are named *cor_percp_1* through *cor_percp_4*.
- Also for questions cr031a through cr031d, a small number of respondents using the Spanish language questionnaire were inadvertently provided with a six-value answer option that had been replaced by a four-value option in the English language questionnaire, a change that was missed during translation. We were unable to reliably map the six values to the four values, so have removed and replaced them with a missing value of ".d Deleted value" in the published UAS data files as well as in the UAS Covid longitudinal files. Access to the removed responses is available upon request. Starting with UAS 248, all respondents see the four-value response option. These variables have the names *cor_percp_1* through *cor_percp_4* in the longitudinal files.

3.2 Longitudinal Files Content

The variables included in each longitudinal file can be classified into 4 main groups: UAS Covid surveys questionnaire data, UAS My Household survey data, variables derived from UAS Covid survey questionnaire data, and information about the surveys themselves. An overview of each follows.

3.2.1 UAS Covid Survey Questionnaire Data

The items asked in the UAS Covid surveys aim to measure the impact of the coronavirus pandemic on physical and mental health, personal finances, employment, education, outlook on the future, and many other aspects of daily life. To view the individual questions asked in each Covid survey, please use the links to the survey codebooks questionnaires, and reports found on the [National Survey Data page](#) and [Los Angeles County Survey Data page](#).

The content of the Los Angeles County short-form surveys is primarily a subset of the larger long-form national surveys. However, there are some items that are only asked in the short-form surveys, such as questions that pertain specifically to residents of Los Angeles County.

In the national sample and Los Angeles County sample longitudinal files, if a variable is not included in a wave, its values have been set to missing code “.z variable not in survey” to simplify any missing data analysis performed by data users.

Currently, no special outlier detection process is performed while creating the longitudinal files. Future versions may incorporate data cleaning techniques to identify and flag highly influential data observations.

3.2.2 My Household Survey Data

The My Household survey is administered quarterly to all UAS respondents. To provide background information about the respondent and household, the survey asks about key demographics, including age, ethnicity, education, marital status, work status, state of residence, family structure, and many others.

The most recent My Household survey response data at the time each respondent completes a UAS survey is included in the data set of that UAS survey. This data is included in the Covid longitudinal files corresponding to each wave in which the respondent has participated.

For more information about the My Household survey, please visit the [My Household Survey Data page](#).

3.2.3 Derived Variables

This section describes the derived variables that were constructed specifically for the Covid longitudinal data sets by using original UAS Covid survey data. Currently, only one related group of derived variables is included in the files. Future releases of the longitudinal files may contain additional derived variables.

- The individual UAS Covid surveys each include questions from the Patient Health Questionnaire-4 (PHQ-4). The PHQ-4 is a brief-screening tool used by clinicians for detecting both anxiety and depressive disorders (Kroenke, K. et al., 2009). The questions are presented in the UAS Covid surveys with answer choices 1 to 4. However, in the original PHQ-4 the answer choices are labeled 0 to 3 and then summed to arrive at a total score. To facilitate scoring of the UAS Covid PHQ-4 data, each response is first re-scaled

by subtracting 1. These derived re-scaled variables have the names: *phq4_anxious_r*, *phq4_worry_r*, *phq4_depressed_r*, and *phq4_littleinterest_r*. These variables are then summed to create the derived variable, *phq4_score*.

3.2.4 Survey Information Variables

The following variables are included in both the national sample and the Los Angeles County sample longitudinal files as identifiers of each UAS Covid survey:

- *wave*: Indicator of the wave number associated with each UAS Covid survey. See [Table 2A](#) for information about how the wave number corresponds to the UAS survey number in the current national sample longitudinal file, and see [Table 2B](#) for this information pertaining to the Los Angeles County sample longitudinal file.
- *survey_source*: Indicates the UAS survey number associated with each UAS Covid survey. For example, if the data for an observation in either longitudinal file is from UAS 230, then the value of *survey_source* is 230.
- *surveytype*: This variable has a value of 1 if the UAS survey source for a data observation is a long-form survey, and a value of 2 if the UAS survey is a short-form survey. **This variable is only found in the UAS Covid-19 Los Angeles County Sample Longitudinal File.** Each observation in the national sample longitudinal file is from a long-form survey.

3.3 Sample Weights

The longitudinal files include wave-specific weights that allow each sample to be representative of the reference population along several demographic dimensions. These dimensions include gender, race/ethnicity, age, education, household size, and household income as well as census region and urban/rural characteristics of the area of residence. A complete description of the UAS weighting procedure can be found [here](#).

Values used to generate the sample weights are taken from the My Household survey information for each respondent in each wave of data.

The longitudinal files contain the weight variable:

- *final_weight*: Relative final post-stratification weights ensuring representativeness of the survey sample with respect to the U.S. population 18 years of age or older. The weights are calculated separately for the national sample and the Los Angeles County sample and are wave specific.

REFERENCES

Kroenke, K. & Spitzer, R. & Williams, J. & Löwe, B. (2009). An Ultra-Brief Screening Scale for Anxiety and Depression: The PHQ-4. *Psychosomatics*. 50. 613-21. 10.1176/appi.psy.50.6.613.

APPENDIX A. VARIABLES WITH CROSSWAVE CONTENT DIFFERENCES

Table A.1 Variable Content Differences in the Longitudinal Files

Wave	UAS Survey	Content Changes
2	UAS 235 & UAS 236	<ul style="list-style-type: none"> Change to source of data contained in variable <i>job_current</i>. This became a constructed variable, calculated from responses to <i>lr001</i>, <i>lr003</i>, <i>lr003a</i>, and <i>lr005</i>. In UAS 230, it was not constructed from several variables, but instead contained the response to a single question (which was <i>lr004</i> in the original UAS 230 survey). Change to response options for the following variables: <i>beh_avoidhighrisk</i>, <i>beh_avoidpublic</i>, <i>beh_avoidrest</i>, <i>beh_cancelldrappt</i>, <i>beh_facemask</i>, <i>beh_pleasuretravel</i>, <i>beh_pray</i>, <i>beh_socialactivities</i>, <i>beh_stockpilefood</i>, <i>beh_visitdr</i>, <i>beh_washhands</i>, <i>beh_worktravel</i>, <i>beh_workactivities</i>, <i>beh_workhome</i> The response option "Unsure" was added.
	UAS 235 only	<ul style="list-style-type: none"> Change to question text for variable <i>coping_exercise</i>. The phrase "extra exercise" in the question text was changed to simply "exercise" in UAS 235.
3	UAS 240 & UAS 241	<ul style="list-style-type: none"> Change to question text for the following variables: <i>cor_tested</i>, <i>covid_diagnosed</i>, <i>think_havecorona</i>, <i>cor_care</i>, <i>cor_care_hosp</i>, <i>cor_care_local</i>, <i>cor_care_pcp</i>, <i>cor_care_urgent</i>, <i>cor_care_where</i>, <i>cor_care_none</i>, <i>cor_care_other</i>, <i>cor_contact</i>, <i>cor_contact_comm</i>, <i>cor_contact_emp</i>, <i>cor_contact_fam</i>, <i>cor_contact_hosp</i>, <i>cor_contact_local</i>, <i>cor_contact_online</i>, <i>cor_contact_oth</i>, <i>cor_contact_pcp</i> These questions are concerned with whether respondents had experienced or informed others about experiencing being tested for, diagnosed with, infected with, or sought medical care for coronavirus. The time frame for these questions was changed to ask about any experiences since the previous interview. Previously, respondents were asked if they had ever in their lives had one of these experiences.
	UAS 240 only	<ul style="list-style-type: none"> Change to response options for variable <i>ins_randomizer_specific</i>. Options available limited to 1 (5%) to 4 (20%) available. In UAS 230, options 5 (30%) to 11 (90%) were also available.
4	UAS 242 & UAS 243	<ul style="list-style-type: none"> Change to response options for variable <i>lr003</i>. Added 2 new answer options for the employment status question asked when Respondent did not previously have a job. "4 I am retired" & "5 I am not in the labor force (not currently working and not looking for work)"

	UAS 242 only	<ul style="list-style-type: none"> Change to response options for the following variables: <i>cl010aa_1_, cl010aa_2_, cl010aa_3_, cl010bb_1_, cl010bb_2_, cl011b_1_, cl011b_2_, cl013_1_, cl013_2_, cl013_3_, r_cl006b, r_cl010aa, r_cl010bb, r_cl011b, r_cl013.</i> Added answer option "12 institution does not plan for in-person enrollment in the fall" for each of these questions.
5	UAS 244 only	<ul style="list-style-type: none"> Change to response options for variable <i>lr010</i>. Added the answer option "6 Twice a month" to the question about how often the respondent is paid by employer.