Tracking COVID-19's Effects by Race and Ethnicity: Phase 2

Technical Appendix

Alena Stern, Ajjit Narayanan, and Yipeng Su
Last updated October 26, 2020

This document describes the variable definitions and methodology behind the Tracking COVID-19's Effects by Race and Ethnicity: Phase 2 feature initially published October 26, 2020. This feature uses data from phase 2 of the federal Household Pulse Survey to measure how the COVID-19 pandemic has affected US households. It will be updated biweekly until the final phase 2 data release on November 18, 2020. The feature presents the race- and ethnicity-disaggregated averages against state, metropolitan statistical area (MSA),[1] and national averages. It allows users to quickly identify where statistically significant racial and ethnic disparities exist, as well as whether those disparities are widening or narrowing over time.

For information on the Tracking COVID-19's Effects by Race and Ethnicity: Phase 1 feature originally published by the Urban Institute on July 1, 2020, please see the associated technical appendix (PDF). For a further discussion of the differences between phase 1 and phase 2 of the Household Pulse Survey as it relates to these two features, please see the "Phase 1 and Phase 2 Differences" section starting on page 8.

Variable Definitions

We use the Household Pulse Survey (Pulse Survey) Public Use Files and construct indicator variables for each indicator in the tracker. We describe each indicator and how we computed it below, along with the universe of respondents who answered the relevant question(s). In all cases, the denominator for each indicator variable is all respondents who answered the question(s). For more detail on how we coded up these indicator variables, please check out our Github repository.

We include a broader set of variables in the data releases on the Urban Institute Data Catalog. Definitions for the variables included in the data releases but not in the feature can be found in a separate section that starts on page 12.

Credit Card/Loan Spending

Respondents were marked as having used credit card or loan spending if they reported that they or someone in their household used credit cards or loans to meet their spending needs within the past seven days.

Universe: All respondents.

Employment Income Loss

Respondents were marked as having lost income if they reported that they or someone in their household had experienced a loss of employment income since March 13, 2020.

Universe: All respondents.

Expected Employment Income Loss

Respondents were marked as expected to lose income if they reported that they or someone in their household expected to lose employment income in the next four weeks because of the COVID-19 pandemic.

Universe: All respondents.

Food Insufficiency

Respondents were marked as food insufficient if they reported that

1. the food in their household in the past week was often not enough to eat

   OR

2. the food in their household in the past week was sometimes not enough to eat.

Universe: All respondents.

Health Insurance Coverage

Respondents were marked as uninsured if they reported that

1. they did **not** have any of the following:
   » employer-provided health insurance
   » insurance purchased directly from an insurance company, including marketplace coverage
   » Medicare
   » Medicaid or any government assistance plan for people with low incomes or a disability
   » TRICARE or other military care
   » VA Health Insurance

   OR

2. they did have health insurance only through the Indian Health Service.[2]

Universe: While all respondents answered this question, we restrict our analysis to all respondents under age 65.[3] The Pulse Survey asks respondents to report their birth year, not their age. We consider all respondents born in 1956 or later as under 65. The Census Bureau uses the same definition to produce the uninsured counts available in table 3 of the Pulse Survey detailed health tables.

Mental Health

Respondents were marked as displaying signs of anxiety or depression if within the past seven days, they

1. were experiencing symptoms of anxiety, calculated by summing the responses to the following two questions based on an assigned numerical scale (not at all = 0, several days = 1, more than half the days = 2, nearly every day = 3):
   » feeling anxious, nervous, or on edge
   » not able to stop or control worrying

   If the total score was 3 or higher, then the respondent was identified as experiencing symptoms of anxiety.

   OR

2. were experiencing symptoms of depression, calculated by summing the responses to the following two questions based on an assigned numerical scale (not at all = 0, several days = 1, more than half the days = 2, and nearly every day = 3):

   » having little interest or pleasure in doing things

   » feeling down, depressed, or hopeless

   If the total score was 3 or higher, then the respondent was identified as experiencing symptoms of depression.

This definition follows the National Center for Health Statistics definition.

Universe: All respondents.

Previous Mortgage Payments

Respondents were marked as caught up on their mortgage payments if they responded that their household is currently caught up on mortgage payments.

Universe: Respondents who reported that they own their house with a mortgage or loan (including home equity loans). Respondents who reported that they own their home free and clear are excluded.

Previous Rental Payments

Respondents were marked as caught up on their rent payments if they responded that their household is currently caught up on rent payments.

Universe: Respondents who reported that they rent their home. Respondents who reported that they occupy their home without payment of rent are excluded.

Public Health Insurance Coverage

Respondents were marked as having public health insurance coverage if they reported that they have any of the following:

- Medicare

- Medicaid or any government assistance plan for people with low incomes or a disability

- VA Health Insurance

Universe: While all respondents answered this question, we restrict our analysis to all respondents under age 65. The Pulse Survey asks respondents to report their birth year, not their age. We consider all respondents born in 1956 or later as under 65. The Census Bureau uses the same definition to produce the publicly insured person counts available in table 3 of the Pulse Survey detailed health tables.

Savings Spending

Respondents were marked as having used savings spending if they reported that they or someone in their household used money from savings or sold assets to meet their spending needs in the past seven days.

Universe: All respondents.

Stimulus Payment Spending

Respondents were marked as having used stimulus payment spending if they reported that they or someone in their household used "stimulus (economic impact) payment" to meet their spending needs in the past seven days.

Universe: All respondents.

UI Benefit Spending

Respondents were marked as having used UI benefit spending if they reported that they or someone in their household used unemployment insurance (UI) benefit payments to meet their spending needs in the past seven days.

Universe: All respondents

Upcoming Mortgage Payments

Respondents were marked as having no or slight confidence they can pay their mortgage next month or having deferred payment if they reported

1. no confidence in their ability to pay their mortgage next month

   OR

2. little confidence in their ability to pay their mortgage next month

OR

3. that they had already deferred their next month's mortgage payment.

Universe: Respondents who reported that they own their house with a mortgage or loan (including home equity loans). Respondents who reported that they own their home free and clear are excluded.

Upcoming Rental Payments

Respondents were marked as having no or slight confidence they can pay their rent next month or having deferred payment if they reported

1. no confidence in paying rent next month

OR

2. little confidence in paying rent next month

OR

3. that they had already deferred their next month's rent payment.

Universe: Respondents who reported that they rent their home. Respondents who reported that they occupy their home without payment of rent are excluded.

Methodology

We update the feature biweekly when a new public use file representing another two weeks of data is released. We use R to download all public use files for phase 2 through the current release. We join each public use data file with the replicate weights file on the unique respondent identifier (scram) provided by the Census Bureau. The Household Pulse Survey public use files report data on race and ethnicity in two separate variables: rhispanic, which is 1 if respondents are not of Hispanic, Latino, or Spanish origin and 2 if they are of Hispanic, Latino, or Spanish origin; and rrace, which has four options: Asian alone, Black alone, white alone, and any other race alone, or races in combination. These data are based on recoding of questions from the Household Pulse Survey that ask for respondents to provide more detailed information on Hispanic, Latino, or Spanish origin and race. We produce a combined race and ethnicity variable with the following race/ethnicity groups that correspond to the race/ethnicity groups in the published Household Pulse Survey data tables:

- Asian alone, not Hispanic: rrace is 3 and rhispanic is not 2

- Black alone, not Hispanic: rrace is 2 and rhispanic is not 2

- Hispanic or Latino (may be of any race): rhispanic is 2

- Two or more races + other races, not Hispanic: rrace is 4 and rhispanic is not 2

- White alone, not Hispanic: rrace is 1 and rhispanic is not 2

We also append a week number column and the full state and MSA names to the public use files. We then produce Boolean variables for each metric per the definitions provided on the previous four pages. In each case, individuals who meet the definition are coded as 1, individuals who are included in the denominator but did not meet the definition are coded as 0, and individuals who are not included in the denominator (most often those who did not answer the question, though there are some exceptions as outlined above) are coded as NA. Accordingly, the percentages shown in the feature are those coded as 1 divided by those coded as 1 or 0. These data are published weekly in the Urban Institute Data Catalog.

Each data point in this feature represents two weeks of data collected by the Household Pulse Survey. This is a departure from phase 1 feature, which used the single-week releases of the Household Pulse Survey to produce two-week rolling averages (weeks 1 + 2, weeks 2 + 3, weeks 3 + 4, etc.). We find that the standard errors in a single two-week period in phase 2 are comparable to the standard errors in a two-week rolling average in phase 1.

STANDARD ERROR CALCULATION

We calculate the standard errors of our estimates following the Census Bureau's specification (see the weekly Household Pulse Survey source and accuracy statements for details). The Census Bureau created 80 replicate weights to calculate the standard error and variance of an estimate.[4] These weights can be used to calculate the variance of a statistic of interest as follows:

$$var(\hat{\theta}) = \frac{4}{80}\sum_{i=1}^{80}(\theta_i - \hat{\theta})^2$$

where

- $\hat{\theta}$ is the estimate of the statistic of interest calculated using the population weights provided by Census (pweight);

- θ_i is the replicate estimate of the same statistic calculated using one of the 80 replicate weights;

- 4 is derived from $1/\llbracket(1-f)\rrbracket\char`\^2$ where f is Fay's adjustment—accordingly, we use a Fay's adjustment of 0.5 to achieve a value of 4; and

- 80 is the number of replicate weights.

To calculate the statistics and standard errors, we use the survey and srvyr packages in R. We use the as_survey_rep() command to create the survey using the modified weights and balanced repeated replication (BRR) and the svyby and svymean functions from the survey package to calculate the means and standard errors.[5] The srvyr package does not allow Fay's adjustment to be set with BRR weights. To replicate the Fay's adjustment of 0.5, we multiply the resulting standard errors by 2. The results are consistent with the STATA svy package's BRR weights with Fay's adjustment and the Census Bureau estimates of standard errors.

We then use the standard errors calculated for each metric to calculate 95 percent confidence intervals, using the formula 95% CI = $\hat{\theta} \pm 1.96 * SE_{\hat{\theta}}$.[6] We calculate the statistical significance of the mean for a given race/ethnicity group versus the geography mean. Contrasting an overall population mean (e.g., the California mean) with a population subgroup mean (e.g., Asian population in California mean) is equivalent to contrasting the subgroup mean and the population without the subgroup (all other race/ethnicity groups in California mean). We use the svycontrast function from the survey package to calculate the mean and standard error of the difference between the race/ethnicity group mean and the mean for all other race/ethnicity groups in that geography. We then use this mean and standard error to conduct a two-sided *t*-test at the 0.05 level. Where the absolute value of the *t*-score is greater than 1.96, we reject the null hypothesis that the two means are equal.

In some instances, the feature shows overlapping confidence intervals of the population mean and subgroup mean along with a statistically significant difference of means. While it is true that two statistics with non-overlapping confidence intervals are necessarily significantly different, the converse is not true (PDF); the confidence intervals of two statistics can overlap while still rejecting the null hypothesis that the two statistics are equal. Moreover, the population mean includes members of the subgroup, so the two means are not independent. Phrased differently, the means of the subgroup and population are often closer together than the means of the subgroup and all other groups. Our calculation of statistical significance of the difference between subgroup and population mean takes this into account.

Phase 1 and Phase 2 Differences

When creating the phase 2 feature, we decided not to directly compare the phase 1 and phase 2 estimates because several key differences in the survey design make direct comparison difficult.

Questionnaire: Phase 2 of the Household Pulse Survey added many new questions, two of which (previous rent payments and previous mortgage payments) are in the feature. In addition, several questions from the phase 1 survey were not included in the phase 2 survey, including four from the Phase 1 feature (last month's rent/mortgage payment, use of stimulus on expenses, and class cancellation).

Survey Period: Phase 2 of the Household Pulse Survey has a two-week survey period while phase 1 of the Survey had a one-week survey period. As a result, phase 2 data are released biweekly while phase 1 data were released weekly.

Nonresponse Rate: Phase 2's unit nonresponse (nonresponse of a sampled individual to the entire survey) rates and item nonresponse (nonresponse of a survey respondent to a specific question) rates differ from phase 1. Because of the longer survey window, unit nonresponse for phase 2 is lower. For example, the unit nonresponse rate is 7.7 percent for week 12 (the final week of phase 1) and 10.6 percent for week 13 (the first week of phase 2). The longer response window could not only increase overall unit response rate, but qualitatively change the type of respondents who answer the survey in observable and unobservable ways. However, when we compare the racial and ethnic composition of survey respondents between weeks 12 and 13, we find a similar composition of respondents:

TABLE 1

Composition of Respondents by Race and Ethnicity

Survey week 12 versus week 13

Race/Ethnicity	Week 12 (phase 1)	Week 13 (phase 2)
Asian alone	4.2%	4.9%
Black alone	7.4%	8.0%
Hispanic or Latino	9.3%	9.4%
Two or more races or other race alone	3.8%	3.7%
White alone	75.3%	74.0%

The many new questions in phase 2 created a significantly longer questionnaire and increased item nonresponse—especially to questions later in the survey. We compare the item response rates overall and by race and ethnicity for the variables used in the feature between week 12 and week 13 in the

tables below. Each entry can be interpreted as the proportion of survey respondents of a given race/ethnicity group that answered the relevant question(s) to calculate the variable based on the definitions above.

TABLE 2

Item Response Rate by Variable

Survey week 12 versus week 13

Variable	Week 12 (phase 1)	Week 13 (phase 2)
Uninsured	90.0%	83.5%
Insured public	90.6%	84.1%
Income loss	99.4%	99.5%
Expected income loss	99.3%	99.3%
Food insufficiency	97.8%	90.9%
Depression and anxiety symptoms	91.5%	84.9%
Spending[a]	98.3%	90.2%
Tenure[b]	89.8%	82.1%

[a] All the spending variables have the same response rate because they are calculated from the same survey question.
[b] We use the response rate to the tenure question that identifies whether a respondent pays rent or mortgage to approximate the response rate to the previous rent/mortgage payments and next month's rent/mortgage payment response rate because respondents only receive those questions based on their response to the tenure question. In our analysis, nearly all respondents who answer the tenure question proceed to answer the previous and next month's payment questions.

We see in table 2 that item response for the questions on income loss—which are at the beginning of the survey—are nearly identical across the two weeks; item response for the questions on insurance, food insufficiency, spending, and housing, which appear later in the survey, are 6–8 percentage points lower. When we look at item response by race, we generally see that item response is lower for Black and Hispanic respondents than other groups across both weeks, though in week 13 we see the gap widen for the insurance, mental health, and housing questions.

TABLE 3

Item Response Rate by Variable and Race/Ethnicity

Survey week 12 versus week 13

Variable	Week 12 (Phase 1)					Week 13 (Phase 2)				
	Asian	Black	Hispanic	Two or more	White	Asian	Black	Hispanic	Two or more	White
Uninsured	89%	86%	86%	89%	91%	81.5%	76.3%	77.4%	82.5%	85.2%
Insured public	90%	87%	88%	89%	91%	82.1%	77.4%	78.7%	83.1%	85.6%
Income loss	100%	99%	99%	99%	99%	99.6%	99.4%	99.4%	99.7%	99.5%
Expected income loss	99%	99%	99%	99%	99%	99.2%	99.1%	99.1%	99.6%	99.3%
Food insufficiency	98%	97%	97%	98%	98%	88.5%	86.4%	86.7%	90.4%	92.2%

Depression and anxiety symptoms	91%	89%	89%	91%	92%	83.0%	78.5%	79.8%	84.2%	86.4%
Spending	98%	98%	97%	99%	98%	87.6%	85.9%	85.8%	89.9%	91.4%
Tenure	89%	86%	87%	89%	91%	79.8%	74.5%	76.0%	80.9%	83.9%

These disparities in item response by race and ethnicity may mean that our estimates for Black and Hispanic populations in phase 2 are likely to have larger margins of error than our estimates for other race and ethnicity groups in phase 2. Increased item nonresponse can also cause concerns if there are systematic differences between those respondents who answer all survey questions and those that drop out throughout the survey. To begin to examine these differences, we look at the differences in the average employment income loss as defined above between the respondents who did and did not answer the questions to calculate each variable in our feature. We use employment income loss because nearly all respondents in both weeks 12 and 13 answered this question.

TABLE 4

Employment Income Loss by Whether Respondent Answered Questions for Each Variable

Survey week 12 versus week 13

	Week 12 (Phase 1)		Week 13 (Phase 2)	
Variable	Responded	Did not respond	Responded	Did not respond
Depression and anxiety symptoms	41.4%	41.0%	37.4%	42.5%
Food insufficiency	41.4%	40.5%	37.7%	43.5%
Insured public	41.4%	40.9%	37.4%	42.2%
Spending	41.5%	36.1%	37.9%	41.0%
Tenure	41.4%	41.4%	37.3%	42.1%
Uninsured	41.3%	42.2%	37.3%	42.7%

Table 4 illustrates that in phase 2, those respondents who did not respond to the later questions in the survey had higher rates of employment income loss on average than those respondents who did respond to the questions, while in phase 1 the difference between respondents and non-respondents is much smaller with the exception of the spending question. We also find a similar pattern in week 14 of the phase 2 data. If employment income loss is correlated with the other variables measured in the feature, it may mean that the item nonresponse introduces bias into the estimates calculated from the survey data.

We encourage users of these data to take these limitations into account when drawing conclusions from the Household Pulse Survey estimates and before drawing direct comparisons between the phase 1 and phase 2 estimates. To view these response rate metrics for all weeks of phase 2, please see the Urban Institute Data Catalog.

Definitions of Additional Variables in Data Release

The variables below are included in the dataset release on the Urban Institute Data Catalog but are not included in the phase 2 feature.

Eviction Risk

Respondents were marked as being at risk of eviction in the next two months if they reported

1. somewhat likely that their household will have to leave their home or apartment within the next two months because of eviction

   OR

2. very likely that their household will have to leave their home or apartment within the next two months because of eviction.

Universe: Respondents who reported that they rent their home and that they are not caught up on rent payments. Respondents who reported that they occupy their home without payment of rent are excluded.

Difficulty Meeting Expenses

Respondents were marked as having difficulty paying for usual household expenses if they responded that it has been somewhat difficult or very difficult for them or their household to pay for usual household expenses in the past seven days, including food, rent or mortgage, car payments, medical expenses, student loans, and so on.

Universe: All respondents

Foreclosure Risk

Respondents were marked as being at risk of foreclosure in the next two months if they reported

1. somewhat likely that their household will have to leave their home within the next two months because of foreclosure

   OR

2. very likely that their household will have to leave their home within the next two months because of foreclosure

Universe: Respondents who reported that they own their home and that they are not caught up on mortgage payments. Respondents who reported that they occupy their home free and clear are excluded.

SNAP Spending

Respondents were marked as having used SNAP benefits to meet their spending needs in the past week if they selected "Supplemental Nutrition Assistance Program (SNAP)" in response to the question "Thinking about your experience in the last 7 days, which of the following did you or your household members use to meet your spending needs? Select all that apply."

Universe: All respondents

Telework

Respondents were marked as having engaged in telework if they selected the option "Yes, at least one adult substituted some or all of their typical in-person work for telework" in response to the question "Did any adults in this household substitute some or all of their typical in-person work for telework because of the coronavirus pandemic, including yourself? Select only one answer."

Universe: All respondents

Time Spent on Learning Activities

Respondents were marked as having student(s) in the household that spent less time on all learning activities relative to a school day before the pandemic if they reported that students spent

1. much less than a school day before the coronavirus pandemic

OR

2. a little bit less than a school day before the coronavirus pandemic.

Universe: All respondents that reported having children in the household enrolled in a public or private school

Unmet Mental Health Need

Respondents were marked as having unmet mental health service needs if they responded "yes" to the question "At any time in the last 4 weeks, did you need counseling or therapy from a mental health professional, but DID NOT GET IT for any reason? Select only one answer."

Universe: All respondents

Notes

[1] The Household Pulse Survey reports data for the 15 largest MSAs in the US, listed below. The metropolitan names in the web feature are shortened for ease of use.

Full MSA name	Shortened name
Atlanta-Sandy Springs-Alpharetta, GA	Atlanta
Boston-Cambridge-Newton, MA-NH	Boston
Chicago-Naperville-Elgin, IL-IN-WI	Chicago
Dallas-Fort Worth-Arlington, TX-OK	Dallas/Fort Worth
Detroit-Warren-Dearborn, MI	Detroit
Houston-The Woodlands-Sugar Land, TX	Houston
Los Angeles-Long Beach-Anaheim, CA	Los Angeles
Miami-Fort Lauderdale-Pompano Beach, FL	Miami/Fort Lauderdale
New York-Newark-Jersey City, NY-NJ-PA	New York
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	Philadelphia
Phoenix-Mesa-Chandler, AZ	Phoenix
Riverside-San Bernardino-Ontario, CA	Riverside
San Francisco-Oakland-Berkeley, CA	San Francisco
Seattle-Tacoma-Bellevue, WA	Seattle
Washington-Arlington-Alexandria, DC-MD-VA-WV	Washington, DC/Arlington

[2] The Census Bureau does not consider Indian Health Service coverage comprehensive, so it classifies people receiving such coverage as uninsured. See Edward R. Berchick, Jessica C. Barnett, and Rachel D. Upton, *Health Insurance Coverage in the United States: 2018*, Current Population Report P60-267(RV) (Washington, DC: US Government Printing Office, 2019).

[3] Given the universal role that Medicare plays for those over 65, we present estimates for nonelderly adults.

[4] From page 10 of the August 31 source and accuracy statement: "These methods primarily measure the magnitude of sampling error. However, they do measure some effects of nonsampling error as well. They do not measure systematic biases in the data associated with nonsampling error. Bias is the average over all possible samples of the differences between the sample estimates and the true value."

[5] In cases where the mean is 0 percent or 100 percent (all respondents in the subgroup are coded 0 or all respondents in the subgroup are coded 1 for the given indicator and week interval), the calculated standard error will be 0. These cases appear in the feature as a point without an associated margin of error.

[6] While it is mathematically possible for the confidence interval to extend above 100 percent or below 0 percent, we clip the confidence interval at 0 and 100 in the feature for ease of reading.

Acknowledgments

For more information on this project, see urban.org/features/tracking-covid-19s-effects-race-and-ethnicity-phase-two.

ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org