

Evaluating the generalizability of the COVID States survey — a large-scale, non-probability survey

Jason Radford, Jon Green, Alexi Quintana, Alauna Safarpour, Matthew Simonson, Matthew Baum, David Lazer, Katya Ognyanova, James N. Druckman, Roy Perlis, Mauricio Santillana, John Della Volpe¹

Abstract

COVID-19 fundamentally changed the world in a matter of months. To understand how it was impacting life in the United States, we fielded a non-probability survey in all 50 states concerning people's attitudes, beliefs, and behaviors, designed to be representative at the state level. Here, we evaluate the generalizability of this study by assessing the representativeness and convergent validity of our estimates. First, we evaluate the representativeness of the sample by comparing it to baseline estimates and auditing the size of the weights we use to reduce bias. We find our sample is diverse and most weights are below levels of concern with the exception of Hispanic respondents. Second, we assess the convergent validity of our survey by evaluating how our estimates of attitudes, behaviors, and opinions compare to estimates from other surveys and administrative data. Third, we perform a direct comparison of our results to the Kaiser Family Foundation's probability-based COVID-19 Vaccine Monitor. Overall, our estimates deviate from others by 1%-7% with the larger differences stemming from states with small populations and few other data sources and estimates from items with differing question wording or response choices. Here, we put forward a standard for evaluating the representativeness of surveys, non-probability or otherwise.

¹ This study was led by Jason Radford. Jon Green, Alexi Quintana, Alauna Safarpour, and Matthew Simonson contributed substantive sections of research and writing of the paper. Katya Ognyanova, David Lazer, and Matt Baum are the Principal Investigators for the COVID States Project and set the direction for the study. James Druckman, Roy Perlis, Mauricio Santillana, and John Della Volpe provided senior leadership in determining the methods, questionnaire design, and analytic directions. This research has been supported by the National Science Foundation (SES-2029292, SES-2029792), Knight Foundation, Russell Sage Foundation, Peter G. Peterson Foundation, and Amazon

Introduction

For all the focus on big data and new computational methods of measurement, the ubiquity of the internet and computing have also led to a revolution in survey research. Anyone with an internet connection and survey software can field a large survey for less than a thousand dollars using crowd work platforms like Mechanical Turk and Prolific, online advertising with Facebook or Twitter, or sample aggregation companies like PureSpectrum or Lucid (Berinsky, Huber, and Lenz 2012; Coppock and McClellan 2019; Peer et al. 2017). As a result, non-probability online surveys have become a widely used tool for public opinion research.

A central question concerns whether these data sources can provide generalizable inferences about a population. Probability surveys are considered the gold standard because they allow one to clearly estimate sampling error based on statistical theory. This is not possible with non-probability surveys because they violate key assumptions of this theory. There is a burgeoning literature that assesses the extent to which non-probability surveys arrive at estimates close to those of probability surveys (Ansolabehere and Schaffner 2014; Baker et al. 2010, 2013; Bradley et al. 2021; Kennedy et al. 2016; MacInnis et al. 2018).² While these are useful, generalizing from their results is difficult due to substantial variation among different probability samples (e.g. varying levels of non-response) and non-probability samples (e.g., various filters, weighting) (Kennedy et al. 2016). Our goal is not to offer a definitive statement of whether non-probability surveys are “generally” suitable substitutes. Instead, we seek to assess the generalizability of our particular large-scale, non-probability survey designed in a circumstance where probability approaches were logically and financially infeasible.

A first step in data assessment requires asking why the researchers collected the data. For example, with survey data, researchers recognize that probability methods often prove infeasible with hard-to-reach-populations (i.e., groups that have low prevalence and/or high rates of non-response). For that reason, researchers employ different methods for such populations (Tourangeau 2014). Alternatively, with experimental data, the goal is to generalize a causal inference, meaning that the demographic representativeness of the sample only matters if the inference is moderated by misrepresented groups and the misrepresentation goes unmodeled (Druckman 2022). Here, we focus on a distinct circumstance: data collected to capture dynamics over-time and across space. Such data introduces the challenge that some of the spatial units are hard-to-read. Moreover, frequent over-time measures across space quickly become extraordinarily expensive. Yet, given that many social, political, and economic phenomena occur over-time and across-space (e.g., political campaigns, recessions, social movements), such data are often vital.

Perhaps no event made this clearer than the emergence of Sars-Cov-2 in late 2019 (COVID-19). COVID-19’s rapid spread across the world in early 2020 required immediate scientific action to study its spread, model human behavior, and understand its impact on public policy. In the United States, a comprehensive approach required accounting for differential state action (given the federalized nature of the response) over time (given the non-linear evolution of the pandemic). This meant gathering data from some hard-to-reach groups (e.g., a sufficient sample from small states, given varying non-response rates) and many large data collections

² Non-probability sampling is a sampling method in which not all members of the population have an equal chance of participating in the study

over-time. While monthly probability samples from each state may be an ideal, logistical and resource realities made that simply infeasible. While we do not minimize the value of other data collection approaches with regard to COVID-19, data across space and over time have inherent value as well.

We collected such data over-time (21 waves thus far) and across space (state-level) using non-probability samples. We did so given the pressing need and the resource realities. Our efforts -- the COVID States Project -- began in April 2020³ using the vendor PureSpectrum. We sought to understand the national and state-level changes in American society during the pandemic, over time. Based on these data, we wrote reports and policy briefs evaluating the impact of policy actions, changes in scientific knowledge, and world events to inform the public and guide policy makers (available at covidstates.org). We have also begun publishing our work in peer reviewed outlets (Druckman et al. 2021; Perlis, Ognyanova, et al. 2021; Perlis, Santillana, et al. 2021). An essential task then is to address whether our approach provides sound estimates given their purpose (e.g., point estimates in states, over time trends). As intimated, some question whether non-probability based inferences are reasonable (Baker et al. 2013; Bradley et al. 2021) while others offer evidence to the contrary (Enns and Rothschild 2021). The reality, as mentioned, is that reasonableness varies based on goals, specific studies, and the evaluative techniques employed (Kennedy et al. 2016).

In what follows, we seek to evaluate the generalizability of our survey results. We believe our approach is essential since there continue to be insufficient standards and reporting of auditable and validated results. Our bottom line is: 1) evaluations of non-probability surveys requires assessment of the sample representativeness and external validity relative to administrative and probability survey baselines, and 2) our own effort fared quite well on these criteria, suggesting that when needed, non-probability approaches to collect data over-time and across-states are feasible. Before turning to a more detailed background of our specific survey effort and our validation efforts, we offer a brief preview.

Validating non-probability surveys

For the COVID States Project, we sought to create estimates of opinions, beliefs, and behaviors representative for each state over time. This fine-grained polling is meant to allow policy makers, public health officials, and the public to make the highest-quality decisions based on the most relevant data for them. To that end, we recruited tens of thousands of participants from an online subject pool from around the United States to participate in surveys. These surveys typically occurred every 4-6 weeks. We maintained a core set of questions while changing some content as conditions changed. Our question is how generalizable are estimates made from the data we have collected?

The first way we assess generalizability is by examining the representativeness of our sample. For example, does the sample look like the population on key measures of interest like race and education? The degree of unrepresentativeness can provide us with insight into the overall generalizability of the survey while individual dimensions of unrepresentativeness, say under-sampling conservatives, can provide us with theoretical explanations for why estimates may be more or less generalizable and in what directions our estimates may be biased.

The fundamental limit to representativeness is that we cannot compare a sample to the population along all possible dimensions. We have to pick a few points of comparison. In this case, there are generally established norms for which demographics to use in the United States for surveys, especially surveys about political issues: race, education, gender, income, age, and political ideology. We present those here. However, there is always the potential for unobserved differences making a sample unrepresentative.

The second way we assess generalizability is through convergent validity: comparing our estimates to those from other sources. For example, we can compare how many people report getting a vaccine for COVID-19 in our survey to government-provided counts of vaccination in the population. No source is perfect. In the validation, we focus on administrative data where it exists and use the Societal Experts Action Network (SEAN) COVID-19 Survey Archive to identify surveys with comparable estimates. SEAN only includes probability-based surveys; hence our survey-to-survey comparisons are always against probability-based surveys.

The central challenge for using convergent validity to assess survey quality is that the most of the estimates we care about are those without comparable baselines. (Otherwise why do a survey in the first place?). Thus, our goal is not just to validate whatever questions we can, but to provide indirect evidence for the validity of the items that have no comparisons. In the Covid States survey, the primary topics are the pandemic and the medical and political response to it. Thus, for this study, we focus on validating two sets of comparison questions - political beliefs and behaviors and health-related beliefs and behaviors.

Validation Roadmap

In what follows, we first describe the process by which we recruited subjects. Our goal was to develop samples of each state within the United States that were representative by age, gender, and race. (While we use state-level comparisons where possible in our tests of convergent validity, given the lack of state-level surveys, most of the comparisons we present are at the national level.) We report the demographic characteristics of our overall sample, our weighting procedure, and the distribution of weights from each wave. We then report on baseline comparisons for a variety of political and pandemic-related items asked in our survey, on other surveys, and collected by other methods. Lastly, we report on an in-depth comparison of our estimates against those from the Kaiser Family Foundation's probability-based poll. We examine top-line results for behaviors and attitudes as well as differences among subgroups like age, race, and gender.

Sample Methodology:

Participant Recruitment

We recruited respondents through PureSpectrum, a self-described sample automation platform which aggregates and deduplicates respondents from multiple online survey research firms ("suppliers"). Researchers set a price, which can be adjusted at any time, and suppliers determine algorithmically how many respondents to send depending on survey length and

completion rate, expected profit margin, and capacity to meet any demographic quotas the researcher specifies. The cost and rate of respondents fluctuates in response to supply (e.g., more respondents appear to log in on weekends) and demand (e.g., number of surveys to choose from). Depending on the supplier, panel members either receive an email notification or are presented with a queue of potential surveys to take when they log on a website or mobile app. Some suppliers offer respondents cash or points redeemable for cash or coupons, while others rely on volunteers or online gamers seeking in-game rewards. In order to minimize self-selection bias, we chose not to reveal the topic of the survey until respondents had passed the screening questions and arrived at our questionnaire (hosted externally in Qualtrics).

To facilitate state-by-state comparisons, we run 51 parallel surveys each wave in each state and the District of Columbia. This allowed us to set demographic quotas in each state based on census data for race, gender, and age. While in large states we found it easy to attain our target sample size of 500 respondents for as little as \$0.60 per complete, we were forced to pay more in smaller states where the suppliers' pools of panels were smaller to begin with.⁴ Because supplier algorithms tended to be averse to difficult-to-fill quotas, we left our demographic quotas in the smallest states unnested, while double or triple nesting gender, age, and race in larger states. For the difficult states, PureSpectrum engaged two additional suppliers which had greater coverage in those states but charged a much higher rate. We also created a shortened version of the survey, excluding any questions which we did not plan to analyze at the state level, and directed the majority of our small state respondents to that version, since respondents are more likely to begin a survey if the supplier reports it has a shorter average completion time. Finally, we ran a nationwide minority augment survey to boost the number of Black, Hispanic, and Asian American respondents for subgroup analysis, as well as surveys targeting return respondents who had participated in prior waves.

Table 1 lists the dates, sample size, median survey duration, and sample type of each wave. We use median rather than mean duration because respondents will occasionally wander away from a survey and return hours later, skewing the distribution. Note that three of our surveys were small nationally representative surveys with stricter quotas, and wave 15 was a special survey following the January 6, 2021 U.S. Capitol insurrection which we limited to respondents whom we had previously surveyed in November. All other state surveys have between 18,000 and 26,000 respondents, with 400-500 respondents in the vast majority of states (N=400 results in a 95% confidence interval of ± 4.9 which is often sufficient for our purposes). A handful of low-population states in most waves had sample sizes falling below 350 and were footnoted in our reports accordingly. With a median survey duration of 17-23 minutes across waves, the average cost per completed interview was between \$1-\$2 in all but the most difficult states.

Table 1. Waves of the COVID States Survey
--

⁴ In fact, the closer we got to the 2020 election, the more we needed to pay, likely due in part to suppliers cutting back on recruitment during the pandemic and increased competition with other researchers launching election season surveys.

Wave	Start Date	End Date	N	Median Duration (minutes)	Sample Type
1	4/16/20	4/30/20	19,489	18	State
2	5/2/20	5/15/20	20,305	19	State
3	5/16/20	6/1/20	18,103	21	State
4	6/5/20	6/12/20	2,090	21	National
5	6/12/20	6/28/20	22,470	17	State
6	7/3/20	7/10/20	1,561	20	National
7	7/10/20	7/27/20	19,058	19	State
8	7/31/20	8/7/20	2,029	17	National
9	8/7/20	8/26/20	21,196	20	State
10	9/4/20	9/28/20	20,315	20	State
11	10/2/20	10/23/20	18,002	20	State
12	10/23/20	11/4/20	12,540	19	State
13	11/3/20	12/1/20	24,017	22	State
14	12/16/20	1/11/21	25,640	20	State
15	1/13/21	1/20/21	2,044	19	National ⁵
16	2/5/21	3/1/21	21,500	23	State
17	4/1/21	5/3/21	21,733	21	State

Survey Content

Our research aim was to track public opinion and behavior in the United States during the pandemic. After initial screening questions and household demographics, we asked whether the respondent had been diagnosed with COVID-19. Those who responded “yes” were asked a series of questions about their symptoms and the duration of the illness. We also examined the spread of the virus within households by asking who in the household has been diagnosed with COVID-19 and the relationships between household members including who shares a bedroom. Thus, by fielding a large sample survey at low cost, we are able not only to gather detailed information on a diverse samples of the population,⁶ but to also provide epidemiological tracking

⁵ Sample limited to respondents who participated in Wave 13.

⁶ According to the Centers for Disease Control, only 10.1% of Americans had been diagnosed with COVID by April 15, 2021 which coincided with the midpoint of Wave 17 where we found an 11.1% positivity rate.

of households over time. All respondents were asked about access to testing, interactions with contact tracers, pandemic-induced hardships and concerns, and evaluation of state and federal pandemic responses. We also measured knowledge and exposure to misinformation about the virus, news sources, adherence to social distancing and self-isolation guidelines, vaccination status and attitudes. Finally, to understand the impact of the disease itself and the resulting lockdowns on public health, we asked about mood, loneliness, and social support, including the PHQ9, a 9-item diagnostic questionnaire used in primary care settings to screen for depression (Kroenke and Spitzer 2002).

Several sections of the survey focused not on COVID-19 but on the political events that unfolded during 2020. With the outbreak of widespread protests in the summer of 2020, we added a special section aimed at protesters and collected unprecedented data on protestors, a small subset of the population that would be hard-to-reach using alternative sampling strategies. We tracked support of political candidates throughout the fall, and following the November 2020 general election, we added a Spanish-language version of the survey to better understand why large numbers of Hispanics voted for Donald Trump. In response to record-setting firearms purchases during the pandemic, we added a supplementary section for gun-buyers to understand their motivations. We also launched two special surveys for return respondents following unexpected news events to examine changes in attitudes. Following the Capitol Insurrection of January 6, 2021, we recruited 2,044 respondents who had participated in the November post-election wave to examine how well their prior attitudes toward the legitimacy of the election and acceptability of political violence predicted their reaction to the insurrection. During the FDA's two-week pause of the Johnson & Johnson vaccine in April 2021, we recruited respondents who had taken the survey just prior to the pause to take the survey again in order to measure within-respondent changes in willingness to get vaccinated.

At the conclusion of the survey, we asked respondents to supply their Twitter handle if they are willing, following common approaches (Hughes et al. 2021). Respondent-supplied handles were filtered for quality such as removing email addresses, entries longer than the maximum handle length, or entries that use characters that are not allowed in Twitter handles; before being stored separately for timeline collection. We also offered an additional monetary incentive for respondents who were willing to install a browser extension we designed for both conducting targeted audits, passive monitoring, and collecting data on what websites know about people, such as Google account activity or ad preferences. The results of research related to the browser extension will be detailed elsewhere.

Data Analysis

Data Weighting Procedure

We employ two types of weighting schemes in our analyses. In both, we use raking with the “survey” package in R (Lumley 2020) which is a form of calibration adjustment across multiple weights (Baker et al. 2013). For state-by-state comparisons, we reweight respondents according to 2019 American Community Survey's estimates for gender, age, race, education, and urbanicity within each state. Weights are truncated so that, in effect, no respondent can be

counted more than five times in the reweighted data.⁷ For analyses at the national level, we re-weight respondents according to interlocked gender-age-race subgroups, as well as education, urbanicity, and region. As detailed below, Hispanics, men, less-educated and older respondents were more likely to be underrepresented in the sample and thus receive high weights. Region and urbanicity required little adjustment.

Our weighting approach is relatively straightforward, especially when compared to a variety of model-based alternatives (Valliant 2020; Yang, Kim, and Song 2020). This is by design, as our goal here is to examine how our results compare to external estimates using conventional methods. Moreover, as Valliant (2020) discusses, all weighting procedures make similar (strong) assumptions regarding the relationship between the sample and target populations. If such assumptions are violated, more elaborate weighting schemes will not by themselves represent a solution. If such assumptions are met, as the sample size increases, different weighting schemes will likely converge on similar answers. The intuition behind our approach is that when the extent to which such assumptions may be violated is unclear (as is the case here, and in most current survey contexts) the optimal strategy is to triangulate across multiple surveys and ground-truth benchmarks (Bradley et al. 2021).

Data Filtering

One of the biggest challenges of online panels is preventing inattentive respondents from contaminating the sample. Since respondents are paid by the survey and not by the question or by the minute, there are adverse incentives for respondents to speed through without reading—skipping questions, clicking random responses, or employing assistive technology (bots, autocomplete, etc.). This problem becomes particularly acute when studying rare phenomena. For instance, if a question asks “Have you attended a protest in the past month?” half of all respondents who click randomly will say “Yes,” thus inflating estimates of protest attendance, perhaps substantially. Furthermore, since researchers often pay more for hard-to-reach subpopulations, there is a risk that savvy, experienced respondents will lie about their demographics and location in order to qualify for a high-paying survey.⁸ PureSpectrum takes steps to mitigate some of this risk; for instance, when respondents report their zip code, if the location of their IP address is not in the state containing that zip code, they are flagged for follow-up. Respondents cannot re-enter the same survey twice within a given wave, regardless of whether they initially qualify. Finally, PureSpectrum uses respondents' self-reported demographic traits to deduplicate respondents who are signed up for multiple suppliers to prevent them from taking the survey twice and also blacklists respondents who change their answers to certain demographic questions from one survey to the next.

In addition to PureSpectrum's built-in quality filters, we implemented several additional steps to improve data quality. Our survey contained two attention check questions (e.g., “Just so we know you are still with us and paying attention, please select “never” below”), a pledge to take the survey seriously to encourage more thoughtful responses, a CAPTCHA test to filter out

⁷ To be precise, weights are trimmed above at either 5 or the 98th percentile, whichever is lower, and below at either 0.2 or the 2nd percentile, whichever is higher. For national weights, we trim at the 1st and 99th percentiles, resulting in a maximum weight between 5 and 6 for most waves.

⁸ There is also a risk of respondents for low-income countries using a VPN to pretend to be in the United States or farming out the survey via Mturk to other workers (Kennedy et al. 2020).

bots, and a speeder check to disqualify anyone finishing the survey implausibly fast. After disqualifying respondents who failed any of those filters, we then set up a series of “flags,” tripping any two of which would result in disqualification: reporting more than 10 household members, claiming to have engaged in more than five activities outside the home in the past 24-hours (from a list of 13), professing to be a follower of more than two religions, being employed in multiple high-COVID-risk professions, straightlining (choosing the nth option each time) on more than 14 consecutive questions,⁹ and reporting opposite ends of the political spectrum for party and ideology. The intuition behind the flags is that while the sample may include respondents who are half-Jewish, half-Muslim or who work part-time in hospital and part-time in a grocery store, the chances that respondent meets two or more of these rare conditions is exceedingly slim compared to the chances that they are clicking randomly. The numeric cutoffs for the flags were determined in the initial waves by comparing the rate of flag-tripping for individuals who failed an attention or speed check to the rest of the pool, in order to filter out respondents whose answers most resembled the attention check violators.

Finally, beginning in the 10th wave, we added the following text entry question: “Other than the state you live in, in what other U.S. state do you have the most friends and family members?” Respondents who gave an acceptable response passed the filter, those giving an irrelevant response were disqualified, and those who skipped the question or gave a response we did not anticipate were added to the aforementioned list of flagged respondents, such that they needed to fail a second flag filter to be disqualified. Lists of acceptable and irrelevant responses were developed by hand-coding the initial waves of respondents (any term relating to geography or indicating “N/A”, “don’t know,” “no friends,” etc. were deemed acceptable), as well as using fuzzy matching to state names to allow for misspellings. This filter was implemented in response to concerns over answers to other free response questions, but since it was not included in the first 10 waves, we do not use it in longitudinal analyses that include them. We highly recommend that researchers implement a filter like this from the get-go when using online samples, and that they choose a question with a limited number of plausible responses (e.g., state names).

Results

Demographic Representativeness

A survey is representative when it perfectly resembles the population on all measured and unmeasured characteristics. In practice, no sample (non-probability or otherwise) perfectly resembles the population and we use weighting to reduce the biases in our estimates stemming from this imperfection. In this section, we examine the underlying, unweighted sample to identify potential areas of under-coverage - where the sample may be too small for drawing inferences. We also investigate potential limits of our weighting procedure by looking at the demographic groups where we had to trim our weights. The results suggest that our sample hews closely to the American population but points to potential under-representation of Hispanics and those with less than a high school degree.

⁹ Also, having a straightline average greater than 1.3 consecutive questions.

A word of caution: this assessment is based on the entire sample across all waves and provides a general overview of the dimensions of sample bias in the survey. However, these results should not be understood as validating any particular subsample including for any particular state for any particular wave. Subsequent studies on subsamples will involve their own representativeness checks.

Sample Diversity

Although we sampled by state-level race, gender, and age; the COVID States survey largely mirrored the population of the United States on these and other dimensions. In Table 2, we show what proportion of the sample matches each key characteristic we measured and could compare to the population at large as measured by either the census or probability-based samples. Given that we conducted multiple waves with some participants providing data multiple times, Table 2 is broken down for all data, all unique participants, and those who provided data more than once. The comparisons allow us to document any potential panel bias in our subsequent results. These are aggregate statistics across all waves and states, providing a general sense of any sampling bias. There is likely to be some degree of heterogeneity in these sample biases at the state level.

	All Subjects, All Waves (unweighted)	Unique Participants (unweighted)	Repeat Participants (unweighted)	U.S. Population
Age				
18-24	12.5	16.0	6.2	12 ⁺
25-34	19.9	23.3	14.9	18 ⁺
35-44	20.5	22.4	18.8	16 ⁺
45-55	15.1	14.5	16.9	16 ⁺
55-65	15.1	12.2	19.6	17 ⁺
65+	17.0	11.6	23.7	21 ⁺
Income				
< \$25,000	23.7	25.4	18.5	17 ⁺
\$25 - \$49,999	24.9	23.4	26.1	20 ⁺
\$50 - \$74,999	16.6	16.1	19.2	16 ⁺
\$75 - \$99,999	13.7	13.1	14.7	12 ⁺
> \$100k	21.9	22.2	21.5	34 ⁺
Education				
Less than HS	3.0	3.5	2.0	10 ⁺
HS Grad	21.7	22.1	20.6	28 ⁺
Some College	33.7	33.8	33.8	26 ⁺
Bachelor Degree	25.9	24.8	27.7	22 ⁺
Graduate Degree	15.7	15.9	15.7	3.5 ⁺
Gender				
Male	32.9	32.7	31.8	49 ⁺
Female	67.1	67.3	68.2	51 ⁺

Race/Ethnicity				
White	73.9	72.5	76.6	75 ⁺
Black	10.3	10.5	9.9	14 ⁺
Hispanic	6.7	7.5	5.1	19 ⁺
Asian	5.5	5.5	5.6	7 ⁺
Other	3.6	4.0	2.7	4.3 ⁺
Religion				
Christian	72.9	69.0	94.1	71 [*]
Nonreligious	21.0	24.2	4.6	22 [*]
Jewish	1.5	1.6	.4	2 [*]
Muslim	1.0	1.2	.2	.9 [*]
Buddhist	1.2	1.4	.3	.7 [*]
Hindu	.5	.5	.09	.7 [*]
Party				
Democratic	35.8	35.6	37.4	31 ^{G1}
Republican	29.2	27.9	31.3	31 ^{G1}
Independent	29.1	29.6	27.8	36 ^{G1}
Other	5.5	6.4	3.3	NA
Ideology				
Liberal	31.8	33.1	30.5	24 ^{G2}
Moderate	36.7	37.1	36.1	35 ^{G2}
Conservative	30.4	28.6	32.9	37 ^{G2}
N	272,092	190,508	39,997	
Standard errors were 2% or less. The Race/Ethnicity item is unconventional as it was taken from PureSpectrum who provided it exclusively, meaning participants could be Hispanic, Black, White, Asian, or Other but could not be multiple.				
⁺ Based on 2019 American Community Survey estimates				
[*] Based on the Pew Religious Landscape study (Pew Research Center 2014)				
^{G1} Based on Gallup Party Affiliation (Gallup 2021)				
^{G2} Based on Gallup's 2019 Report (Saad 2020)				

Overall, we substantially undersampled several demographic groups - those with less than a high school degree (3% compared to 10%), those with only a high school degree (22% to 28%), those with incomes over \$100,000 per year (22% to 34%), those who identify as conservative (30 to 37%) and men (33% to 49%). Our measure of race and ethnicity is unconventional. This item was collected by PureSpectrum who asked participants to choose one and included hispanic as one of the choices. Thus, our statistics are not directly comparable to the Census which allows participants to select multiple races and asks about hispanic ethnicity separately (hence the Census statistics add up to more than 100%).

Next we address panel bias. While Covid States is not a panel survey, some participants did participate in multiple waves providing more data than others. Overall, there was relatively little repeat participation: 91% of all participants participated once (79.0%) or twice (12%). Only 5,605 people (2.9%) participated in five or more waves. If we compare every individual who participated to those who participated more than once, we can identify the panel bias - the types of people who are more likely to contribute their data again and again. We find older participants

and those who identify as Christian were substantially more likely to participate multiple times. Overall, panel effects should be limited given 1) the relatively small number of participants who participated multiple times and 2) the small demographic differences in who participated multiple times.

Table 3: Percent of Participants with Trimmed Weights by Demographic Group		
	Pct Trimmed (National)	Pct Trimmed (State)
Age		
18-24	0.3	<0.1
25-34	0.9	<0.1
35-44	0.5	<0.1
45-55	1.1	0.1
55-65	0.9	<0.1
65+	1.4	0.1
Income		
< \$25,000	1.5	.01
\$25 - \$49,999	1.0	0.1
\$50 - \$74,999	0.7	0.1
\$75 - \$99,999	0.5	<0.1
> \$100k	0.3	<0.1
Education		
Less than HS	18.4	0.1
HS Grad	0.8	0.1
Some College	0.3	<0.1
Bachelor Degree	0.2	<0.1
Graduate Degree	0.1	<0.1
Gender		
Male	1.9	0.1
Female	0.4	<0.1
Race/Ethnicity		
White	0.4	<0.1
Black	1.4	0.1
Hispanic	6.0	0.3
Asian	0.9	<0.1
Other	0.2	0.2
Religion		
Christian	0.9	<0.1
Muslim	0.7	0.1
Buddhist	0.8	0.0
Nonreligious	0.8	<0.1
Hindu	0.9	0.0
Jewish	0.4	0.0
Party		
Democratic	1.0	0.1
Republican	0.6	<0.1
Independent	0.9	0.1
Other	1.1	0.1

Ideology		
Liberal	0.9	<0.1
Moderate	1.0	<0.1
Conservative	0.8	<0.1

Survey Weights

For the national weights, we trimmed the weights for 4,678 surveys (1.8% of all responses) resulting in a trimmed range from .12 to 7.6. For the state weights, we trimmed the weights for 1,910 surveys (.7% of all responses) resulting in a range from .22 to 3.3. To explore the impact of our weighting schema and our trimming rules, we examine the percent of each demographic group who had their weights trimmed for being too large (see Table 3).

First, there were no demographic groups whose weights were trimmed at the state level. In general, the proportion of the sample with trimmed weights was always larger for the national weights than the state weights. This was due to two things. First, our sampling frame was representative at the state-level rather than the national level. Second, there were more weights created for the national data and some were interlocked. For example, white women under 25 had a different weight than black women under 25.

Second, sample size did not necessarily correspond to the propensity for groups to have their weights trimmed. Some groups, like religious minorities, are small in our sample and in the population. Moreover, even though we examine the impact of weighting on all of our demographic groups, we only reweighted on a subset of characteristics. In this case, we see larger trims for the groups we actually weighted based on. We do not see any spillover into demographic categories that were not weighted like political party or ideology.

Overall, demographic groups had few members with trimmed weights with two exceptions. First, a large proportion of people with less than a high school education (18%) had weights that were trimmed. Second, a sizable minority of hispanics (6%) had their weights trimmed. Taken together, 50% of hispanics with less than a high school education (397 out of 802) had weights that were so large they had to be trimmed. Thus, we should be careful in generalizing national-level estimates to this subpopulation.

Public Health Comparisons

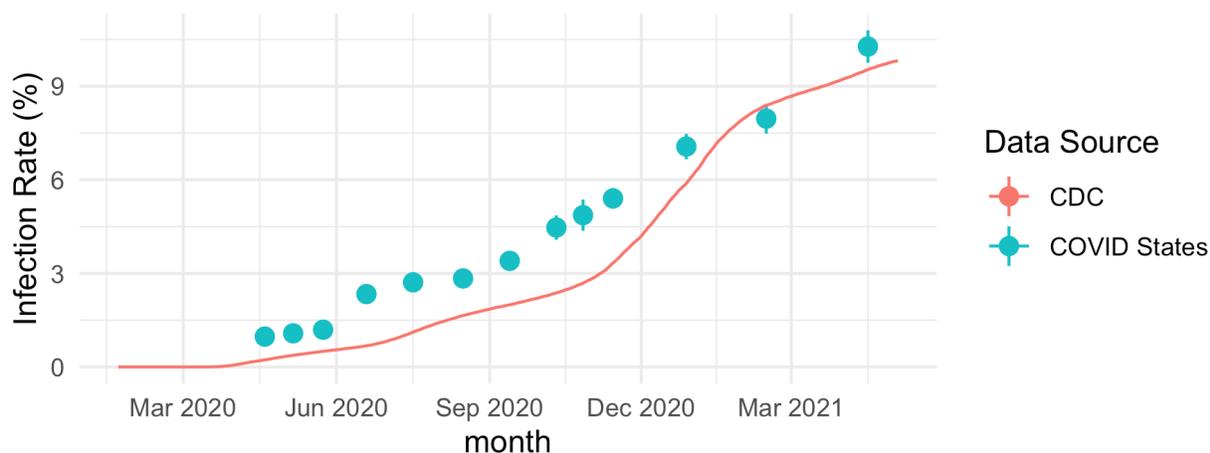
In this section we evaluate how closely the COVID States survey follows other measures of public health phenomena. We focus on measuring the prevalence of COVID-19 and vaccination rates as compared to official state data. Then we compare our estimates of people's attitudes, behaviors, and mental health experience to results from other surveys. In some cases, we approximate or improve on these baselines and, where we diverge, it largely comes from differences in the way we pose our question as compared to how other survey groups pose their questions.

COVID-19 Prevalence

True prevalence of SARS-CoV-2 in the U.S. has been difficult to measure, even for public health agencies. Many epidemiologists believe true prevalence has been widely underreported,

particularly early in the pandemic when tests were in short supply (Irons and Raftery 2021; Kalish et al. 2021; Lu et al. 2021). While both public health agencies are likely to miss asymptomatic or mild infection cases, surveys have the added risk of overestimating rates of symptomatic infection, since, in the absence of access to testing or medical professionals, respondents may misattribute to SARS-CoV-2 the symptoms of other illnesses. We attempt to avoid such overreporting by asking specifically “Have you been diagnosed with coronavirus (COVID-19)?” and coding as positive only those who respond “Yes, I was diagnosed by a medical professional” (other options include “No, I was not diagnosed but I think I may have it now” as well as a definite “no” or unsure). That said, some patients may have been told by medical professionals over the phone or at hospital waiting rooms to presume they were positive if they were showing symptoms rather than coming in for a test and risking exposing others. Thus, we should not be surprised if our estimates are systematically higher than case counts compiled by the CDC from state and local health agencies.

Estimates of COVID-19 Infection Rate in the U.S. Over Time



Bars indicated 95% confidence intervals.
 CDC rates reported daily based on statistics from state and local health agencies.
 COVID States estimates are plotted at weighted mean date of each wave.

Figure 1: CDC and COVID States Estimates of COVID-19 Infections

As shown in Figure 1, our estimates, though systematically higher, consistently track with case rates reported by the CDC (CDC Case Surveillance Task Force 2021). Our reported cumulative infection rate never decreases. With the exception of the penultimate wave, our estimates range from 0.7 to 2.2 percentage points higher than that of the CDC, which is consistent with the theory that symptomatic cases not reported to public health agencies were widespread in the first few months of the pandemic (prior to the start of our survey) but have not accumulated since.

This is further shown when examining responses to our question about COVID testing among respondents who did or did not report having been diagnosed with COVID-19 by a medical professional (Figure 2). During the early stages of the pandemic, we observe high rates of respondents reporting having been diagnosed with COVID-19 by a medical professional but

not having received a positive test. Over time, testing status and diagnosis status increasingly align, consistent with expanded testing capacity.

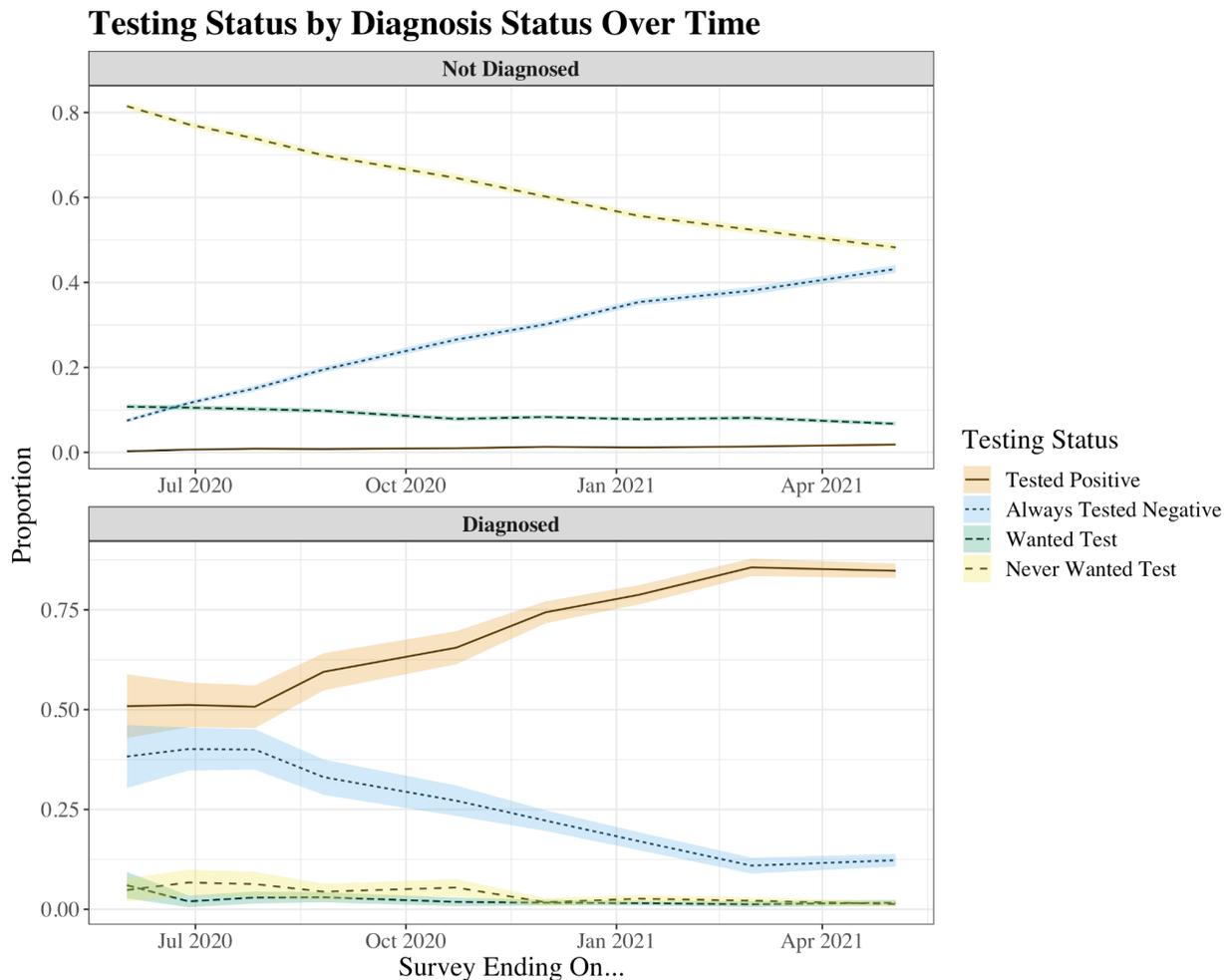
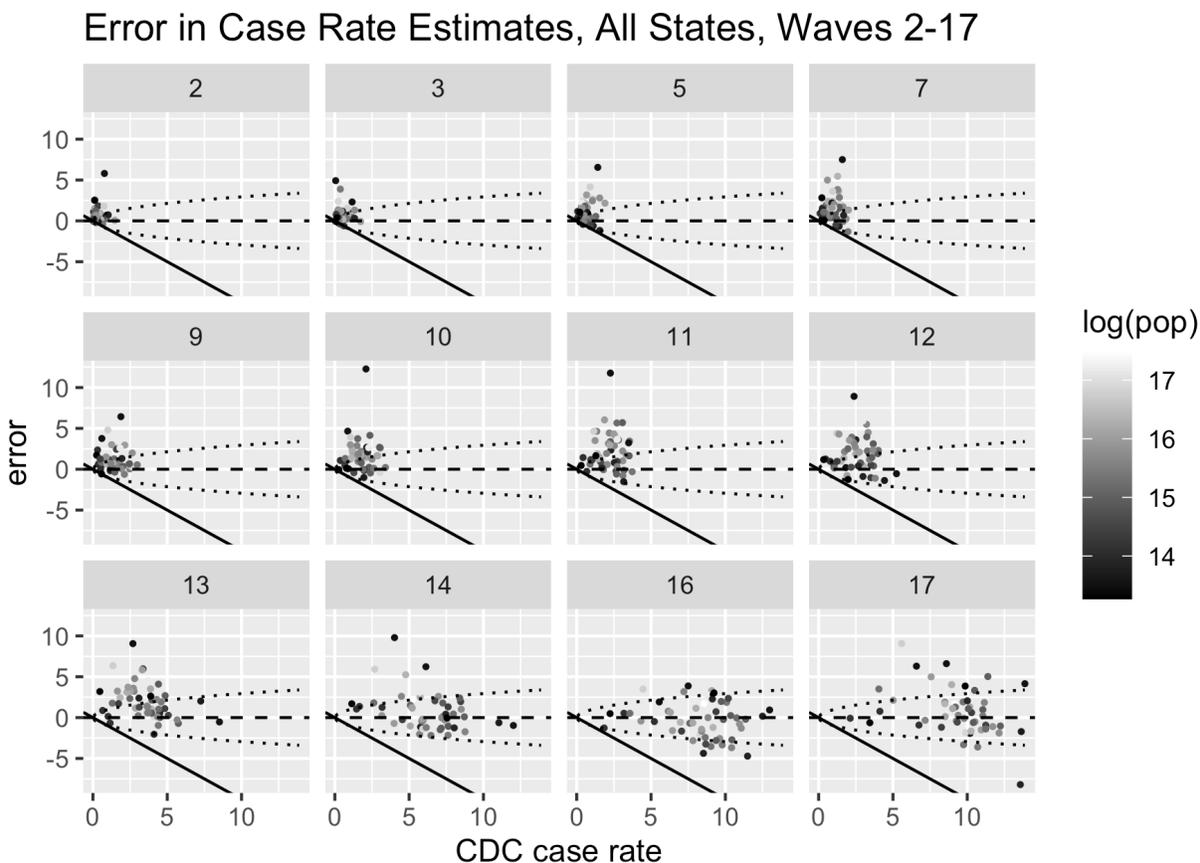


Figure 2: COVID States Testing and Test Positivity Rates



Each panel indicates a wave (wave 1 is looks similar to wave 2 and has been dropped due to space considerations). The solid line indicates the most negative possible error value, while dashed line shows x-axis. The dotted curves indicate the zone of the 95% margin of error, assuming $N=400$ (a typical value for most states in most waves)

Figure 3: Error in state-level comparisons between CDC and Covid States survey

Figure 3 demonstrates the discrepancies between our survey's COVID-19 cumulative incidence rates and those reported by the CDC in each of the 50 states plus the District of Columbia. In general, our survey appears to nearly always match the CDC's case rate in a given state to within 5 percentage points in any given wave. While the typical state's error has not substantially changed over time (rising only marginally in waves 7–13 before falling again), our estimate of cumulative infection rates increasingly converged with the CDC as rates increased over time. For CDC case rates higher than 4%, approximately three-quarters of states-wave observations fall within a 95% margin of error.¹⁰ As indicated by the shading, a state's underlying population is not strongly associated with the accuracy of our predictions. Note that our systematic overestimation may also be due in part to the fact that it is impossible to obtain

¹⁰ This number rises steadily for cutoffs lower than 4.0, then fluctuates between 75% and 80%. Note that the standard error formula yields an ever-widening margin of error as the underlying proportion being estimated approaches 0.5. Thus, our survey is not necessarily more accurate for higher rates or later waves; rather the 95% margin of error grows more accommodating.

an estimate lower than 0% (and hence an error greater than the CDC case rate itself, as indicated by the black diagonal in each panel).

In sum, our survey captures the upward trend in cumulative COVID-19 cases both on the state and national level and gradually converges to the CDC estimate over time. While we generally overestimate the CDC's reported case rate, we know the CDC data are an undercount of diagnoses of "presumed" COVID-19, the information on which patients are told to act regardless of its clinical accuracy. At the state-level, once the CDC's estimated case rate rises above 4% in a given state, we are able to predict it within two standard errors three-quarters of the time, regardless of its underlying population.

Vaccination Rate

Starting in Wave 14, we asked participants "Have you or anyone you know received a COVID-19 vaccine?" We produced national estimates based on those who selected "I was vaccinated for COVID-19." In Figure 4, we plot these estimates against records gathered by the Centers for Disease Control (Centers for Disease Control and Prevention 2021). Our estimates capture the overall S-curve trend in vaccination rates over the first half of 2021 and fall within 3 percentage points of the CDC's rates each time. Since our survey waves each last approximately four weeks, we plot our estimates at the weighted median date—the date when the middle-most respondent took the survey after reweighting on demographics. Furthermore, the CDC's reporting is also uneven and subject to change, since it depends on state and local health departments which sometimes submit backlogged or duplicated data (Associated Press 2021; Wingrove 2021)

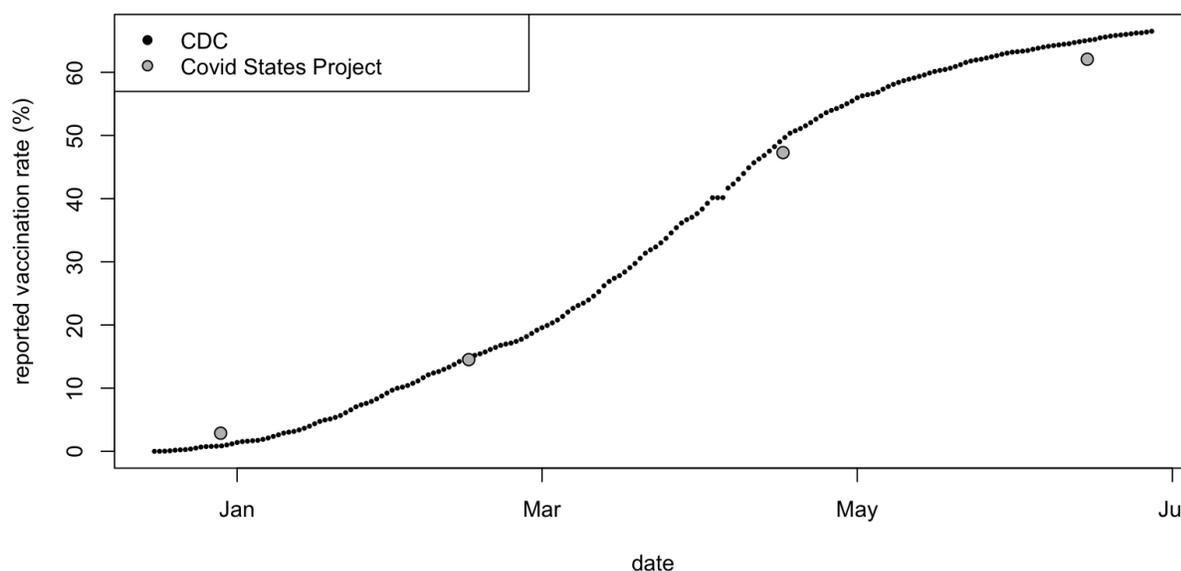


Figure 4: CDC and COVID States Vaccination Rates

Public Health Attitudes

Our survey regularly monitored general attitudes regarding the COVID-19 pandemic, including people's support for specific public policies and attitudes about the vaccine. One public health attitude question we asked was a battery adapted from a question Pew asked their American Trends Panel in mid-March of 2020 regarding how different people and institutions were reacting to the outbreak -- whether they were overreacting, reacting about right, or not taking the outbreak seriously enough (Pew Research Center 2020b).

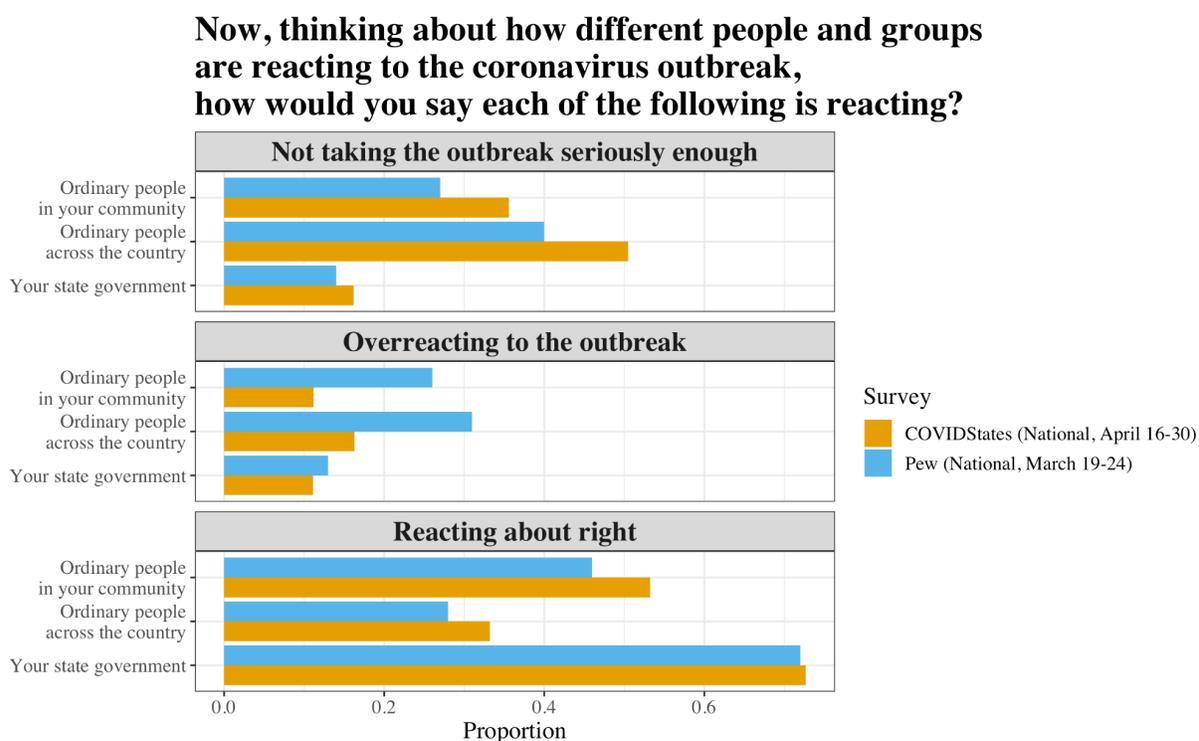


Figure 5: Comparing Covid States and Pew Measures of the Public Reaction to the Pandemic

We see general agreement between our results and theirs particularly on the question of how the respondent's state government was reacting. We see large shifts in perceptions of how ordinary people were responding. In our data, respondents generally felt more like others were not taking the outbreak seriously enough. However, these make sense given the timing of our survey. The Pew survey was fielded just as the first wave of the pandemic began. Ours was fielded three weeks later as the first wave of the pandemic had spread around the country. Unfortunately, this question was not asked again, so we cannot compare our measures over time.

Approval of COVID-19 restrictions

The second set of attitude questions we asked revolved around participants' support of COVID-19 restrictions like mask wearing, avoiding large gatherings, and staying home. In our survey, we ask: "Do you approve or disapprove of the following measures which federal, state, and local

governments could take to prevent the spread of coronavirus (COVID-19) in the next 30 days?”. The possible answers are “Somewhat approve”, “Strongly approve”, “Somewhat disapprove” and “Strongly disapprove.” While three polls asked questions about these attitudes, the question wording, option choices, and timing varied substantially among them (see Table 4).

In general, our participants were more supportive of restrictions. However, there were systematic differences in how we posed the questions. AP-NORC allowed participants to provide a neutral opinion (“neither favor nor oppose”). While 9-14 percent more of our respondents approved of restrictions compared to AP-NORC, only 2-3 percent less of our respondents disapproved of restrictions. This means that the neutral category cannot account for all the difference but that it likely accounts for much of it. In the one case where we see the same question asked in multiple waves by the same survey (the restrictions on U.S. travel), both surveys document the same changes in attitudes as fewer people are strong supporters and more disapprove of restrictions from May 2020 to June 2020.

The largest differences in our measures comes from differences in question and option framing. For example, the NPR/PBS/Marist poll in December 2020 asked if a national stay-at-home-order was a good idea or not while we asked if people approved of governments “asking people to stay home”. While 84.9% of our participants approved, only 51% thought a stay at home order was a good idea. We believe this captures a substantive difference in public opinion rather than any measurement error.

Table 4. Comparing Public Approval of Covid Restrictions across Surveys							
Topic	Covid States Question	Covid states Date	Covid States results	Other survey Question	Source	Other Date	Results
Travel Restrictions	Restricting international travel to the U.S.	4/16 - 4/30/20	Somewhat approve + Strongly approve, 94.6%, Strongly disapprove + Somewhat disapprove 5.4%	Restricting international travel to the U.S.	(Pew Research Center 2020c)	3/19 - 3/24/20	Necessary: 95%, Unnecessary: 5%
	Restricting travel within the U.S.	5/16 - 6/1/20	Strongly approve 43.8%, Somewhat approve 32.8% Somewhat disapprove 15.3% Strongly disapprove 8.1%	Restricting travel within the U.S.	(AP-NORC 2020d)	5/14 - 5/18/20	Strongly favor: 37%, Somewhat favor: 27%, Neither favor nor oppose: 15%, Somewhat/Strongly oppose: 20%
		6/12/ 6/28/20	Strongly approve 40.1%, Somewhat approve 35.1%, Somewhat disapprove 16.2%, Strongly disapprove 8.6%	Restricting travel within the U.S.	(AP-NORC 2020b)	6/11 - 6/15/20	Strongly favor: 20%, Somewhat favor: 31%, Neither favor nor oppose: 19%, Somewhat/Strongly oppose: 29%
Mask Mandate	Requiring people to wear masks when in public	12/16/20 - 1/11/21	Strongly approve 70.8% , Somewhat approve 15.9%, Somewhat disapprove 6.4%, Strongly disapprove 6.9%	Requiring Americans to wear face masks when they're around other people outside of their homes	(AP-NORC 2020a)	12/3 - 12/7/20	Strongly favor: 59 Somewhat favor: 14 Neither favor nor oppose: 12 Somewhat oppose: 8 Strongly oppose: 8

Mask Mandate		2/5/21 - 2/28/21	Strongly approve 70.1% Somewhat approve 16.1% Somewhat disapprove 6.5% Strongly disapprove 7.2%		(AP-NORC 2021)	2/25 to 3/1/21	Strongly favor: 54 Somewhat favor: 15 Neither favor nor oppose: 14 Somewhat oppose: 8 Strongly oppose: 8
Social Distancing	Asking people to stay at home and avoid gathering in groups	12/16/20 - 1/11/21	Strongly approve 59.3% Somewhat approve 25.6% Somewhat disapprove 8.9% Strongly disapprove 6.2%	Due to coronavirus, do you think a national stay-at-home order, if recommended by public health officials, is a good idea or a bad idea?	(NPR/PBS NewsHour/ Marist 2020)	12/1 to 12/6/20	Good Idea: 51%, Bad Idea: 45%, Unsure: 4%
Business closures	Requiring most businesses other than grocery stores and pharmacies to close	4/16/ - 4/30/20	Strongly approve 51.3% Somewhat approve 31.3% Somewhat disapprove 12.1% Strongly disapprove 5.2%	Requiring most businesses other than grocery stores and pharmacies to close	(Pew Research Center 2020c)	3/19 to 3/24/20	Necessary: 71%, Unnecessary: 28%, No Answer: 1%
Canceling Events	Canceling major sports and entertainment events	4/16/ - 4/30/20	Strongly approve 74.7% Somewhat approve 17.3% Somewhat disapprove 5.5% Strongly disapprove 2.5%	Canceling major sports and entertainment events	(Pew Research Center 2020c)	3/19 to 3/24/20	Necessary: 91%, Unnecessary: 9%, No Answer: 0%

Closing Schools	Closing K-12 schools	4/16 - 4/30/20	Strongly approve 71.8% Somewhat approve 19.8% Somewhat disapprove 5.4% Strongly disapprove 2.9%	Closing K-12 schools	(Pew Research Center 2020c)	3/19 to 3/24/20	Necessary: 90%, Unnecessary: 10%, No Answer: 0%
				Closing Schools	(AP-NORC 2020c)	3/26 - 3/29/20	Strongly Favor: 58%, Somewhat Favor: 25%, Neither Favor Nor Oppose: 10%, Somewhat Oppose: 3%, Strongly Oppose: 3%
Note: These survey results were found via the Societal Experts Action Network (SEAN) COVID-19 Survey Archive at https://covid-19.parc.us.com/client/index.html#/							

Compliance with policies

Through our different survey waves, we asked a battery of questions on compliance with health recommendations related to COVID-19 prevention. The battery starts with the question: “In the last week, how closely did you personally follow the health recommendations listed below?” and then lists a series of behaviors. The possible answers are “Very closely”, “Somewhat closely”, “Not very closely” and “Not at all closely”. Among the behaviors we ask for, we found similar questions in other surveys for “Avoiding public or crowded places” and “Avoiding contact with other people”.

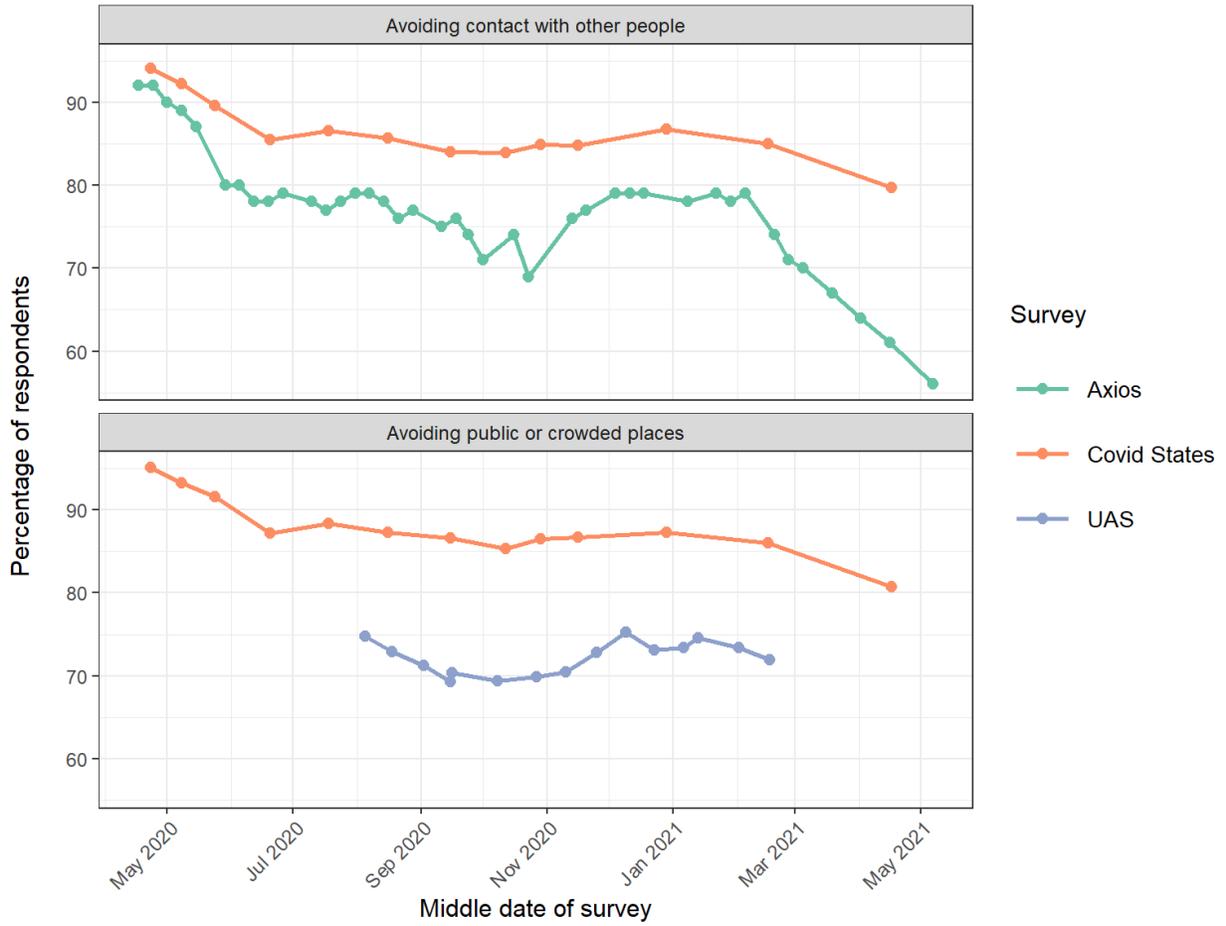
The Understanding America Study (UAS)¹¹, a panel of households at the University of Southern California (USC) representing the United States has been asking the following question since July 2020: “Which of the following have you done in the last seven days to keep yourself safe from coronavirus? Avoided public spaces, gatherings, or crowds”. The possible answers are “Yes”, “No”, and “Unsure”. Similarly, the Axios-Ipsos polls (Ipsos 2020, 2021) on COVID-19 have been asking the following question: “Have you done the following in the last week? Social distanced – that is stayed at home and avoided others as much as possible” using the Ipsos Knowledge panel, since April 2020. The possible answers are “Yes” or “No”.

While these questions match our questions well, the possible answers do not, so we compared the percentage answering “Very closely” or “Somewhat closely” in our surveys to the percentage answering “Yes” in the UAS and Axios-Ipsos surveys. Figure 6 shows the result of this comparison. We consistently overestimate the percentage of positive answers, however, the mismatch between the question answers could partially explain these differences. In addition, the trends for both surveys are similar. In Figure 7, we compare the difference in percentage of positive answers from one wave to the next wave of our survey, with this difference for the waves of the UAS and Axios-Ipsos polls that are closer to each of our waves. We compare relative differences, that is, we divide them by the standard deviation of the differences corresponding to a given question and survey. In general, these relative differences are small and in the same direction. In other words, the periods during which our estimates decline and increase correspond to similar periods for the UAS and Axios-Ipsos polls, and, in relative terms, the differences are of the same order.

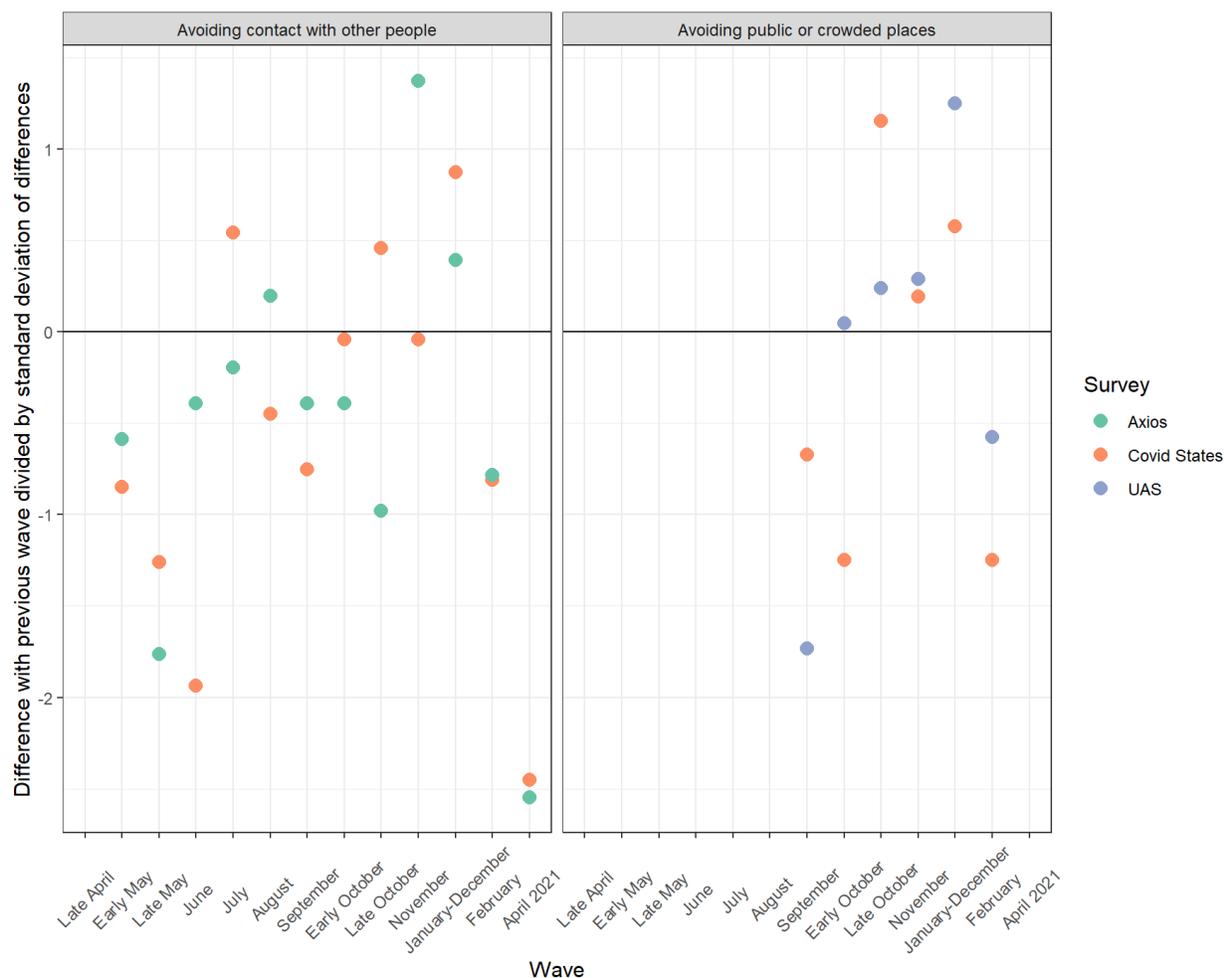
¹¹ Data available at <https://uasdata.usc.edu/index.php>. The Understanding America Study is maintained by the Center for Economic and Social Research (CESR) at the University of Southern California. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of USC or UAS. The collection of the UAS COVID-19 tracking data is supported in part by the Bill & Melinda Gates Foundation and by grant U01AG054580 from the National Institute on Aging, and many others.

Adherence to health recommendations

Percentage answering 'Yes' for the UAS and Axios surveys and percentage answering 'Very Closely' or 'Somewhat Closely' for the Covid States surveys



Comparison of trends in the Covid States and UAS/Axios surveys regarding adherence to health recommendations



Mental Health Indicators

Beyond its direct public health effects, COVID-19 has also generated a considerable degree of stress and social isolation in the U.S. public. This has led to concerns regarding mental as well as respiratory health. Our survey periodically includes the Patient Health Questionnaire-9 (PHQ-9) for either all respondents or a random subset, so as to monitor both the prevalence and correlates of depressive symptoms.

Here, we compare our national estimates of minor depressive symptoms to the Census Household Pulse survey, which is the only analogous nationally representative survey to include questions from this battery. The Pulse survey asks the short-form PHQ-4, which contains two items related to depression and two related to anxiety: “Little interest or pleasure in doing things” and “Feeling down, depressed, or hopeless.” Following convention in the literature (Kroenke et al. 2009), we code responses to these two items on a 0-3 scale and define

respondents as exhibiting minor depressive symptoms if the sum of these two responses is greater than or equal to 3.

Figure 8 plots the means and 95% uncertainty intervals from each wave of each survey that showed these two items to respondents, representing the final date of the relevant survey wave on the x-axis. As the figure shows, the estimated prevalence of minor depression does not perfectly match between surveys, but is typically close and for the most part exhibits similar time trends. The biggest topline deviations tend to be in the summer of 2020, where a smaller number of COVID States respondents were shown the items and there is more uncertainty. However, all estimates in both surveys fall within an eight percentage point window throughout the entire period. In general, the time trends in the COVID States survey indicate more stability as well, though estimates in both surveys tend to increase in October and December of 2020 and decrease in the early months of 2021.

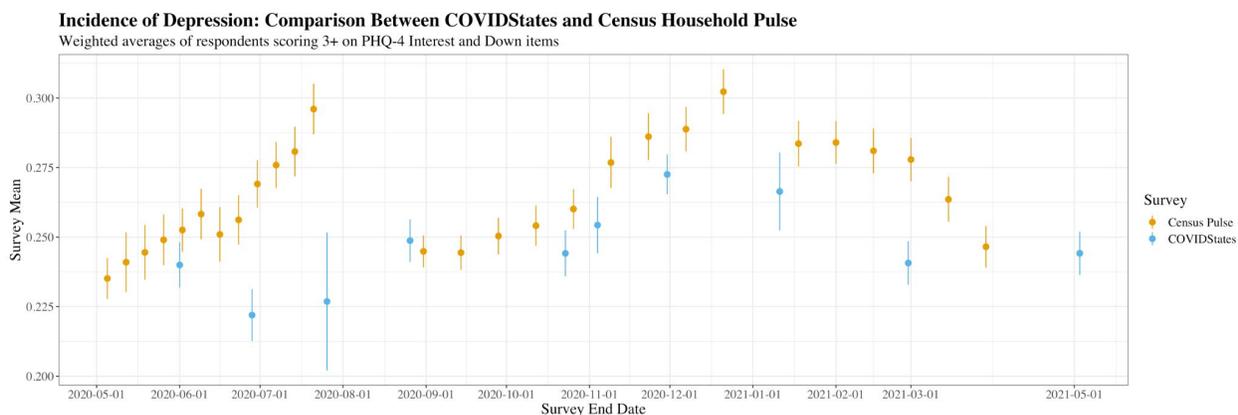


Figure 8: Incidence of Depression between Covid States and Census Household Pulse Survey

Generally speaking, we take this as encouraging evidence that our survey is picking up a similar signal as that picked up in the Census Household Pulse survey. Both surveys' estimates are nearly always within a close range, only once deviating by more than five percentage points (which was itself the most uncertain estimate in the COVID States data) and typically exhibiting similar wave-over-wave trends.

Political comparisons

In this section, we evaluate how well the COVID States survey aligns with other measures of political behavior, attitudes, and outcomes. First, we focus on the election in 2020, comparing our state-level estimates of vote share to others. Second, we compare our estimates of participation in political activities with a particular focus on participation in protests in response to racism and police violence during the summer of 2020. Finally we examine political attitudes comparing our feeling thermometer questions to those posed by the ANES.

State-by-State Comparison with 2020 election polls

No public opinion matter is polled as closely as a presidential election. Hence, these polls make for a perfect reference to gauge potential biases in our survey procedures compared with the state of the art of survey methodology.

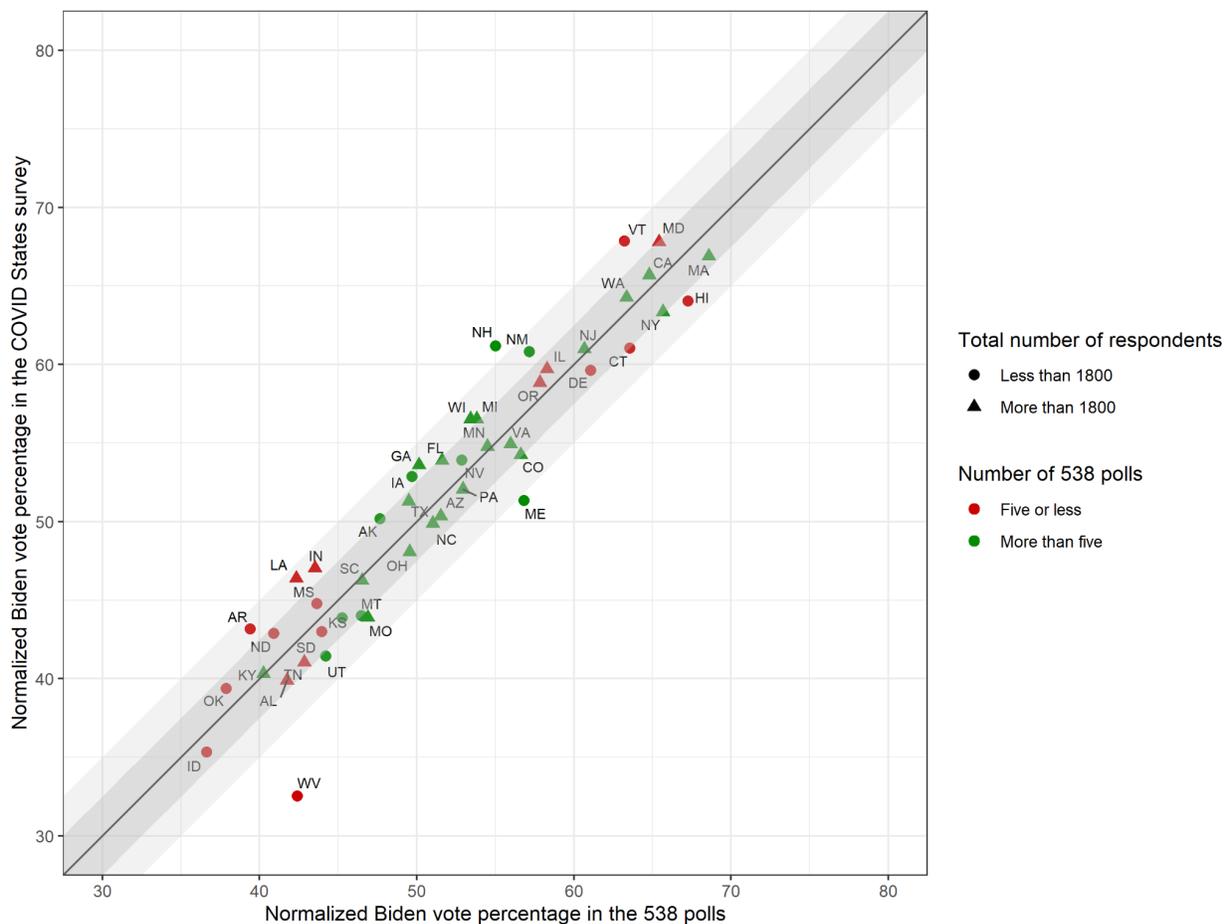
For this comparison, we used the polls listed at FiveThirtyEight, a polling aggregator and forecaster. We extracted the results of all 2020 presidential elections polls since the start of our first survey, April 16th 2020. For this comparison, we removed polls from a pollster with ratings of C/D or worse, and those without a pollster rating¹². For the comparison, the remaining polls' percentage of vote for Biden and Trump was averaged. No polls following this criteria were found for Wyoming and Rhode Island, so we excluded these states from the comparison. In addition, the only polls available for Nebraska were for congressional districts only, so we also excluded this state.

To calculate the vote intention from our survey, we pooled the data from our 9 pre-election state waves, keeping only the most recent answer from respondents appearing in more than one wave. We then reweighted the data using the same weighting procedure as described before. Then, we calculated the percentage of respondents who would vote for Trump or Biden over the number of respondents stating they are registered to vote and intending to vote or unsure about it. This filtering left a total of 91,908 respondents over the total of 124,558 respondents in the aggregated dataset, averaging 1802 respondents per state.

A large subsample of respondents, 10.5% of the sample at the national level, reported being unsure of whom they would vote for. As a result, the total intention to vote for either Trump or Biden is significantly lower in our survey than in election polls. Hence, our estimates of vote for both Biden and Trump are lower for all states than the estimates from the FiveThirtyEight polls. To solve this issue, we compared the percentage of vote for Biden calculated over the percentage of vote for either Biden or Trump, what we call normalized Biden vote. The result of this comparison is shown in Figure 9.

Overall, the values obtained from our survey line up well with the election polls, both in terms of absolute deviation from the election polls average and in terms of a consistent direction for the deviations. The absolute difference is higher than 5% for only three states: West Virginia, New Hampshire and Maine which have fewer respondents than the average state in our aggregated sample. In addition, the West Virginia average from the FiveThirtyEight polls is based on a single poll. The median absolute difference is 1.8%, and for 31 of the 47 states considered the absolute difference is below 2.5%. As a point of comparison, we calculated the standard deviation of the percentage of vote for Biden (over total vote) for the different polls of each state. The median standard deviation of these values is 2.3, and for Maine and New Hampshire it is 2.4 and 3.6, respectively. Hence, the intention to vote from our survey deviates from the average of election specialized polls for normalized Biden vote similarly to how percentage of Biden vote for these polls deviates from their average. In addition, the states for which our survey lines up worse are states for which election polls had an above average standard deviation.

¹² Fivethirtyeight.com rates pollsters on a scale from A to D mainly based on the accuracy of their past predictions (see <https://projects.fivethirtyeight.com/pollster-ratings/>). They regularly update their ratings and we used ratings as they were on election day.



Political Participation

To measure political participation, we adapted the American National Election Survey's (ANES) Political Participation questionnaire. We asked about participation in the past six months rather than the past 12 months and changed some of the questions slightly. We began asking these questions starting in Wave 10, September 2020. In Table 5, we compare our pooled responses from September through November 4th. We see similar magnitudes in political participation for similar questions - volunteering, attending meetings, contacting representatives, and contributing money. We see larger differences in rates of posting comments online (14 points) and making a donation (6 points different). These differences may be due to question wording and difference in 12 vs 6 month timing. Our questions about attending events are likely

incomparable given their different wording. Given the substantial differences in our question wording, these comparisons do not provide the apples-to-apples comparisons we would like. But they do offer insight into the areas of agreement and disagreement with existing gold-standard data sources. We have better comparisons with a more targeted question: estimating protest participation in the summer of 2020.

ANES Item	2020 Result	COVID States Item	Pooled Result Sept-Nov 2020
Contacted member of US Senate or House of Rep	14.1%	Called or wrote to an elected official	11.4%
Posted comment online about political issue	37.4%	Posted about politics on social media	22.7%
Contribute money to individual candidate running for public office	16.6%	Made a donation to a candidate, party, or other political organization	10.8%
Do any (other) work for party or candidate	2.9%	Volunteered for a candidate, political party, or other political organization	5.1%
Go to any political meetings, rallies, speeches, dinners	5.3%	Attended a rally or protest	7.0%
Attend online political meetings, rallies, speeches, fundraisers	11.7%	Attended a town hall held by an elected official	3.6%

We also measured political participation in protests that arose during the summer of 2020. For our June 12-28 wave (Wave 5) we asked participants “When, if ever, did you last attend a rally, protest, or vigil about racism and/or police violence?” with the possible responses “In the past month” “More than a month but less than a year ago” “More than a year ago” and “Never.” Based on these responses, we estimated that 4.6% (± 0.32) of adults attended a protest about “racism and/or police violence” over the past 30 days. Pew’s American Trends Panel’s June 4-10 survey reported 6% (± 1.6) “in the last month” (Pew Research Center 2020a). NORC’s June 11-14 survey (2020), which asked about protest attendance in “the past few weeks,” estimated turnout at 7% (± 3.7). Thus, our estimate falls within Pew and NORC’s margins of error. Two other pollsters found slightly higher turnout rates. Civis Analytics (2020) over two waves (June 12-25) found 8.6% participation “in the last four weeks,”¹³ while the

¹³ No margin of error was reported, but their sample size of 4446 implies an unweighted margin of error of ± 0.82 .

Kaiser Family Foundation (2020) (June 8-14) reported 9% (± 3) of the population attended “in the past three months.”¹⁴

We also examined the correlates of protest participation (Table 6). Like Civis, we found slightly higher rates of protest participation from men than women (1.5 points in ours, 0.8 points in Civis). Pew, NORC, Civis, and our survey all find that Blacks had about twice the likelihood of participating as Whites (7.4:3.8 in our survey, 10:5 in Pew, 13:6 in NORC, and 16:7 in Civis).¹⁵ We estimate decreasing rates of protest participation in each older age bracket. We find less participation for Republicans and those approving of President Trump than do Pew and Civis (NORC and Kaiser don’t report these breakdowns). Both Civis and our survey found the probability of attendance to 1) increase with education and income; 2) be roughly twice as high for adults living with children under 18 than those without; 3) higher for single than married individuals, who are in turn higher than divorcees, and 3) higher in the Northeast with similar rates across the Midwest, West, and South. (Other surveys did not report these correlations). Kaiser and Pew also report the composition of protesters themselves, broken down by age, education, party, race, and urbanicity. While their estimates tend to diverge widely, our estimates usually fall in between the two and generally agree ordinarily if not cardinally.

Demographic	COVID States	Kaiser ⁺	Pew ⁺⁺
High school or less	26.6 [23.0, 30.1]	28	n/a
Some college	28.8 [25.6, 32]	20	n/a
College	44.6 [41.2, 48.0]	53	n/a
Democrat	50.8 [47.3, 54.4]	42	n/a
Democrats/Lean D	62.2 [58.7, 65.6]	n/a	79
Republican	16.0 [13.4, 18.5]	6	n/a
Republican/Lean R	20.0 [17.2, 22.8]	n/a	17
Independent	27.3 [24.0, 30.5]	46	n/a
Ages 18-29	46.3 [42.8, 49.8]	52	41
Ages 30-49	40.7 [37.3, 44.2]	33	38
Ages 50+	13.0 [10.3, 15.7]	11	21

¹⁴ We also asked participants “When, if ever, did you last attend a rally, protest, or vigil about reopening, quarantine, or coronavirus restrictions?” We estimate 0.6% of the population attended a protest about quarantine, on par with the 1% reported by Kaiser.

¹⁵ Kaiser reports a smaller ratio of 13:10, but their margin of error for Blacks is ± 9 .

Urban	41.4 [37.8 44.9]	n/a	41
Suburban	48.8 [45.2 52.3]	n/a	42
Rural	9.8 [8 11.7]	n/a	17
White	53.2 [49.7 56.8]	n/a	46
Black	19.5 [16.8 22.1]	n/a	17
Hispanic	17.7 [14.3 21.2]	n/a	22
Asian	5.5 [4.3 6.7]	n/a	8
+ From Kaiser Family Foundation 2020 ++ From (Pew Research Center 2020a)			

Feeling Thermometer

Our survey included a battery of “feeling thermometer” items, asking respondents to rate various individuals, groups, and institutions on a 0-100 scale. These items are common in survey research, especially in political surveys. Here, we compare six such items that overlap with those asked on the American National Election Studies (ANES) 2020 post-election wave -- both in date range and question wording. The ANES is widely considered the gold-standard for regular, publicly-available survey research in American Politics. As feeling thermometer items are particularly prone to mode effects (Homola, Jackson, and Gill 2016), we consider the ANES’s online-only sample for an apples-to-apples comparison.

In Figure 10, we see general agreement between the feeling thermometer ratings in our survey and those in the ANES, including when responses are broken out by respondents’ seven-point partisan identification. While there are cases where base rates differ (most notably, feelings toward racial groups tend to be slightly higher in the COVID States survey than in the ANES), partisan trends in feeling thermometer ratings are consistent across surveys. Democrats holding warmer views toward scientists, Black people, and Hispanic people and Republicans holding warmer views toward white people and the police.

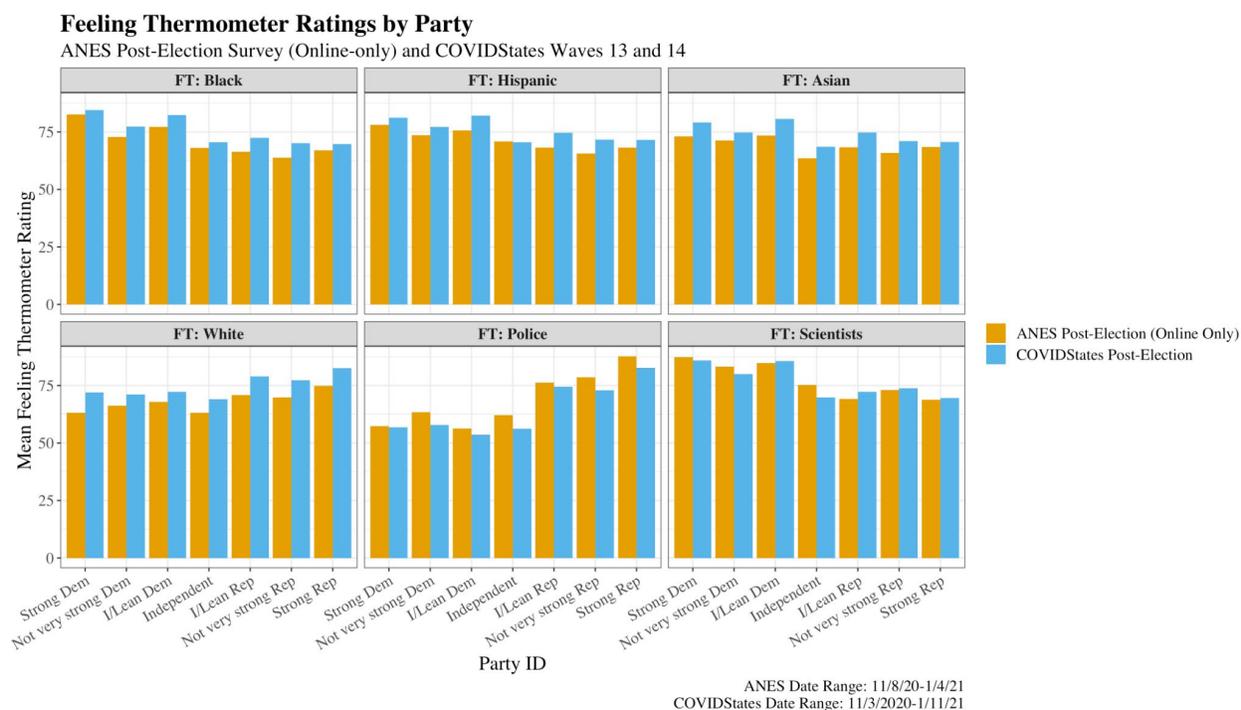


Figure 10: Comparison of Feeling Thermometer between COVID States and 2020 ANES

Summary

Our measures of political attitudes, beliefs, and behaviors fit well with prevailing standards. Our polls matched other polls in predicting Biden vote share at the state level. While our measure of everyday political activity diverged somewhat from the ANES due to item construction issues, our estimate of participation in racial justice and anti-pandemic restriction protests align broadly with other studies, both in point estimates and in correlations with who was likely to attend. Finally, our measure of sentiment towards others closely tracked that measured by the ANES.

Direct Comparison to Kaiser Family Foundation’s CVM Poll

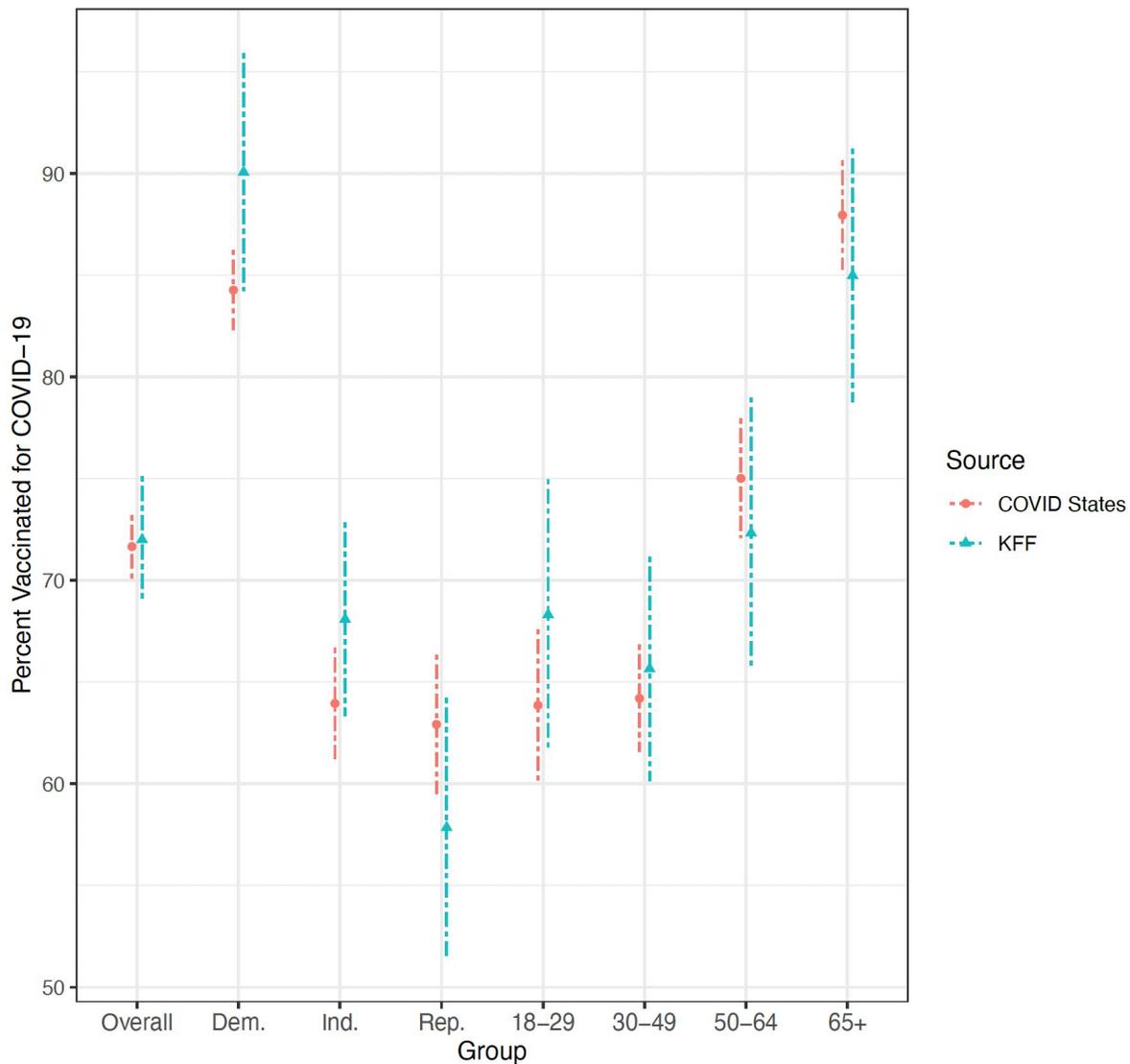
Thus far, we have compared our results to surveys asking similar questions around the same time. This raises questions about the temporal validity of our results and noise from variations in question prompt and response choices. That’s why, in September 2021, we conducted an assessment of the consistency of our data relative to the data collected by the Kaiser Family Foundation’s (KFF) COVID-19 Vaccine Monitor (CVM) (Kaiser Family Foundation 2021). We fielded a set of shared questions during overlapping time periods to provide a validation on more comparable data. We compare our poll results across a range of responses and across demographic subgroups.

KFF’s CVM is a source of probability, public health polling data widely considered to be the “gold standard” in public opinion research. The experiment asked several identical questions as KFF’s September CVM, including questions on whether the respondent is vaccinated, whether they feel getting vaccinated is mostly a personal choice or everyone’s responsibility to protect the health of others, and items asking support for employer, college, and K-12 schools

vaccine mandates. To assess the consistency of COVID States' results with KFF's results, we first subsetted our data to include only respondents who completed the survey during the time period in which KFF's September CVM was in the field (September 13-22, 2021), yielding a sample size of 5,233 respondents. We then weighted our data using demographic characteristics to match the U.S. population with respect to race/ethnicity, age, gender, education, and living in urban, suburban, or rural areas.

The results of this study indicated that the COVID States method performs well compared to KFF when it comes to self-reported vaccination rates. The overall share of respondents already vaccinated for COVID-19 was identical to the share reported by KFF in its September CVM (72%). The share of individuals vaccinated by gender, age, educational attainment, and race were within 5 percentage points of the results reported by KFF in their September CVM. As shown in Figures 11 and 12, confidence intervals for COVID States estimates of self-reported vaccination rates overall and by demographic subgroups (including party, age, education, race, and gender) are within the range of the margin of sampling error (MOSE) for the September CVM KFF estimates for all groups.¹⁶

¹⁶ KFF data was downloaded from the Roper Center and demographic groups were recoded to match COVID States demographic groups as closely as possible. However, note that the race question is notably different between KFF and COVID States.



Notes: Error bars for KFF are the Margin of Sampling Error (MOSE) for each group. Error bars for COVID States are 95% error intervals around percentage.

Figure 11. Self-Reported COVID-19 Vaccination Rates Overall, by Party, and Age, COVID States versus KFF’s September CVM.

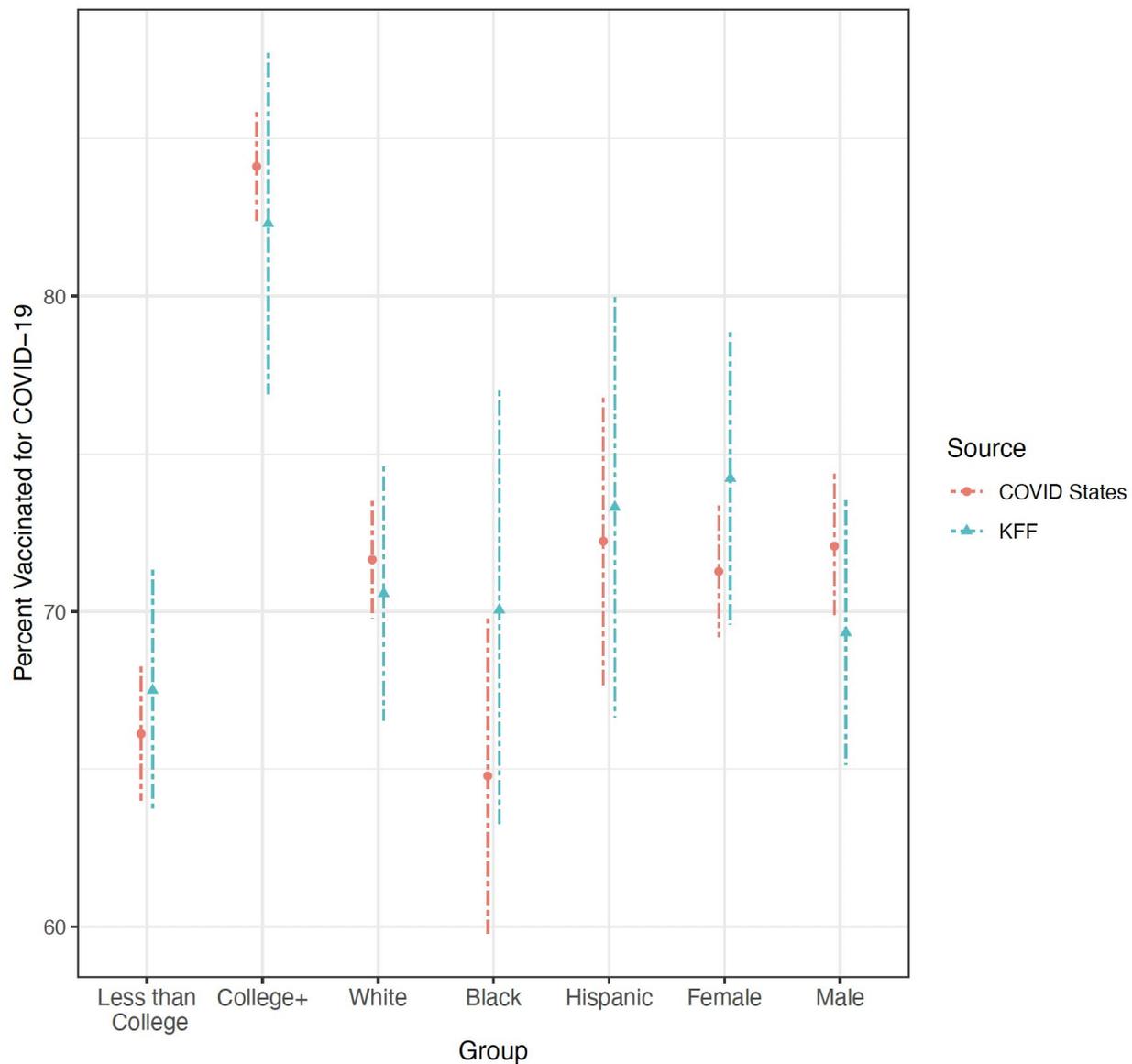


Figure 12. Self-Reported COVID-19 Vaccination Rates by Education, Race and Gender, COVID States versus KFF's September CVM.

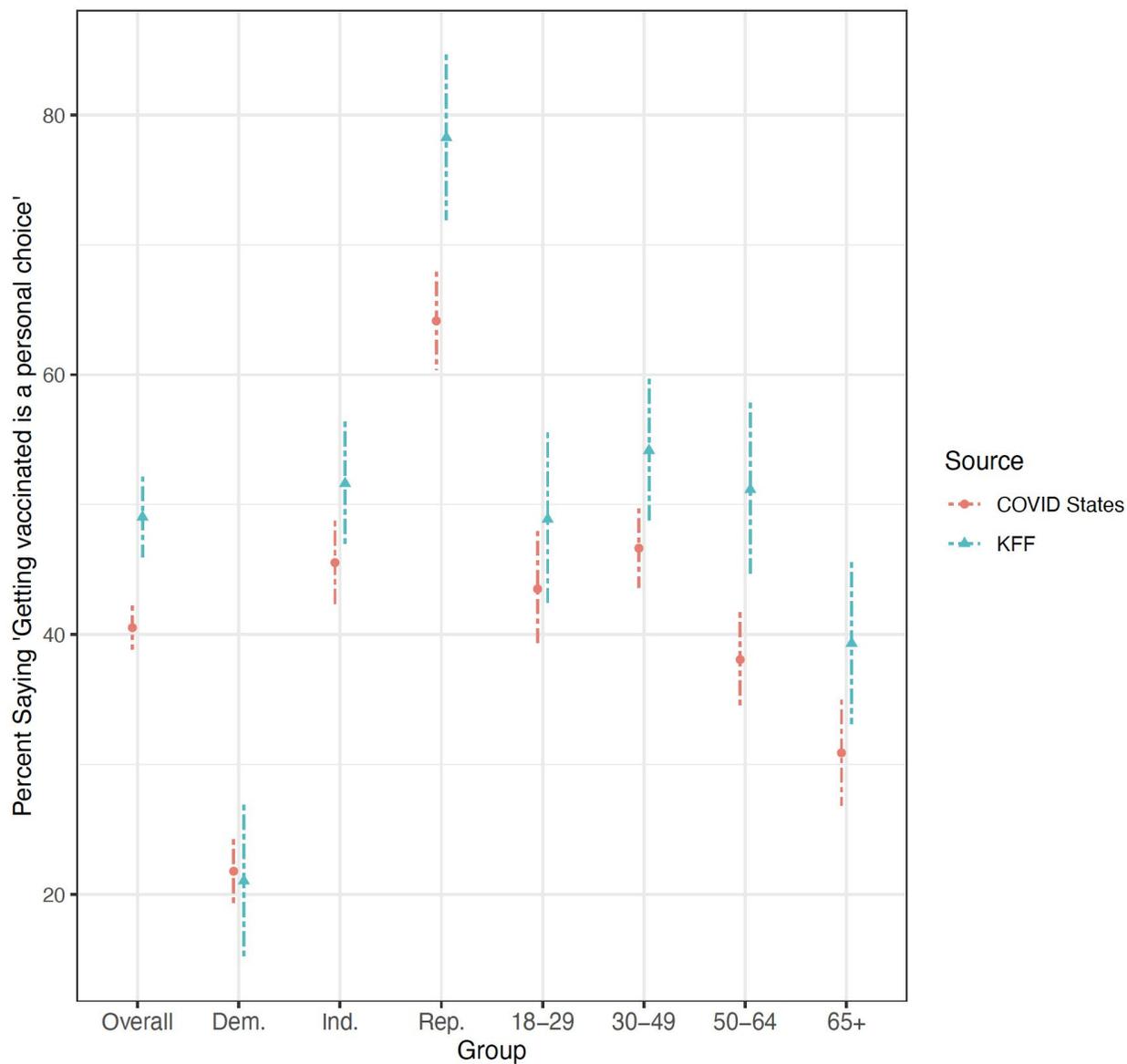


Figure 13. Share of Respondents Who Say Getting Vaccinated is a Personal Choice Overall, by Party, and Age, COVID States versus KFF's September CVM.

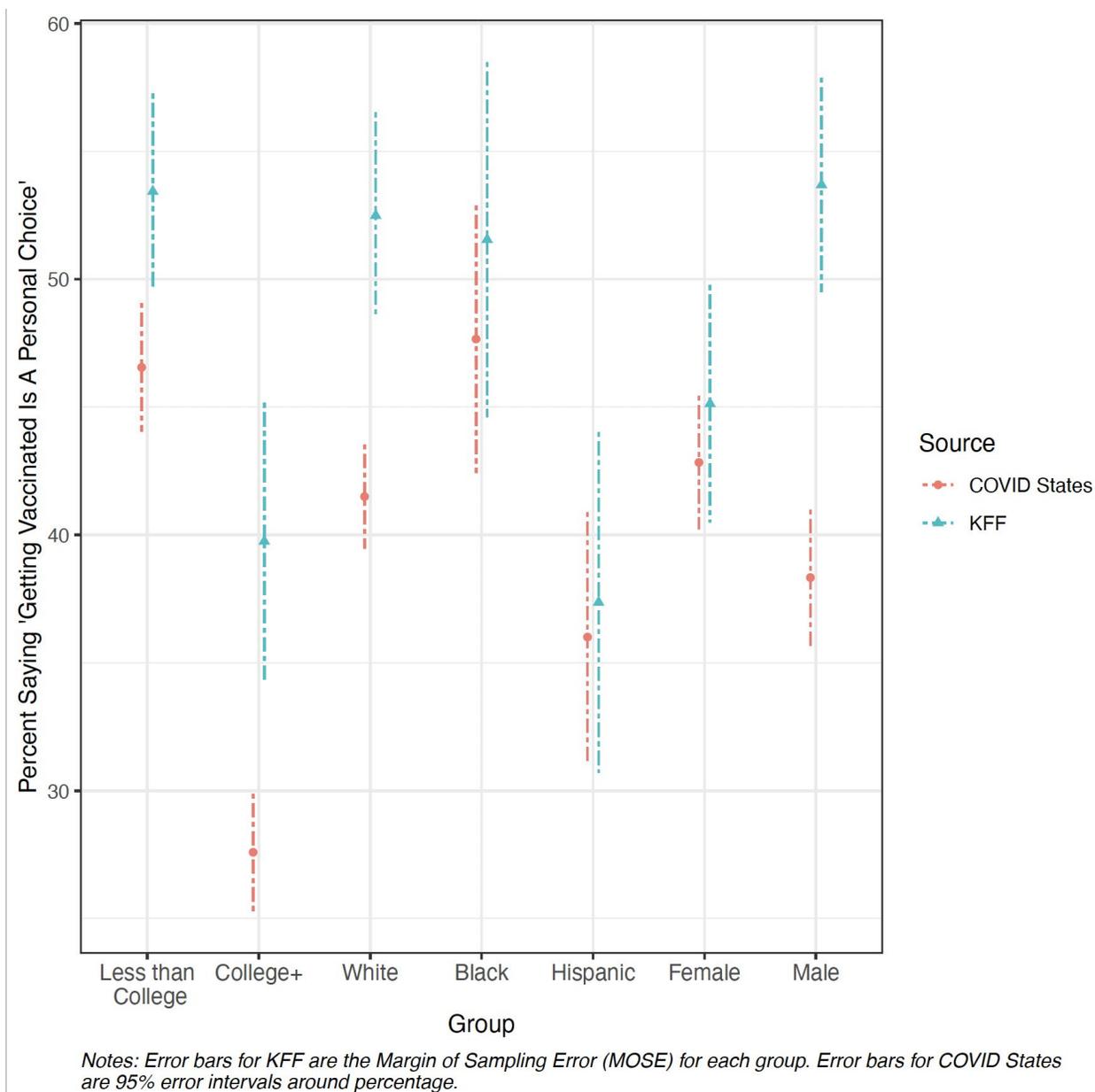


Figure 14. Share of Respondents Who Say Getting Vaccinated is a Personal Choice by Education, Race, and Gender, COVID States versus KFF’s September CVM.

Unlike the roughly equivalent performance of COVID-19 vaccination rates, significant discrepancies exist overall, among Republicans, among 50-64 year olds, high school educated, college educated, Whites and men (see figures 13 and 14). Among all these groups, COVID States finds a significantly lower share of respondents reporting that getting vaccinated is a personal choice relative to KFF’s CVM. Given the highly partisan nature of the question, we suspect that our survey underestimates the share of ideological conservatives across these groups relative to KFF’s sample.

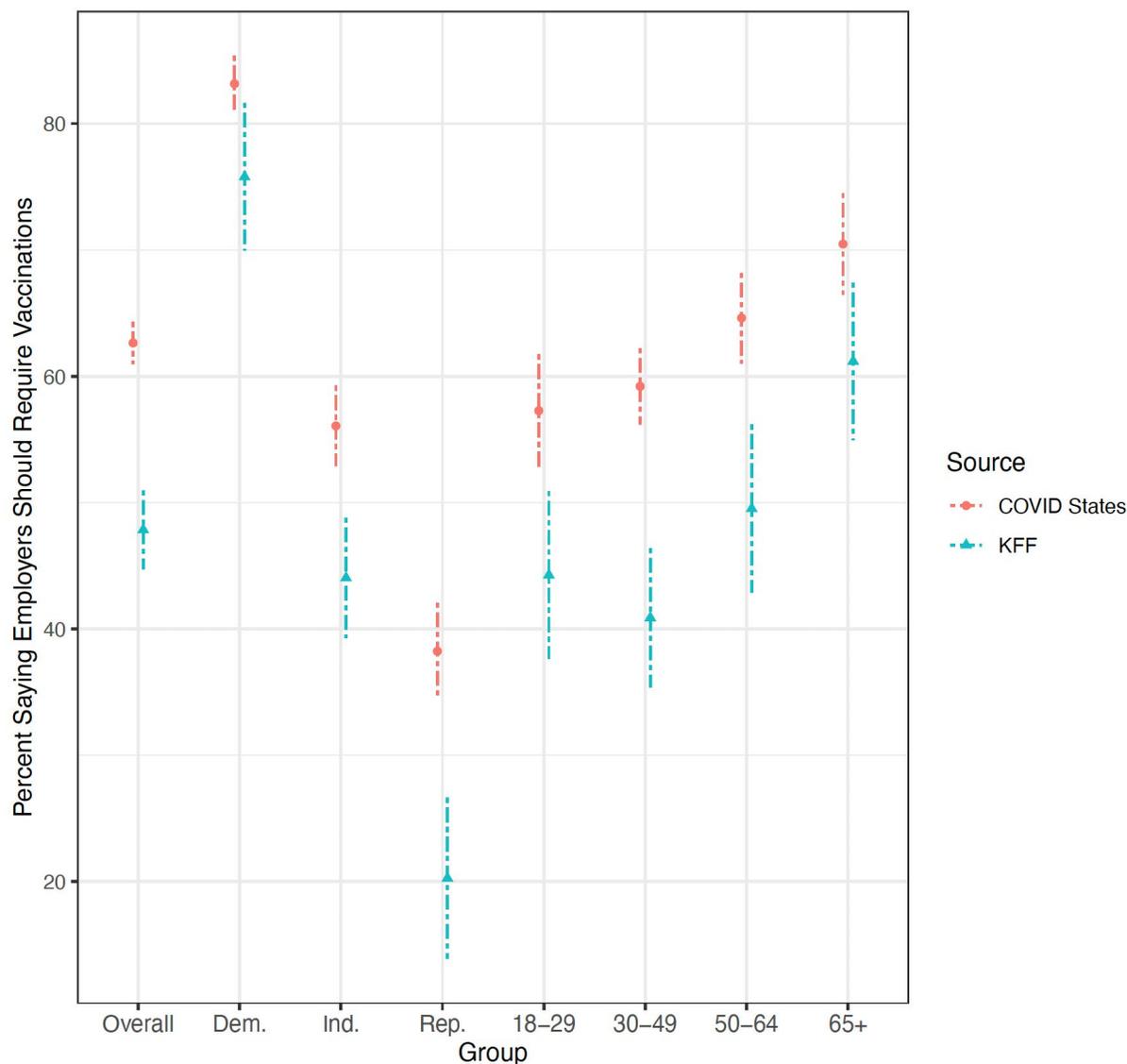
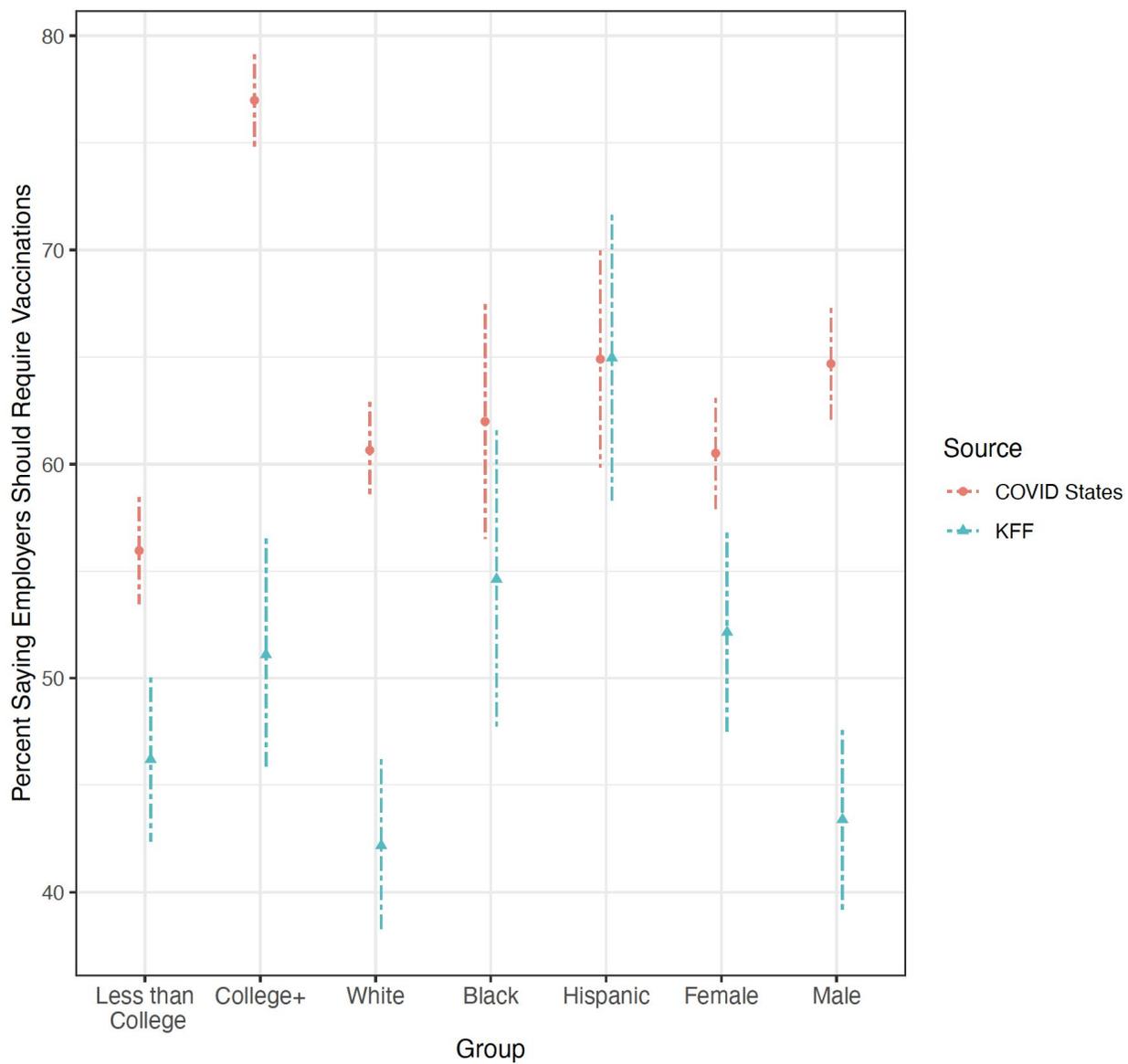


Figure 15. Share of Respondents Who Support Employer Vaccine Requirements Overall, by Party, and Age, COVID States versus KFF’s September CVM.

Similarly, as shown in Figures 15 and 16, we observe significant differences overall and across subgroups so respondents in support for employer vaccination requirements. Across all groups, COVID States finds significantly higher shares of respondents supporting employers requiring employees to get vaccinated. The only overlapping estimates are among Black and Hispanic respondents.



Notes: Error bars for KFF are the Margin of Sampling Error (MOSE) for each group. Error bars for COVID States are 95% error intervals around percentage.

Figure 16. Share of Respondents Who Support Employer Vaccine Requirements by Education, Race, and Gender, COVID States versus KFF's September CVM.

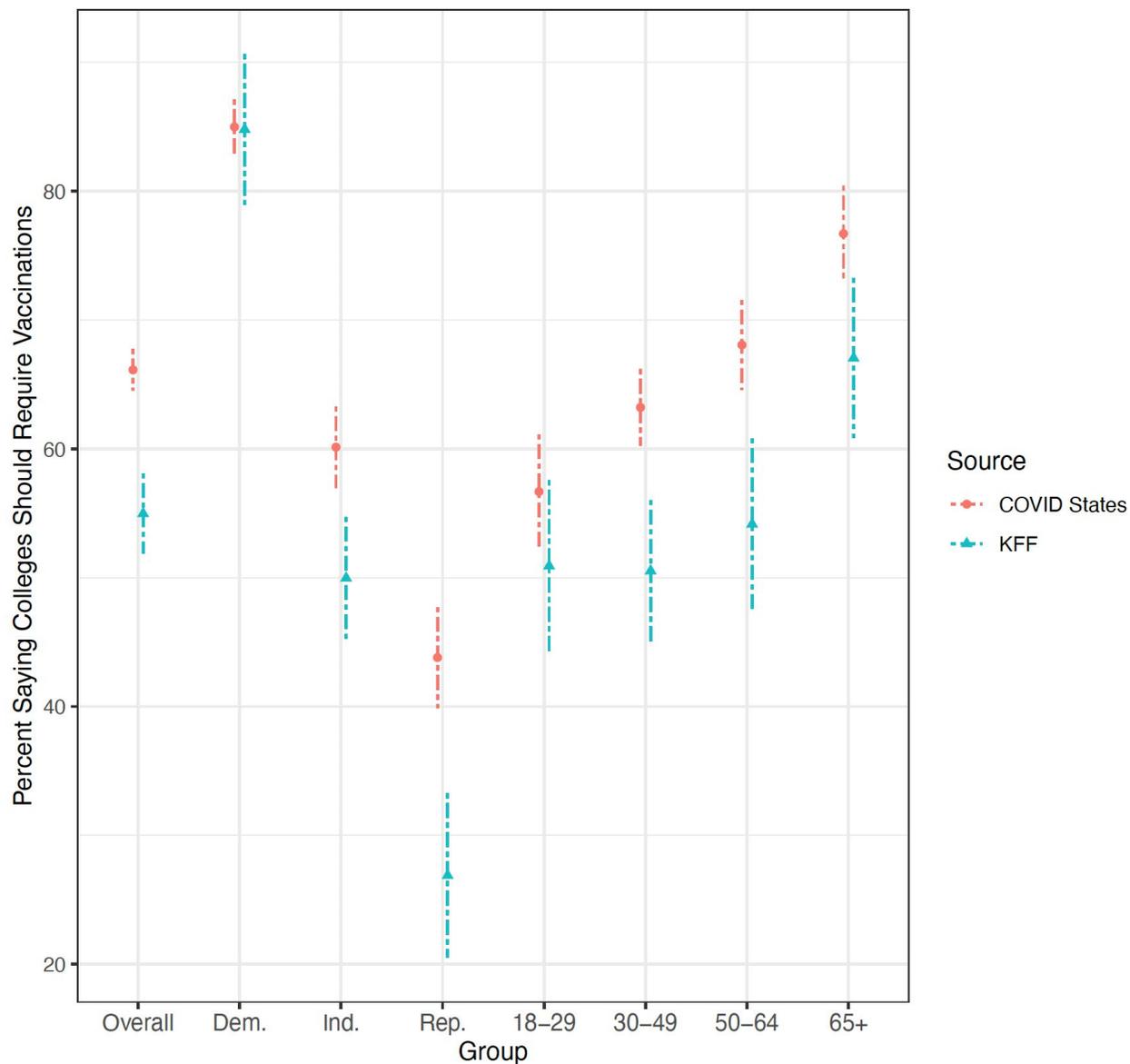


Figure 17. Share of Respondents Who Support College Vaccine Requirements Overall, by Party, and Age, COVID States versus KFF's September CVM.

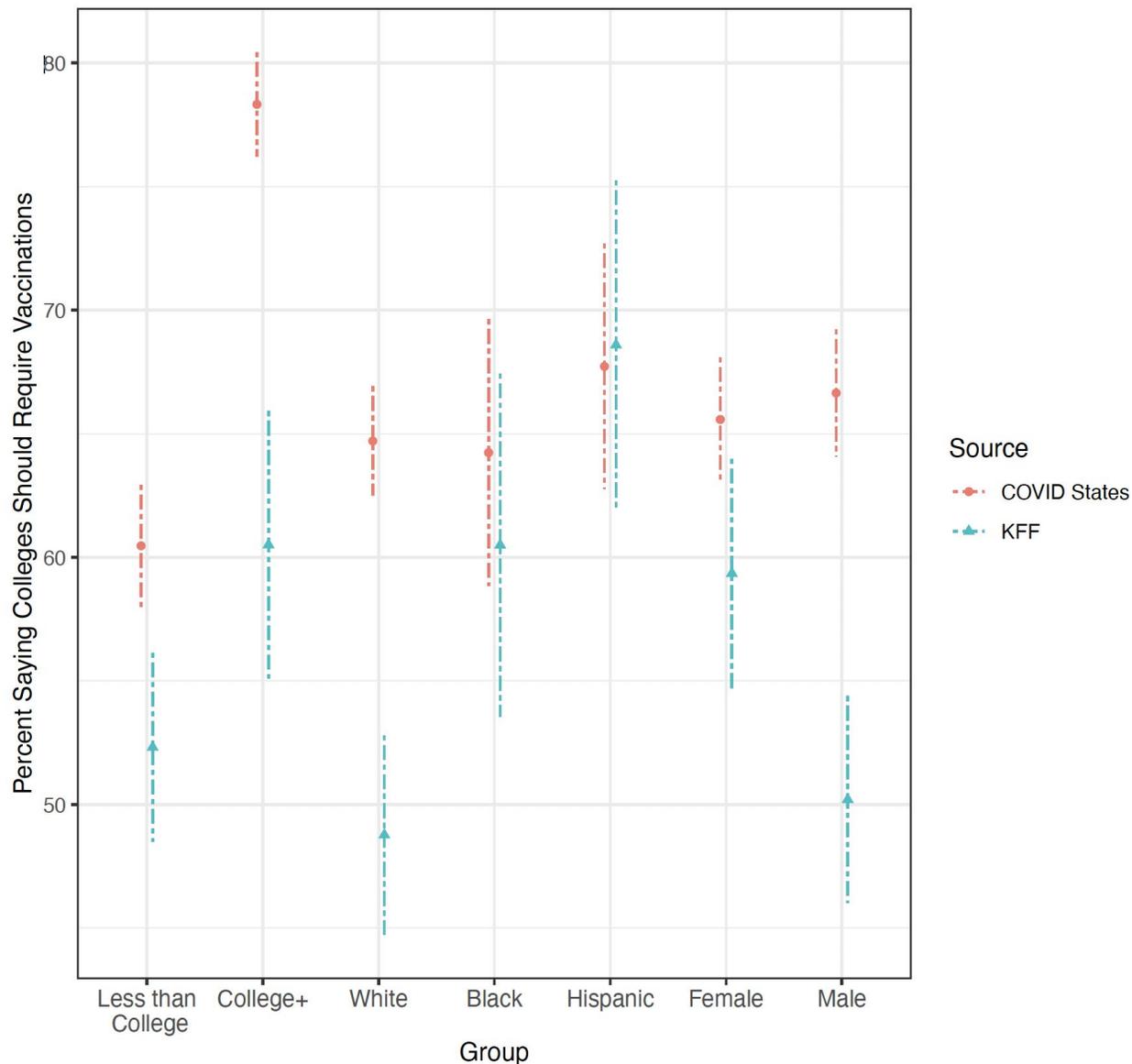


Figure 18. Share of Respondents Who Support College Vaccine Requirements by Education, Race, and Gender, COVID States versus KFF's September CVM.

As shown in Figures 17 and 18, COVID States estimates for support for college vaccine requirements are also significantly higher than KFF estimates overall, among independents, Republicans, ages 30 and up, those with less than a college education, those with some college education or more, Whites, and men. COVID States similarly significantly overestimates (relative to KFF) the share of respondents overall who support vaccine requirements in K-12 schools, as well as among Republicans, 30-64 year olds, college educated, non-college educated, Whites, and men.

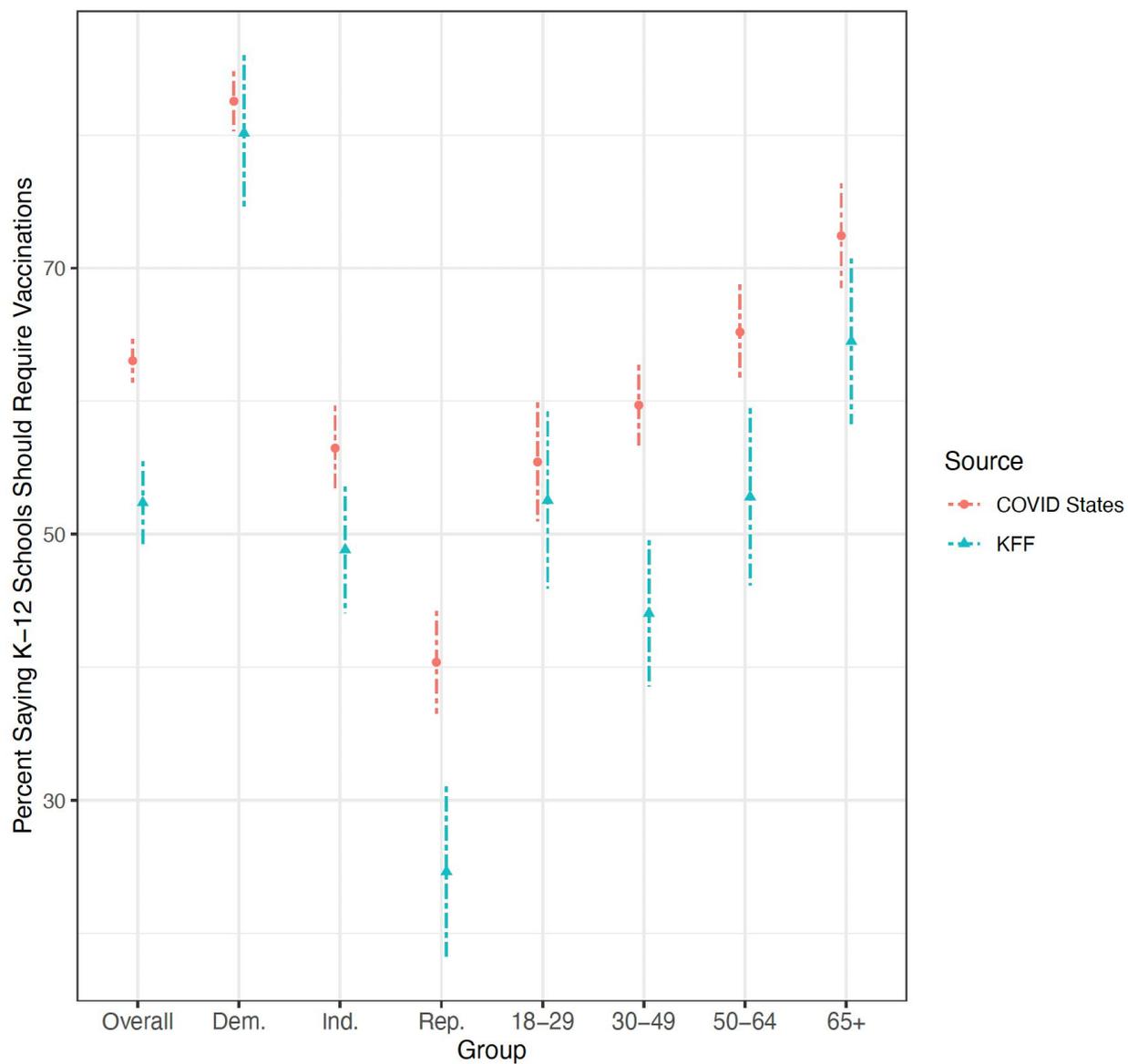


Figure 19. Share of Respondents Who Support K-12 Schools Vaccine Requirements Overall, by Party, and Age, COVID States versus KFF's September CVM.

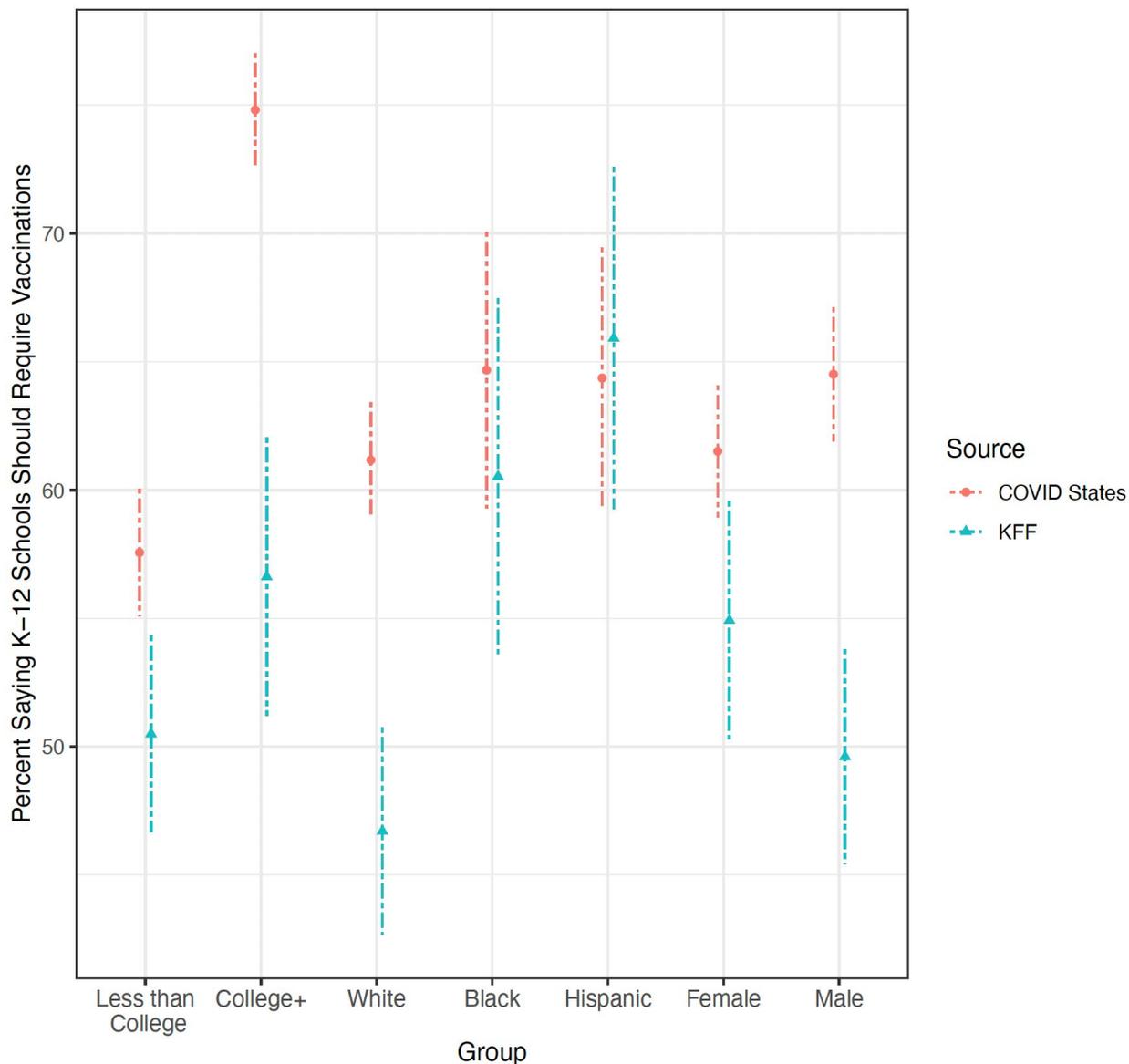


Figure 20. Share of Respondents Who Support K-12 Schools Vaccine Requirements by Education, Race, and Gender, COVID States versus KFF's September CVM.

Taken together, COVID States estimates of vaccination rates are consistent with KFF vaccination rates, however COVID States diverges from KFF in terms of attitudes towards vaccines and vaccine mandates. Specifically, COVID States finds a significantly higher share of respondents support vaccine mandates and a significantly lower share who say that getting vaccinated is mostly a personal choice (rather than everyone's responsibility to protect the health of others). This pattern could reflect more liberal leaning respondents relative to KFF's sample. This explanation is partially borne out by the data: 32% of COVID States respondents during this period self-identified as liberal and 28% as conservative, compared with 27% liberal

and 34% conservative among KFF's respondents which is closer to other national estimates (Saad 2020).

Discussion

Overall, estimates from our survey closely track estimates from large scale administrative data and probability-based survey data: COVID prevalence, vaccination rates, depressive symptoms, and state-level election polls. On these, our estimates diverge between 1 and 7 percentage points, with larger differences occurring from less standardized survey items. The results for election polls and vaccination rates are particularly important because they provide broad validation that estimates from our survey are representative at the state-level. Finally, we see similar patterns of change over time and the same correlations between key measures as other surveys. This suggests that our data can broadly be used for representative estimates and statistical analyses of temporal changes and relationships between measures.

However we do find some potential sources of bias. First we substantially undersample Latinos and people without high school degrees, leading to a potential over-reliance on weights. Researchers performing analyses of results driven by these groups should exercise caution. Second, we also find some large discrepancies in measuring support for vaccine mandates particularly among conservatives and conservative-leaning demographic groups. While our under-sampling of conservatives is not large, we do not re-weight our data on ideology. This means all of our estimates have this imbalance baked in. As such, estimates from our data should take into account the under-representation of ideological conservatives. Moreover, the under-representation of certain groups can affect statistical estimates of relationships between items. In looking for temporal or statistical relationships then, we encourage analysts to take ideology into account when conducting their analysis and interpreting their results.

Conclusion

Through bias and error, all surveys, whether based on non-probability or probability samples, fail to be representative to one degree or another. The generalizability of a survey can only truly be evaluated empirically after it has been conducted. Here we've demonstrated what that evaluation should involve: 1) auditing the representativeness of the respondents and weights applied and 2) comparing its estimates to those from other data sources.

Despite substantial efforts by the American Association for Public Opinion Research's Transparency Initiative to develop standards for survey transparency, many surveys are not conducted or reported in such a way that their results can be replicated or validated. This makes it impossible to know whether any given estimate from that survey may be model-dependent, biased, or the result of error. In fact, many pollsters consider their weighting schema to be proprietary information, the "secret sauce" that makes their surveys more generalizable. In reality, this stance makes it more difficult to assess the value of these polls and understand why polls vary from one another or miss systematically, such as when they collectively under-predicted conservative turnout in two presidential election cycles in the U.S.

With the COVID States survey, our goal has been to execute the best survey possible and provide the transparency needed for others to assess the generalizability of its estimates.

This allows academic and non-academic audiences alike to interpret any results with an understanding of the strengths and weaknesses of the underlying sample. As non-probability sampling becomes more and more common, we hope these practices become part of the standards for survey reporting. Such transparency will enable these surveys to contribute to our understanding of the social world in proportion to their epistemic value.

References

- Ansolabehere, Stephen, and Brian F. Schaffner. 2014. "Does Survey Mode Still Matter? Findings from a 2010 Multi-Mode Comparison." *Political Analysis* 22(3):285–303. doi: 10.1093/pan/mpt025.
- AP-NORC. 2020a. "Toplines from December 2020 AP-NORC Center Poll."
- AP-NORC. 2020b. "Toplines from June 2020 AP-NORC Center Poll."
- AP-NORC. 2020c. "Toplines from March 2020 AP-NORC Center Poll."
- AP-NORC. 2020d. "Toplines from May 2020 AP-NORC Center Poll."
- AP-NORC. 2021. "Toplines from February 2021 AP-NORC Center Poll."
- Associated Press. 2021. "Data Correction Drops COVID Vaccine Rates in Pennsylvania." *AP News*, November 26.
- Baker, R., S. J. Blumberg, J. M. Brick, M. P. Couper, M. Courtright, J. M. Dennis, D. Dillman, M. R. Frankel, P. Garland, R. M. Groves, C. Kennedy, J. Krosnick, P. J. Lavrakas, S. Lee, M. Link, L. Piekarski, K. Rao, R. K. Thomas, and D. Zahs. 2010. "Research Synthesis: AAPOR Report on Online Panels." *Public Opinion Quarterly* 74(4):711–81. doi: 10.1093/poq/nfq048.
- Baker, Reg, J. Michael Brick, Nancy A. Bates, Mike Battaglia, Mick P. Couper, Jill A. Dever, Krista J. Gile, and Roger Tourangeau. 2013. "Summary Report of the AAPOR Task Force on Non-Probability Sampling." *Journal of Survey Statistics and Methodology* 1(2):90–143.
- Berinsky, A. J., G. A. Huber, and G. S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.Com's Mechanical Turk." *Political Analysis* 20(3):351–68. doi: 10.1093/pan/mpr057.
- Bradley, Valerie C., Shiro Kuriwaki, Michael Isakov, Dino Sejdinovic, Xiao-Li Meng, and Seth Flaxman. 2021. "Unrepresentative Big Surveys Significantly Overestimated US Vaccine Uptake." *Nature*. doi: 10.1038/s41586-021-04198-4.
- CDC Case Surveillance Task Force. 2021. "COVID-19 Case Surveillance Public Use Data with Geography." Retrieved (<https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data-with-Ge/n8mc-b4w4>).
- Centers for Disease Control and Prevention. 2021. "COVID-19 Vaccinations in the United States, County." Retrieved (<https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh>).
- Civis Analytics. 2020. "Public Opinion Data on Black Lives Matter Police Reform." Retrieved (https://www.civisanalytics.com/wp-content/uploads/2020/06/Public_Opinion_Data_BLM_CombinedCrosstabs_ALL.pdf).
- Coppock, Alexander, and Oliver A. McClellan. 2019. "Validating the Demographic, Political,

- Psychological, and Experimental Results Obtained from a New Source of Online Survey Respondents." *Research & Politics* 6(1):205316801882217. doi: 10.1177/2053168018822174.
- Druckman, James N., Katherine Ognyanova, Matthew A. Baum, David Lazer, Roy H. Perlis, John Della Volpe, Mauricio Santillana, Hanyu Chwe, Alexi Quintana, and Matthew Simonson. 2021. "The Role of Race, Religion, and Partisanship in Misperceptions about COVID-19." *Group Processes & Intergroup Relations* 24(4):638–57. doi: 10.1177/1368430220985912.
- Druckman, Jamie. 2022. *Experimental Thinking: A Primer on Social Science Experiments*. Cambridge, United Kingdom ; New York, NY: Cambridge University Press.
- Enns, Peter, and Jake Rothschild. 2021. "Revisiting the 'Gold Standard' of Polling: New Methods Outperformed Traditional Ones in 2020." *Medium*. Retrieved March 2, 2022 (<https://medium.com/3streams/revisiting-the-gold-standard-of-polling-new-methods-outperformed-traditional-ones-in-2020-451650a9ba5b>).
- Gallup. 2021. "Party Affiliation." Retrieved (<https://news.gallup.com/poll/15370/party-affiliation.aspx>).
- Homola, Jonathan, Natalie Jackson, and Jeff Gill. 2016. "A Measure of Survey Mode Differences." *Electoral Studies* 44:255–74. doi: 10.1016/j.electstud.2016.06.010.
- Hughes, Adam G., Stefan D. McCabe, William R. Hobbs, Emma Remy, Sono Shah, and David M. J. Lazer. 2021. "Using Administrative Records and Survey Data to Construct Samples of Tweeters and Tweets." *Public Opinion Quarterly* 85(S1):323–46. doi: 10.1093/poq/nfab020.
- Ipsos. 2020. "Axios-Ipsos Coronavirus Index."
- Ipsos. 2021. "Axios-Ipsos Coronavirus Index, Waves 34-48."
- Irons, Nicholas J., and Adrian E. Raftery. 2021. "Estimating SARS-CoV-2 Infections from Deaths, Confirmed Cases, Tests, and Random Surveys." *Proceedings of the National Academy of Sciences* 118(31):e2103272118. doi: 10.1073/pnas.2103272118.
- Kaiser Family Foundation. 2020. "KFF Health Tracking Poll - June 2020." Retrieved (<https://www.kff.org/racial-equity-and-health-policy/report/kff-health-tracking-poll-june-2020/>).
- Kaiser Family Foundation. 2021. "KFF COVID-19 Vaccine Monitor: September 2021." Retrieved (<https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-september-2021/>).
- Kalish, Heather, Carleen Klumpp-Thomas, Sally Hunsberger, Holly Ann Baus, Michael P. Fay, Nalyn Siripong, Jing Wang, Jennifer Hicks, Jennifer Mehalko, Jameson Travers, Matthew Drew, Kyle Pauly, Jacquelyn Spathies, Tran Ngo, Kenneth M. Adusei, Maria Karkanitsa, Jennifer A. Croker, Yan Li, Barry I. Graubard, Lindsay Czajkowski, Olivia Belliveau, Cheryl Chairez, Kelly R. Snead, Peter Frank, Anandakumar Shunmugavel, Alison Han, Luca T. Giurgea, Luz Angela Rosas, Rachel Bean, Rani Athota, Adriana

- Cervantes-Medina, Monica Gouzoulis, Brittany Heffelfinger, Shannon Valenti, Rocco Caldararo, Michelle M. Kolberg, Andrew Kelly, Reid Simon, Saifullah Shafiq, Vanessa Wall, Susan Reed, Eric W. Ford, Ravi Lokwani, John-Paul Denson, Simon Messing, Sam G. Michael, William Gillette, Robert P. Kimberly, Steven E. Reis, Matthew D. Hall, Dominic Esposito, Matthew J. Memoli, and Kaitlyn Sadtler. 2021. "Undiagnosed SARS-CoV-2 Seropositivity during the First 6 Months of the COVID-19 Pandemic in the United States." *Science Translational Medicine* 13(601):eabh3826. doi: 10.1126/scitranslmed.abh3826.
- Kennedy, Courtney, Andrew Mercer, Scott Keeter, Nick Hatley, Kyley McGeeney, and Alejandra Gimenez. 2016. *Evaluating Online Nonprobability Surveys*. Pew Research Center.
- Kennedy, Ryan, Scott Clifford, Tyler Burleigh, Philip D. Waggoner, Ryan Jewell, and Nicholas J. G. Winter. 2020. "The Shape of and Solutions to the MTurk Quality Crisis." *Political Science Research and Methods* 8(4):614–29. doi: 10.1017/psrm.2020.6.
- Kroenke, K., R. L. Spitzer, J. B. W. Williams, and B. Lowe. 2009. "An Ultra-Brief Screening Scale for Anxiety and Depression: The PHQ-4." *Psychosomatics* 50(6):613–21. doi: 10.1176/appi.psy.50.6.613.
- Kroenke, Kurt, and Robert L. Spitzer. 2002. "The PHQ-9: A New Depression Diagnostic and Severity Measure." *Psychiatric Annals* 32(9):509–15. doi: 10.3928/0048-5713-20020901-06.
- Lu, Fred S., Andre T. Nguyen, Nicholas B. Link, Mathieu Molina, Jessica T. Davis, Matteo Chinazzi, Xinyue Xiong, Alessandro Vespignani, Marc Lipsitch, and Mauricio Santillana. 2021. "Estimating the Cumulative Incidence of COVID-19 in the United States Using Influenza Surveillance, Virologic Testing, and Mortality Data: Four Complementary Approaches" edited by C. Viboud. *PLOS Computational Biology* 17(6):e1008994. doi: 10.1371/journal.pcbi.1008994.
- Lumley, Thomas. 2020. "Survey: Analysis of Complex Survey Samples."
- MacInnis, Bo, Jon A. Krosnick, Annabell S. Ho, and Mu-Jung Cho. 2018. "The Accuracy of Measurements with Probability and Nonprobability Survey Samples: Replication and Extension." *Public Opinion Quarterly* 82(4):707–44. doi: 10.1093/poq/nfy038.
- NORC. 2020. "The June 2020 AP-NORC Center Poll." Retrieved (https://apnorc.org/wp-content/uploads/2020/06/Topline_final_release5.pdf).
- NPR/PBS NewsHour/Marist. 2020. "Toplines from the NPR/PBS NewsHour/Marist Poll."
- Peer, Eyal, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. "Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research." *Journal of Experimental Social Psychology* 70:153–63. doi: 10.1016/j.jesp.2017.01.006.
- Perlis, Roy H., Katherine Ognyanova, Mauricio Santillana, Matthew A. Baum, David Lazer, James Druckman, and John Della Volpe. 2021. "Association of Acute Symptoms of COVID-19 and Symptoms of Depression in Adults." *JAMA Network Open* 4(3):e213223. doi: 10.1001/jamanetworkopen.2021.3223.

- Perlis, Roy H., Mauricio Santillana, Katherine Ognyanova, Jon Green, James Druckman, David Lazer, and Matthew A. Baum. 2021. "Factors Associated With Self-Reported Symptoms of Depression Among Adults With and Without a Previous COVID-19 Diagnosis." *JAMA Network Open* 4(6):e2116612. doi: 10.1001/jamanetworkopen.2021.16612.
- Pew Research Center. 2014. "Religious Landscape Study." Retrieved (<https://www.pewforum.org/religious-landscape-study/>).
- Pew Research Center. 2020a. "Amid Protests, Majorities Across Racial and Ethnic Groups Express Support for the Black Lives Matter Movement." Retrieved (<https://www.pewresearch.org/social-trends/2020/06/12/amid-protests-majorities-across-racial-and-ethnic-groups-express-support-for-the-black-lives-matter-movement/#most-americans-say-theyve-had-conversations-about-race-or-racial-equality-in-the-last-month>).
- Pew Research Center. 2020b. "Most Americans Say Coronavirus Outbreak Has Impacted Their Lives." Retrieved (<https://www.pewresearch.org/social-trends/2020/03/30/most-americans-say-coronavirus-outbreak-has-impacted-their-lives/>).
- Pew Research Center. 2020c. "Worries About Coronavirus Surge, as Most Americans Expect a Recession – or Worse." Retrieved (<https://www.pewresearch.org/politics/2020/03/26/worries-about-coronavirus-surge-as-most-americans-expect-a-recession-or-worse/>).
- Saad, Lydia. 2020. "The U.S. Remained Center-Right, Ideologically, in 2019." Retrieved (<https://news.gallup.com/poll/275792/remained-center-right-ideologically-2019.aspx>).
- Tourangeau, Roger. 2014. "Defining Hard-to-Survey Populations." Pp. 3–20 in *Hard-to-Survey Populations*, edited by R. Tourangeau, B. Edwards, T. P. Johnson, K. M. Wolter, and N. Bates. Cambridge: Cambridge University Press.
- Valliant, Richard. 2020. "Comparing Alternatives for Estimation from Nonprobability Samples." *Journal of Survey Statistics and Methodology* 8(2):231–63. doi: 10.1093/jssam/smz003.
- Wingrove, Josh. 2021. "Vaccine Data Gaps Point to Millions More in U.S. Who Lack Shots." *Bloomberg*, December 18.
- Yang, Shu, Jae Kwang Kim, and Rui Song. 2020. "Doubly Robust Inference When Combining Probability and Non-Probability Samples with High Dimensional Data." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82(2):445–65. doi: 10.1111/rssb.12354.