

2020 Collaborative Multiracial Post-Election Survey (CMPS)

Principal Investigators: Lorrie Frasure, Janelle Wong, Matt A. Barreto, Edward D. Vargas

Data Extract I (September 9, 2021 Release)

Methodology Statement: Primary Samples (PS) only (September 9, 2021 Release)

A total of 14,988 completed interviews were collected online in a respondent self-administered format from April 2, 2021 to August 25, 2021. The initiation date of the survey was determined, in part, in response to national events unfolding in real-time in the aftermath of the 2020 election. We sought to allow events to unfold and include respondent reactions to post-election disturbances in December 2020 and January 2021, given the unprecedented events our nation witnessed.

The survey (and invitation) was available to respondents in English, Spanish, Chinese (simplified), Chinese (traditional), Korean, Vietnamese, Arabic, Urdu, Farsi, and Haitian Creole. Because of the primary interest in the 2020 election, the project started with a large sample of registered voters from online sources that were pre-matched to the voter file. In addition, the data include a sample of non-registered adults, including non-citizens. Overall sample sizes were:

	PS Total	Latino	Black	Asian	White, not-Hispanic
Registered to vote	9,779	2,239	2,634	2,632	2,274
Not registered	5,209	1,767	1,371	1,343	728
PS Total	14,988	4,006	4,005	3,975	3,002

The full data are weighted within each racial group to fall within the margin of error of the adult population in the 2019 Census ACS 1-year data file for age, gender, education, nativity, and ancestry. A post-stratification raking algorithm was used to balance each category within +/- 2 percent of the ACS estimates. Data are not weighted to their national *combined* racial average. That is, Whites account for 20 percent of all cases, and each racial

group roughly 27 percent. If users want to create nationally representative racial composition from the data they can consult the latest ACS data file. Excluding “other” and just counting these four racial groups, the Census ACS 2019 1-year data for the adult population is estimated to be 64% White, not Hispanic, 17% Latino, 13% Black, 6% AAPI.

We recommend that you use the weight variable for *all analysis*, whether multivariate regression or simple cross tabulations. If you are using Stata, please use:

tab Q14 race [aw=weight], col nof

The code above will tabulate *weighted* presidential vote choice by race, showing column percentages by race, but not raw frequency counts.

If you are using R, please see the full code on the last page appendix below.

In spring 2020, scholars were invited to collaborate on the 2020 Collaborative Multi-Racial Post-Election Survey (CMPS). The goal of the project was to build upon the 2016 CMPS which was the first cooperative, 100% user content driven, multi-racial, multi-ethnic, multi-lingual, post-election online survey in Race, Ethnicity and Politics (REP) in the United States. The survey’s main focus is on attitudes about the 2020 Election and candidates, experiences with racism, policy attitudes, immigration, and personal experiences with civic engagement across many facets of American life.

This 2020 CMPS includes over 200 scholars across nearly 100 different colleges/universities, including Contributors who purchased survey content and Collaborators who purchased data access, prior to the release of the individual level data extract. Survey questions were user-generated. Users who contributed survey content could submit questions for just one single racial group, or common questions across all racial/ethnic, or oversample groups, depending on their interest. In cases where two different users submitted very similar questions the PIs worked to create a single common question. Overall, the survey contained over 800 unique questions including split samples, branch-items, and group-specific questions, and the average respondent completed over 500 items. The median completion time was 104 minutes. Respondents were able to pause the survey at any point and return for completion. Respondents who paused the survey were sent reminders to complete the survey.

Data for registered voters came from two primary sources: first, a national database of registered voters matched to email addresses, and second from online panel vendors whose panelists are pre-matched to a national voter registration file. Both sets of sample sources were deduped against each other to ensure any invited participant was only included once.

For the first source, registered voters with email addresses were randomly selected to participate in the study, receiving up to five email requests to participate. Invitations were bilingual for selected racial group samples. For the panel vendor sources, respondents were randomly selected and invited to participate through the panel platform they were enrolled in. For all participants, voter registration status was asked and confirmed in the screening process prior to their participation.

For the non-registered portion, sample came from two primary sources: first, email address lists were obtained and flagged against the voter registration database. Participants who were not located on the current voter file were flagged as potential non-registered. Second, online panel vendors with full adult populations were included, in particular those with expertise or specialization in racial minorities and hard-to-reach populations. For both sources, respondents were randomly selected and invited to participate in the study, receiving up to five requests to participate. Invitations were bilingual for selected racial group samples.

As of September 8, in total, 298,159 email addresses were selected and sent invitations to participate in the survey and 116,834 clicked the initial invitation (39.2%). Among this group, 17,126 clicked the link, but then did not answer any survey questions at all (14.7%) and 19,051 people started the survey but were terminated as not meeting eligibility criteria to participate (16.3%) and 31,386 people were terminated due to fulfilled quotas (26.8%). This left 49,270 people who were eligible and started the main survey (42.2%). Among this group, 30,061 answered some questions, but did not complete the full survey. Among the partial completes, a majority answered less than 10% of the questions and then dropped out, and only 9% made it half-way through the survey before dropping out. Overall, 19,209 participated and completed the full survey, among which 14,988 are included in the Primary Sample (PS -Black, White, Latino, and Asian American) and 4,221 are part of the hard-to-reach Oversample groups (as of September 8).

Respondents were offered a gift card of up to \$40 as compensation for their participation. Quality control efforts included the elimination of respondents who demonstrated clear non-serious responses (i.e. speeding or straight-line responses). Preliminary analysis related to inconsistent/illogical responses (i.e. marriage year less recent than birth year) suggests that the vast majority of the sample is of high quality. Of course, no research study can force a participant to pay close attention and be consistent. Indeed, the real world of opinion formation is messy, busy, noisy and filled with distractions. However, our participants were given multiple reminders to proceed slowly, read questions carefully, and give their best honest answers. Still, each researcher needs to make their own judgement and decisions about whether or not they want to flag or isolate particular

respondents from the sample due to inconsistent response patterns, and the threshold will vary by researchers. This is the case in all surveys.

Programming and data collection for the full project were overseen by Pacific Market Research (PMR) in Renton, WA.

In keeping with best practices and data transparency ethics in the social sciences, the original survey data shall be posted to the Inter-University Consortium for Political and Social Research (ICPSR) within up to 2-years of data release (“ICPSR Posting”).

For publication, please cite:

Frasure, Lorrie, Janelle Wong, Matt A. Barreto, and Edward D. Vargas. 2021. The 2020 Collaborative Multiracial Post-election Survey (CMPS). Los Angeles, CA.

September 9, 2021

```
###
```

APPENDIX: SAMPLE R CODE FOR WEIGHTS

```
library(haven) # likely requires the newest haven install
```

```
# Read in Data #
```

```
cmps <- read_dta(file = "~/filepath/cmps_primary_sample.dta") %>%  
  as.data.frame()
```

```
# DESCR #
```

```
library(descr)
```

```
# Column CrossTab #
```

```
with(data = cmps,  
      descr::crosstab(dep = Q14,  
                      indep = race,  
                      weight = weight,  
                      prop.c=T,  
                      plot=F)  
)
```

```
# DPLYR #
```

```
library(dplyr)
```

```
# Single Variable Frequency #
```

```
cmps %>%  
  count(Q14 , wt = weight) %>%  
  mutate(per = n/sum(n))
```

```
# Two-way CrossTab
```

```
cmps %>%  
  group_by(race) %>%  
  count(Q14 , wt = weight) %>%  
  mutate(per = n/sum(n))
```

```
# Raw Numbers with survey package #
```

```
# this package has a lot of functionality but requires more learning  
library(survey)
```

```
des <- svydesign(id=~1, weights=~weight, data=cmps)
svytable(~Q14+race, des)
```

```
# Single Variable Frequency, weights package #
library(weights)
```

```
wpct(cmps$Q14,
      weight = cmps$weight)
```