URBAN INSTITUTE · ELEVATE · THE · DEBATE

Tracking COVID-19's Effects by Race and Ethnicity

Technical Appendix

Alena Stern and Ajjit Narayanan
Last updated July 1, 2020

This document describes the variable definitions and methodology behind the Tracking COVID-19's Effects by Race and Ethnicity feature initially published July 1, 2020. This feature uses data from the federal Household Pulse Survey to measure how the COVID-19 pandemic has affected US households. It will be updated weekly. 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.

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.

Class Cancellation

Respondents were marked as having classes cancelled if they reported that classes normally taught in person at their child's school were cancelled.

Universe: All respondents with at least one child that attends public or private school. Respondents who reported no children in their household were excluded from the denominator for this metric.

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.

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.

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.

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 weekly when a new public use file representing another week of data is released. We use R to download all public use files through the current week. We join the week's 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: white alone, Black alone, Asian 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 metropolitan statistical area (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, individual 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.

The data in the feature reflect two-week rolling averages (weeks 1 + 2, weeks 2 + 3, weeks 3 + 4, etc.). We use two-week rolling averages to obtain more precise estimates at each point in time, owing to

the considerable standard errors in some cases for the estimates when disaggregated by state/MSA and race or ethnicity.

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 $1/\ [\![(1-f)]\!]$^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. Based on guidance received from Census Bureau researchers, to calculate the standard errors of the two-week rolling averages, we first divide the population weights and replicate weights by 2 (to calculate a *n*-week rolling average, you would divide the weights by *n*). We then 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; two estimates with overlapping confidence intervals may be significantly different. 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.

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 age 65, we present estimates for nonelderly adults.

[4] From page 11 of the June 9 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

This research was funded by the Urban Institute. The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

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

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org

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.