*Journal of Survey Statistics and Methodology (2026) **00**, 1–24*

USING OPT-IN NON-PROBABILITY SURVEYS TO ESTIMATE COVID-19 INFECTION AND VACCINATION RATES

ALEXI QUINTANA-MATHÉ *
ATA A. USLU
JASON RADFORD
JAMES N. DRUCKMAN
KRISTIN LUNZ TRUJILLO
ALAUNA SAFARPOUR
KATHERINE OGNYANOVA
MATTHEW A. BAUM

ALEXI QUINTANA-MATHÉ is a PhD candidate, ATA A. USLU is a PhD student, JASON RADFORD is Principal Research Scientist, MAURICIO SANTILLANA is Professor, and DAVID LAZER is University Distinguished Professor at the Network Science Institute, Northeastern University, 177 Huntington Avenue, Boston, MA 02115, USA. JAMES N. DRUCKMAN is Martin Brewer Anderson Professor at Department of Political Science, University of Rochester, Rochester, NY 14627, USA. KRISTIN LUNZ TRUJILLO is Assistant Professor at Department of Political Science, Boston College, Boston, MA 02467, USA. ALAUNA SAFARPOUR is Assistant Professor at Department of Political Science, Gettysburg College, Gettysburg, PA 17325, USA. KATHERINE OGNYANOVA is Assistant Professor at School of Communication and Information, Rutgers University, New Brunswick, NJ 08901, USA. MATTHEW A. BAUM is Marvin Kalb Professor at John F. Kennedy School of Government and Department of Government, Harvard University, Cambridge, MA 02138, USA. JONATHAN SCHULMAN is a Research Associate at Pew Research Center, Washington, DC 20004, USA. ROY H. PERLIS is a Director at Center for Quantitative Health, Massachusetts General Hospital, Boston, MA 02114, USA; Professor at Department of Psychiatry, Harvard Medical School, Boston, MA 02115, USA. DAVID LAZER is University Distinguished Professor at Institute for Quantitative Social Science, Harvard University, Cambridge, MA 02138, USA.

Alexi Quintana-Mathé and Ata A. Uslu contributed equally to this work.

This research is based on work supported by the National Science Foundation (grants SES-2029292, SES-2029792, SES-2116189, SES-2241887, SES-2241885) and the Centers for Disease Control and Prevention Foundation (grant CDC-RFA-FT-23–0069). This research was partly supported by a grant from the Knight Foundation. We received generous support from the Russell Sage Foundation. The project was also supported by the Peter G. Peterson Foundation. Data collection was supported in part by Amazon. Our work was made possible through the continued financial and logistic support provided by Northeastern University, Harvard University, Rutgers University, and Northwestern University. Any opinions, findings, and conclusions or recommendations expressed here are those of the authors and do not necessarily reflect the views of the funders/sponsors.

Dr. Santillana has received institutional research funds from the Johnson and Johnson foundation, from Janssen global public health, and from Pfizer Pharmaceuticals, Inc. Dr. Perlis serves as a scientific advisor to Genomind, Vault Health, Psy Therapeutics, Circular Genomics, Swan AI Studios, Belle AI, and Mila Health.

*Address correspondence to Alexi Quintana-Mathé, Network Science Institute, Northeastern University,177 Huntington Avenue, 4th Floor, Boston, MA 02115, USA. E-mail: quintanamatha.a@northeastern.edu

https://doi.org/10.1093/jssam/smaf041

JONATHAN SCHULMAN
ROY H. PERLIS
MAURICIO SANTILLANA
DAVID LAZER

Public health crises demand timely surveillance of disease and interventions. This was a substantial challenge during COVID-19 in the United States due to the federalized response. States varied widely in infection and vaccination rates. Continuous state-level probability samples for tracking were unavailable and would have been logistically and financially onerous. Administrative data eventually became inaccurate. We thus evaluate the accuracy of large-scale opt-in non-probability state and national samples (from the Covid States Project) that estimated infection and vaccination rates on a near-continuous basis. The estimates aligned closely with a high-quality national probability sample and state-level administrative data (when such data were reliable). We offer evidence that the success of the surveys, compared to other less accurate non-probability surveys, may have stemmed from the successful recruitment of respondents with low trust in health institutions (i.e., individuals who often avoid surveys). We conclude by discussing how the Covid States Project can inform further data collection efforts requiring extensive spatial and temporal coverage.

KEY WORDS: COVID-19; Non-probability surveys; Online survey; Sampling; Survey methodology.

Statement of Significance

The COVID-19 pandemic posed major challenges for tracking infection and vaccination rates in the U.S. This was especially the case due to geographic variation, which necessitated unique state-level estimates. High-quality probability surveys with temporal and spatial reach were unavailable, and state administrative data eventually became unreliable. We evaluate the success of an alternative, the Covid States Project (CSP), which used opt-in non-probability samples across all states, over time. CSP estimates closely matched national probability surveys, and the administrative benchmarks when they were reliable. The CSP thus offers insights into how to collect geographically fine-grained quality data over time.

1. INTRODUCTION

The United States fared poorly through the COVID-19 pandemic (Nuzzo and Ledesma 2023). The roots of the U.S.'s performance are multifaceted, but one persistent challenge stemmed from its federalized public health infrastructure (Haffajee and Mello 2020). In essence, the country was experiencing 51 distinct pandemics—each state plus DC had unique experiences with distinct infrastructures (DeSalvo et al. 2021; Woolf 2022). Among other challenges, this created difficulties in surveillance that requires near continuous monitoring of infections and interventions (e.g., vaccinations). Surveillance data are crucial for facilitating optimal public health response. While administrative data are common for health surveillance (e.g., Virnig and McBean 2001), survey data also can serve that purpose (Marker et al. 2024).

Indeed, the survey research community provided critical data throughout the pandemic. The Societal Experts Action Network offered invaluable insights using various data types, most notably probability samples (National Academies of Sciences, Engineering, and Medicine 2020). Many of these surveys covered different time intervals and geographic units; however, financial and logistical realities meant an absence of multi-year, across-state surveys of COVID-19 infections and vaccination (two key surveillance outcomes). The Covid States Project (CSP) sought to fill this gap by using non-probability surveys in each state and the District of Columbia to track infections and vaccination, among other items. The project ran surveys quickly, inexpensively, and nearly continuously (which was necessary during a once-in-a-century crisis). A central question is whether the data provided accurate surveillance of infections and vaccination. If so, this approach may offer lessons for future efforts.

Yet, there is reason to be skeptical of non-probability surveys given the literature on their relative inaccuracy (Callegaro et al. 2014; Dutwin and Buskirk 2017; MacInnis et al. 2018; Cornesse et al. 2020; Lavrakas et al. 2022; Jerit and Barabas 2023; Mercer and Lau 2023; Rohr et al. 2024). More acutely, Bradley et al. (2021) documented inaccuracies in two large non-representative samples in their assessments of COVID-19 vaccination rates. The common use of quotas in non-probability samples invariably risks suffering from selection biases and violating the missing at random assumption (i.e., the source of missingness is unobservable; see Kawabata et al. 2024).

An alternative to non-probability samples is integrating probability samples at an aggregate level (e.g., national) with non-probability samples for smaller units (e.g., states) (Wiśniowski et al. 2020; Kim and Tam 2021; Salvatore 2023; Salvatore et al. 2024); with methods like small area estimation (Beaumont and Rao 2021; Rao 2021). However, these rely on either having statistically meaningful samples in the probability survey for the unit of interest or assuming some degree of homogeneity across units to "borrow strength" across areas (Molina and Rao 2023). During COVID-19, these conditions were rarely met since state-level probability samples were scarce for many

states, and states behaved heterogeneously in disease spread and vaccination coverage. Another alternative is using Multilevel Regression and Post-Stratification (MRP) (Downes and Carlin 2020; Ghitza and Gelman 2013) with national probability samples and state-level census data to post-stratify. However, MRP also assumes cross-states homogeneity and/or that demographics adequately characterize states regarding the outcomes of interest. In our case, these are problematic assumptions: as we show later, factors—including, say, institutional trust—are significantly associated with vaccination rates above and beyond demographics. The assumptions are alleviated with more data per state, implying that MRP is no substitute for decent-sized, quality state samples (but an alternative if these are not available).

Given these limitations of data integration approaches, we evaluate whether the CSP survey produced accurate national and state-level estimates. Non-probability sampling methods vary widely in accuracy (Baker et al. 2013), motivating our examination of their potential for over time surveillance across small spatial units. We compare CSP estimates to a high-quality probability survey and administrative sources, finding that the Covid States Project closely tracked both vaccination and cumulative infection rates, even at the state level. We also present evidence that part of this success may be due to the recruitment strategy of the CSP, which may have vitiated unit non-response/selection bias among those with lower institutional trust.

2. METHODS

2.1 CSP Survey

The Covid States Project was developed during the pandemic to provide state-level survey data over time in the U.S., recruiting through an opt-in non-probability platform. Most participants were recruited via PureSpectrum, an online polling firm aggregating over 20 panel providers across all 50 states and DC. PureSpectrum provides initial respondent identification, deduplication, and screening for quality, demographics, and location. Using PureSpectrum API, CSP launched and monitored simultaneous state-level surveys covering all 50 states and DC, using Census-based quotas, qualifications, and targeted sample sizes. PureSpectrum's providers (survey suppliers) display various types of surveys available to potential respondents and the CSP advertised itself as a "General" survey (rather than "Health") without referencing COVID-19 to avoid self-selection. No institutional affiliations were mentioned initially. Sponsor information appeared only in the body of the consent form after respondents clicked through.

CSP supplemented these respondents with a small number of respondents (4.7 percent) recruited via Facebook. While these participants were funneled into the same real-time survey projects, their recruitment messages disclosed

that the survey was conducted by major U.S. universities to study COVID-19. Each CSP recruitment data wave lasted approximately a month, with new waves launched roughly every six weeks. The data used here span June 12, 2020 to January 17, 2023 ($N = 324{,}536$). See Table S3 for fielding periods and sample sizes, and Table S7 for the state-level sample sizes by wave.

The CSP used quotas for race/ethnicity, age, and gender to approximate state-level demographics, with recruitment incentives adjusted in real-time to meet targets. In larger states (e.g., California, Texas, New York), quotas were interlocked by race, age, and gender. In smaller states (e.g., Wyoming, Delaware, and Rhode Island), quotas were used independently. National samples were reweighted using raking across interlocking race/ethnicity-gender-age subgroups, plus education, rurality, and region, to match Census benchmarks (or NCHS, for rurality). State-level analyses used separate weights matched to state-specific distributions on gender, race, age, education, and rurality. Weights were trimmed at the top and bottom 1 percent to limit outlier influence. This quota/weighting approach targeted variables that are linked to vaccination and infection outcomes. (Yasmin et al. 2021; Baghani et al. 2023).

The Covid States Project took several steps to address response validity (a particular concern in online surveys, Callegaro et al. 2014). First, respondents who failed a CAPTCHA test or one of two basic attention checks that straightforwardly ask respondents to choose a specific response option were filtered out (Berinsky et al. 2024). The remaining respondents were evaluated based on their performance on a series of quality checks, including: (i) duplicate responses, (ii) prevalence of item non-response, (iii) short survey completion time, (iv) straight-lining consecutive item responses, (v) selecting a large number of occupations, (vi) selecting a large number of religions, (vii) reporting a household with over 20 family members, (viii) reporting over 10 activities outside their home in the past 24 hours, (ix) identifying as a "Democrat" and "very conservative"; or a "Republican" and "very liberal", and (x) giving a meaningless answer to an open-ended question requesting a state name. Respondents who failed multiple quality checks were removed from the data. This more complex filtering affected a relatively small percentage of respondents (1–3 percent), while most of the completed responses that were removed (around 25 percent) were due to failing the basic attention checks.

Survey attrition was typically 15–20 percent, partly reflecting the survey's length (average median completion time: 23 minutes). Although filtering rates widely vary across online surveys in general (Daikeler et al. 2024), these are still non-trivial frequencies, and they highlight a downside of many opt-in non-probability surveys that invariably capture some invalid responses. Yet, our approach also underscores how to rigorously address this reality. COVID-19 infection questions appeared at the beginning in the survey, while

vaccination items were asked mid-survey. An example questionnaire (Wave 20, 11/03/21–12/02/21) is provided in SM.

2.2 Validation

We evaluate the accuracy of CSP infection and vaccination data using three validation strategies. First, we compare national estimates to those from the Axios-Ipsos Coronavirus Index (Jackson et al. 2022), a near-continuous national probability survey starting in March 2020. Axios-Ipsos uses Ipsos' KnowledgePanel, based on address-based probability sampling of U.S. adults. However, even after aggregating multiple waves, state-level samples were too small to be statistically meaningful; for example, waves 34 to 37 (∼6 weeks) yielded fewer than 40 respondents in 20 states. Thus, our comparisons are limited to national estimates during overlapping time periods.

Second, we compare CSP estimates to administrative data on infections and vaccinations on both national and state levels. These are often viewed as highly valid (Virnig and McBean 2001; Adebisi 2025), particularly when based on presumed census as was the case here (Marker et al. 2024). However, as we will detail shortly, COVID-19 administrative data became unreliable in specific periods. Therefore, we use these sources for state-level validation only during periods where we deem them accurate, that is, when they match the probability survey data nationally.

For infection rates, we relied on data compiled by *The New York Times* from state, county, and regional health departments. These became unreliable though after the February 2022 rollout of mass at-home rapid testing under the Biden administration, as positive results were rarely reported to health authorities that produced the administrative data (Donovan 2022; Rader et al. 2022; Santillana et al. 2024). Vaccination data came from the CDC but were also flawed: records often failed to link multiple doses to the same person, leading to inflated estimates (e.g., if person X received the two shots and a booster, the CDC often would count it as two or three people being vaccinated) (CDC 2023). Consequently, as more people became eligible for more shots, the rates exceeded even 100 percent in some cases (see Table S4).

Third, we evaluate the accuracy of the data by looking at its internal consistency for over time cumulative rates. Cumulative series should never decrease over time. To account for statistical uncertainty, we conclude the criterion is violated if an estimate is lower than the previous estimate in the series, and the 95 percent CI of both estimates do not overlap. This violation signals an inconsistency, as the rates of vaccinated people and cumulative infections can only increase over time. Other threshold criteria could be reasonably employed, and we therefore also generally examine the fluctuations of the estimates over time.

2.3 Measures

We calculate CSP vaccination rates using responses to: *"Have you or anyone you know received a COVID-19 vaccine?"*, computing the weighted share who selected *"I was vaccinated for COVID-19"*. Axios-Ipsos asks a similar multiple-choice item: *"Do you personally know anyone who has already received the COVID-19 vaccine?"*, and we use the weighted percentage selecting *"Yes, I have received the vaccine."* See Appendix A in the Supplementary Online Materials (SOM) for detail on our 95 percent confidence interval calculations. For the administrative data, we sum the number of first doses administered to adults (18+) in the CDC data and divide it by the national/state 18+ population from the 2020 Census. Since each CSP wave spans roughly one month (Table S3), we use the weighted median response date as the reference for vaccination estimates. Axios-Ipsos lacks precise response dates, so we use the midpoint of each wave's fielding period.

Definitions of a "COVID-19 infection" vary across states and over time in the administrative data and differ from survey-based definitions. In CSP surveys, infections are counted only when respondents report at least one positive test, whereas administrative data may include cases without confirmed tests under some conditions (see Appendix A in the SOM for details). For CSP data, we estimate cumulative infection rates using two methods. Method 1 divides the pandemic into multi-month periods and uses the most contemporaneous wave to estimate infections within each period to mitigate recall error. These period-specific infection rates are then accumulated. Method 2 calculates the cumulative rate directly, by counting the reported *total number of COVID-19 contractions throughout the pandemic*, separately in each wave. While Method 1 is generally more reliable and robust (due to minimizing recall bias and leveraging more responses), Method 2 is practical when longitudinal data are unavailable or under time and budget constraints. The Axios-Ipsos survey enables cumulative infection estimates in only six waves (November 19, 2021–June 13, 2022). We describe the calculation procedures for both CSP and Axios-Ipsos, along with the relevant survey questions, in Appendix A in the SOM.

3. RESULTS

We present results for vaccination rates, followed by infections. A natural question is: how closely must CSP estimates track the other data to be considered "accurate"? The answer depends on purpose and concomitant tolerable tradeoffs (e.g., what alternatives exist if CSP data were unavailable?). We do not offer a formal answer but present deviations and our own assessment of their size; readers may view deviations as more or less concerning.

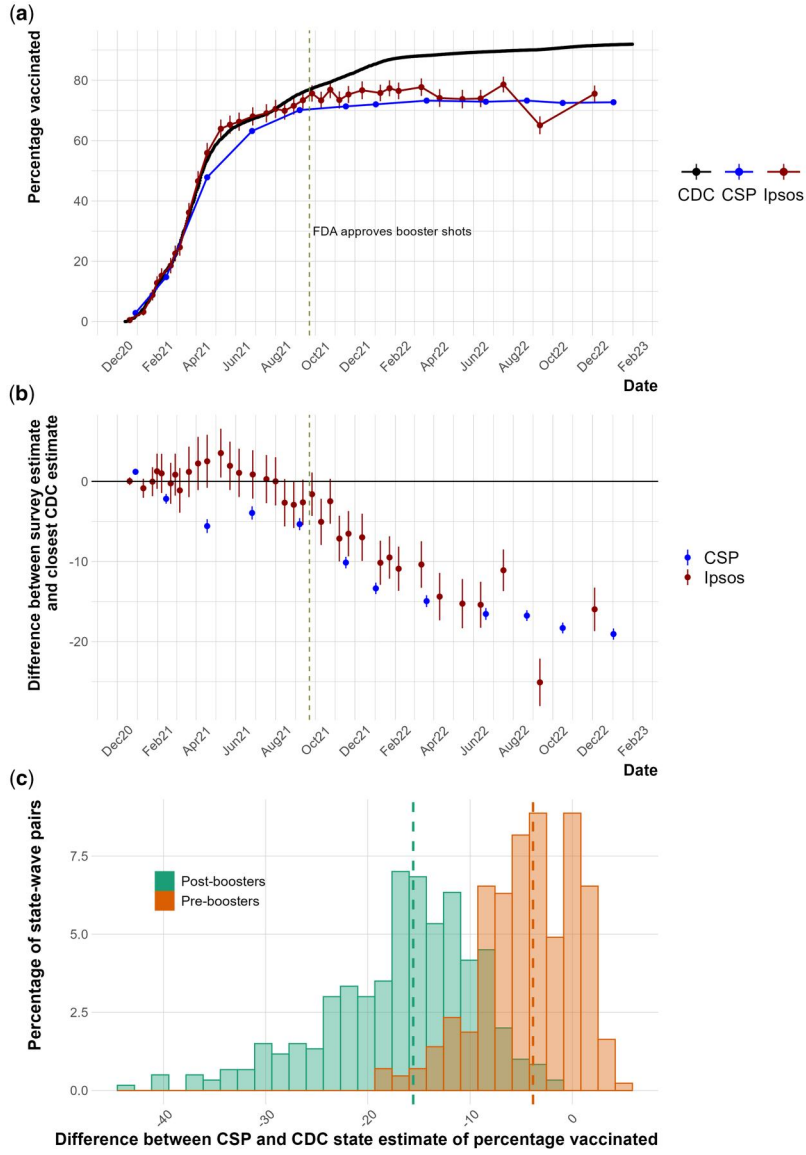

Figure 1. (a) Trends in vaccine uptake estimations, comparing the CDC, the Axios-Ipsos Coronavirus Index, and the CSP. Error bars represent 95 percent CI (see Appendix A in the SOM for details) (b) Difference between Axios-Ipsos/CSP estimates and the CDC estimate of vaccine uptake over time. (c) Histogram of differences between CSP and CDC state estimates across all CSP waves, split into waves before and after booster shots were approved. Vertical dashed lines correspond to medians.

3.1 Vaccinations

Figure 1a and b compare national vaccination rate estimates from CSP, Axios-Ipsos, and CDC data. The first takeaway is that, as discussed, CDC estimates become unreliable after the rollout of booster shots: by December 2021, CDC rates clearly exceed both surveys. Post-booster, CDC deviates from Axios-Ipsos by an average of 10 points and CSP by 16 points, likely due to their failure to link multiple doses to individuals, leading to inflated estimates.

Before boosters, CDC rates deviate from CSP on average by 3.6 points (across 5 waves) and from Axios-Ipsos by 1.4 points (across 20 waves). The CDC estimate is within 95 percent error margins of 18 of the 20 Axios-Ipsos estimates and of none of the 5 CSP estimates; however, this is largely because of very narrow CSP confidence intervals due to large sample sizes. Regardless, the CSP estimates were not substantively far off from the CDC during this period. Moreover, CSP data show that 65 percent of the vaccinated respondents had received a second shot in the April 2021 wave, and above 80 percent in subsequent waves, suggesting that CDC may have overestimated rates as early as spring 2021. Another difference between the CDC and the surveys is the trends after January 2022: CDC estimates continue rising slowly, while both surveys plateau with means (in this period) of 72.8 percent (SD = 0.5) for the CSP and 74.8 percent (SD = 3.8) for Axios-Ipsos.

CSP and Axios-Ipsos vaccination estimates generally align: 9 of 12 CSP estimates deviate 5 percent or less from the closest Axios-Ipsos estimate, and 4 CSP confidence intervals intersect with those of the closest Axios-Ipsos estimate (see Table S8 for differences and margins of error). Two of the three >5 percent deviations stem from a clear Axios-Ipsos outlier in September 2022: the only instance where CSP overshoots. That Axios-Ipsos estimate violates the internal consistency validity criteria insofar as it shows a statistically significant decrease in a cumulative variable. CSP never violates this criterion. The second largest deviation (8.1 percentage points) is in April 2021, when Axios-Ipsos overshoots CDC by 2.5 percent, likely an overestimation similar to the CDC, as noted. Excluding these three likely-off points, all CSP Axios-Ipsos differences remain within 5 percent throughout the whole period. We interpret this as evidence of strong alignment/validation.

Based on national level comparisons, we assume that the CDC data are valid before boosters (despite some overcounting, as discussed), and split the state-level comparisons by the date of booster approval. Figure 1c depicts the distribution of state-level differences between CSP and CDC, before and after the booster approval. Deviations increase substantially post-booster: only 61 of 357 state-wave deviations (17 percent) are below 10 percent in this period. In contrast, 155 of 255 state-wave comparisons (61 percent) pre-boosters are within 5 percent. In Figure S4, we plot state-level CSP and CDC trends; and Figure S5 shows the state-level CSP-CDC differences for the five waves before booster approval. Survey estimates match official data fairly well

during this period: 58 percent of 255 state-wave deviations fall within confidence intervals, with mean and median absolute differences of 4.9 percent and 3.9 percent. At least half of the 51 state deviations fall within the margins of error each wave, and 36 states have 3 or more of 5 deviations within 95 percent error margins. However, some states show bias: DC, SD, NM and VT undershoot the CDC by 10 percent or more in 3 of 5 estimates, and OK and SD have zero deviations within confidence intervals. In addition, 6 states plus DC have only 1 deviation within CIs (AK, DC, NE, NM, PA, VT, WV).

Unsurprisingly, states with smaller samples show greater deviations from the CDC: the 19 states plus DC, which have at least 3 out of 5 differences above 5 percent have on average 428 respondents per wave, while the other 31 states have 494 (see Table S7). CSP state-level estimates generally undershoot the CDC values as all differences above 95 percent error margin after February 2021 are negative, mirroring the national-level undershooting during this period. The internal consistency validation criterion is only violated twice (out of 561 comparisons) at the state level, by the SD estimate of August 2022 (-20 percent difference from the previous estimate), and by the NY estimate of October 2022 (−7 percent difference). Only 22 estimates are more than 5 percent lower than the previous estimate. Vaccination rates reach a plateau after the booster rollout (Figure S4); the standard deviations of the CSP estimates in this period for every state are above 5 only for SD and ND, and the mean standard deviation is 2.5 percent. Overall, this indicates that the CSP state estimates are robust over time.

We conclude that the CSP survey is able to provide reliable estimates of vaccination rates for most U.S. states and is certainly superior to CDC estimates once boosters were available (which, if not capped at 95 percent, exceeded 100 percent in 10 states in January 2023, see Table S4). The comparison to CDC data during the pre-booster period allows for the identification of problematic states (say DC, ND, NM, OK, SD, or VT).

3.2 Infections

Figures 2a and b compare national infection estimates from the CSP (Methods 1 and 2), Axios-Ipsos, and *The New York Times*. As with vaccination data, we first assess when administrative data provide reliable benchmarks before turning to state-level comparisons. Methods 1 and 2 of CSP yield highly similar results: the largest national-level difference between each Method 1 and its closest Method 2 estimate is 3.8 percentage points. Method 1 aligns especially well with official data before 2022, with differences being less than 1 percentage point in all periods except three (September 20: 1 percent; October 20: 1.18 percent; November 20: 1.52 percent). CSP estimates slightly overshoot the official numbers in 2020, a year in which institutionalized testing capabilities were inconsistent (Cohen 2020; Patel 2020). All the official numbers in

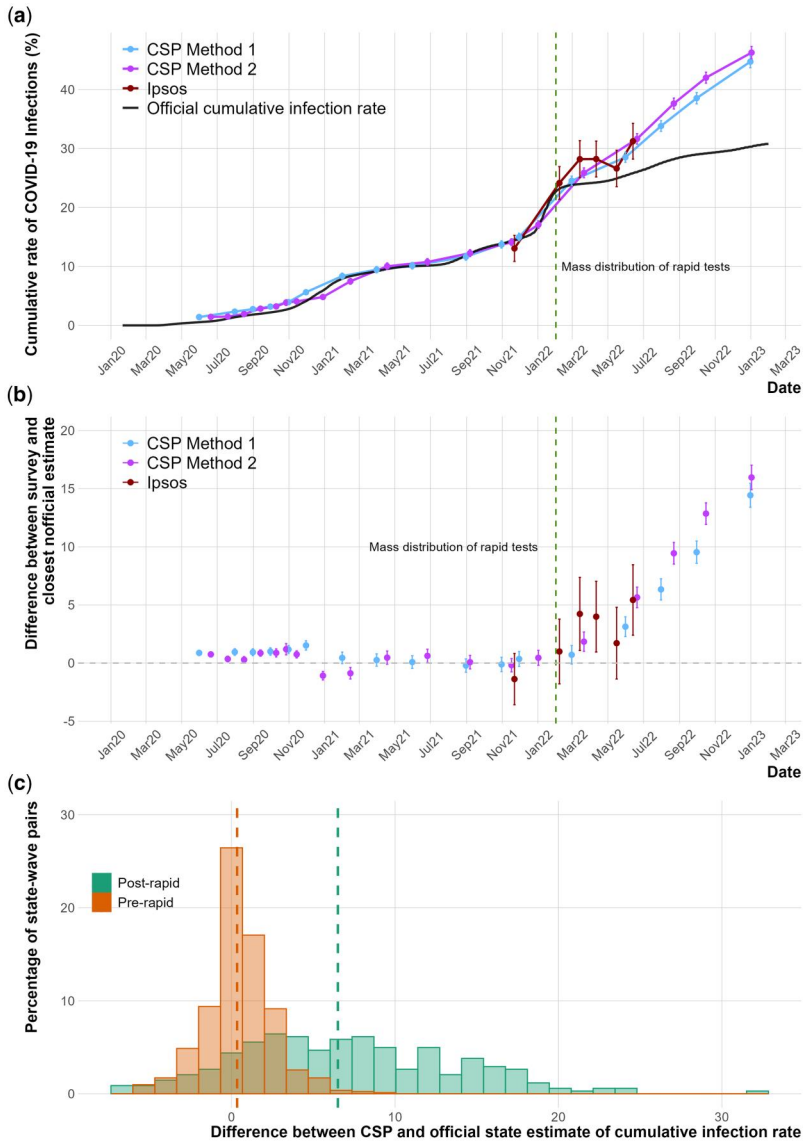

Figure 2. (a) Cumulative COVID-19 infection rates, comparing the two estimation methods of CSP, to Axios-Ipsos, and the official rate. (b) Difference between Axios-Ipsos/CSP estimates and the official number of cumulative COVID-19 infection rates over time. (c) Histogram of differences between CSP estimates and official state numbers across all CSP waves using Method 1, split into waves before and after the mass distribution of at-home rapid tests. Dashed lines show medians. Error bars represent 95 percent CI (see Appendix A in the SOM for details).

2021 are within the 95 percent confidence interval of CSP using Method 1. Method 2 estimates also closely follow the official numbers during 2020 and 2021. In 2020, the differences are less than 1 percentage point, except in September, October, and November 2020. In 2021, half of the official numbers are within the 95 percent confidence intervals of Method 2 (which, as mentioned, we expected to be less reliable).

The estimate from the only Axios-Ipsos survey that was fielded during the first 2 years of the COVID-19 pandemic (December 2021) and asked the number of times respondents were infected aligns with CSP estimates from both methods, as well as the official data. While only the 6 Axios-Ipsos waves shown in Figure 2 included questions allowing a calculation of cumulative infection rates (i.e., asking how many times a respondent had been infected), most of their other survey waves did ask whether respondents tested positive at least once. Therefore, we can use Axios-Ipsos as a comparison through a longer period for an estimand (the percentage of the population that tested positive at least once), closely related to the estimand of interest (the cumulative infection rate). Appendix A in the SOM details this analysis from June 2020 to June 2022 and shows the results in Figure S2. CSP estimates are slightly above Axios-Ipsos values, but they are generally close, with 9 out of 16 CSP estimates within confidence intervals of the most contemporaneous Axios-Ipsos counterpart.

From February 2022 onwards, infection rates from the administrative data progressively diverged from the CSP and Axios-Ipsos surveys—as expected due to the aforementioned distribution of at-home rapid tests. In 2022, only 1 of the estimates from Method 1 (February 28, 2022) captures the official number of infections within its confidence interval, and none from Method 2, with both methods estimating substantially more COVID-19 cases than the administrative data. Similarly, Axios-Ipsos estimates show significantly more infections than the official data in this period. Although they contain both the most contemporaneous CSP and the official numbers in their confidence intervals in 3 of the 6 waves, all Axios-Ipsos estimates are closer to the CSP values than the official data. The results from Axios-Ipsos indicate even more cases than the CSP in 2/2022, 3/2022, and 4/2022. After February 2022, only 1 out of 4 Axios-Ipsos confidence intervals contains the administrative rate, as the gap between survey estimates and the official data grows, pointing to an underreporting of COVID-19 infections in the official numbers in this period. Both the CSP Method 2 and the Axios-Ipsos estimates satisfy the internal consistency criteria. (CSP Method 1 satisfies it by its accumulating design.) In fact, each Method 2 estimate is higher than the previous estimate.

Given these issues in the administrative data, we split the state-level comparisons by the start date of at-home rapid test distribution. Figure 2c displays histograms of the differences between the CSP estimates and the administrative value at the median date of the survey responses in each state, before and after the distribution of rapid tests. Figures S6 and S7 display state trends and

deviations from the official estimates prior to the rapid test distribution, respectively. Using Method 1, all CSP state-level cumulative infection rates from 2020 and 2021, 612 estimates from 12 survey waves in total, are within 10 percent of the official number of cases, with a median and mean difference of 0.34 percent and 0.46 percent, respectively. In 43 states and the DC, all estimates from these 12 CSP waves are within 5 percent of the official numbers, with the exceptions being AK, DE, NC, ND, RI, SC, TX, and WY. Apart from NC, SC, and TX, these again are all low population states with smaller CSP samples (see Table S7 for state sample sizes). At least 11 out of 12 estimates are capturing the official numbers within 95 percent CI in 36 states; and in all states except TX and DE, a majority of the official estimates are within 95 percent CI. These results point to generally little bias in the CSP estimates except for a few states.

Inconsistencies in how states report cases may explain some of the larger differences observed (see Appendix A in the SOM for details). For example, positive results in an antigen test were not counted as an infection in Texas, and thus differences with the CSP may reflect uncommon, aberrational administrative accounting (Platoff et al. 2020). It is worth noting that the states with more biased CSP estimates of cumulative infection rate before rapid tests are not the same as the states with biased estimates of vaccination coverage before the booster rollout. In particular, none of the 9 states (TX, DE, NC, SC, WI, CA, MS, OH, RI) with 7 or fewer cumulative infection rate estimates within 95 percent CI of the administrative data (out of 12 estimates) is also one of the 9 states (OK, SD, NE, NM, AR, DC, PA, VT, WV) with only 0 or 1 vaccination estimates within 95 percent CI of the CDC estimate (out of 5 estimates). After the mass distribution of at-home rapid tests, CSP's state-level estimates significantly (and predictably) overshoot the official data, mimicking the results at the national level: the median and mean differences are 7.1 and 6.5, respectively, and 87 percent of the state-wave CSP estimates (out of 255 state-wave pairs) are larger than the official value.

While we focus on Method 1 in the previous paragraph, Method 2 provides estimates that match the administrative data well at the state level prior to the rapid test campaign. The mean absolute difference between Method 2 estimates and the official data across 663 state-wave estimates before rapid tests is 1.57 percent. For Method 1 it is 1.52 percent (612 estimates). The Method 1 estimates are within CI 86 percent of the time, more than the 82 percent for Method 2. As expected, Method 1 is slightly more accurate, and also more consistent, as this method generates estimates for the same state that are correlated and cannot violate the internal inconsistency criteria (Figure S7). However, this can also let inaccuracies in a single wave extend to estimates further on. For example, the overestimation of the official numbers in the June–July 2020 period in Texas, early in the pandemic, causes 9 out of 12 estimates to have a CI outside of the administrative value for this state. In contrast, the CI of only 5 out of 13 estimates from Method 2 does not cover the official estimate in Texas. Also, only one estimate from Method 2 (out of 918

comparisons), from August 2020 for Rhode Island, violates the internal inconsistency criterion (with a difference of −1.76 with the previous estimate), and 75 percent of these estimates are higher than the previous one in the temporal series. Overall, our results show that a single survey wave from CSP can be used with Method 2 to provide generally accurate state-level estimates of cumulative infection rates, which would be particularly useful when time or budget limits prevent multiple large-scale waves.

An advantage of looking at both the administrative data and the CSP's geographically granular data is it enables the computation of data-driven confidence intervals for the CSP estimates (a desirable feature given that nonprobability samples do not technically satisfy the assumptions underlying confidence interval computation). We use the state-level deviations from the administrative data before the distribution of rapid tests, where we deem it accurate, to derive confidence intervals for the CSP state estimates through the whole period considered. The resulting 95 percent confidence interval is (−4 percent, 5 percent) (see Appendix A in the SOM for details), slightly wider than the average of the conventional 95 percent error margins across the CSP state estimates in the pre-2022 data, of 3 percent. This is worthwhile since it allows for the computation of uncertainty without relying on assumptions about the data generating process (a probability sample), and accounts for bias unaccounted for in standard metrics.

3.3 Evaluating Reasons for CSP's Accuracy

In the introduction, we briefly mentioned Bradley et al.'s (2021) study of survey surveillance of COVID-19 vaccination from January 2021 to May 2021. The authors find that two large-scale surveys, the Delphi-Facebook (Salomon et al. 2021) and Census Household Pulse (Fields et al. 2020), with poor coverage, were inaccurate relative to the CDC data and Axios-Ipsos probability surveys. This raises the question of why the CSP performed well whereas the two surveys in Bradley et al. (2021) did not. As with most evaluations of survey accuracy, we cannot arrive at a definitive answer due to too many factors being in play (Clinton et al. 2021). First, it could be that the respondent quality checks included in the CSP address common sources of error in other online opt-in surveys (Freese and Jin 2025). Second, it could be the specific quotas matter given they encompass key variables related to vaccination, for example, age. Indeed, the quotas minimized the relevance of weights: the CSP unweighted estimates are similarly accurate as the weighted estimates, nationally and at the state level (see Figures S8 and S9). Some of the variables on which the unweighted sample strongly skewed, like gender, were generally unrelated to vaccination and infection (Table S6), making the CSP data "fit for purpose" for these outcomes (Cornesse et al. 2020).

Third, we suspect a partial explanation lies in unit non-response not captured by standard demographic measures, in particular, trust in institutions. Low-trust individuals may decline participation as they view it as "a political act" (Clinton et al. 2021; Silber et al. 2022). Those with lower trust in scientific and health institutions were significantly less likely to vaccinate (e.g., Szilagyi et al. 2021), a finding confirmed in the CSP data (see Appendix A in the SOM). Thus, if a surveillance survey obtains a lower response rate from those with lower trust, then it would overstate vaccination rates.

We start by documenting this point with simulations. The CSP surveys asked respondents about their trust in the CDC (on a 4-point scale, see Appendix A in the SOM for details). We downsample the CSP data with probabilities that depend on the answers to this question and then calculate the vaccination and the cumulative infection rate (using Method 1) with the downsampled datasets. The goal is to simulate what the estimates would be if the sample had fewer low-trusting people. We use four different downsampling schemas: Table 1 displays the probability that respondents are included in the downsampled dataset in each schema depending on their answers to the CDC trust question.

These four downsampling processes are applied a hundred times to each survey wave, before reweighting (this is key since it ensures that the biases we simulate are not accounted for by standard weighting variables) and calculating vaccination and cumulative infection rates. We then calculate the difference between the averaged outcome across the hundred simulations and the value from the original dataset, for each survey wave, and plot them in Figure 3. To provide a measure of uncertainty in the differences, we sum the MOE of the original estimate to the MOE averaged across the 100 simulations. The results make clear that if the samples had been skewed toward higher-trusting respondents (i.e., had greater unit non-response among low-trusters) vaccination rate estimates would have been significantly higher/overestimated, even after applying our weighting scheme.

Table 1. Probability of Sampling in Simulations by Level of Trust in the CDC in Columns 2 to 5. Column 6 displays the fraction of the original dataset preserved for each downsampling probability schema on average across simulations and survey waves

Downsampling probability schema	Answer to trust in the CDC question				Average fraction of original sample preserved
	A lot	Some	Not too much	Not at all	
Low	1	0.9	0.8	0.7	0.90
Mild	1	0.9	0.6	0.5	0.85
High	1	0.9	0.4	0.3	0.80
Very high	1	1	0.2	0.2	0.80

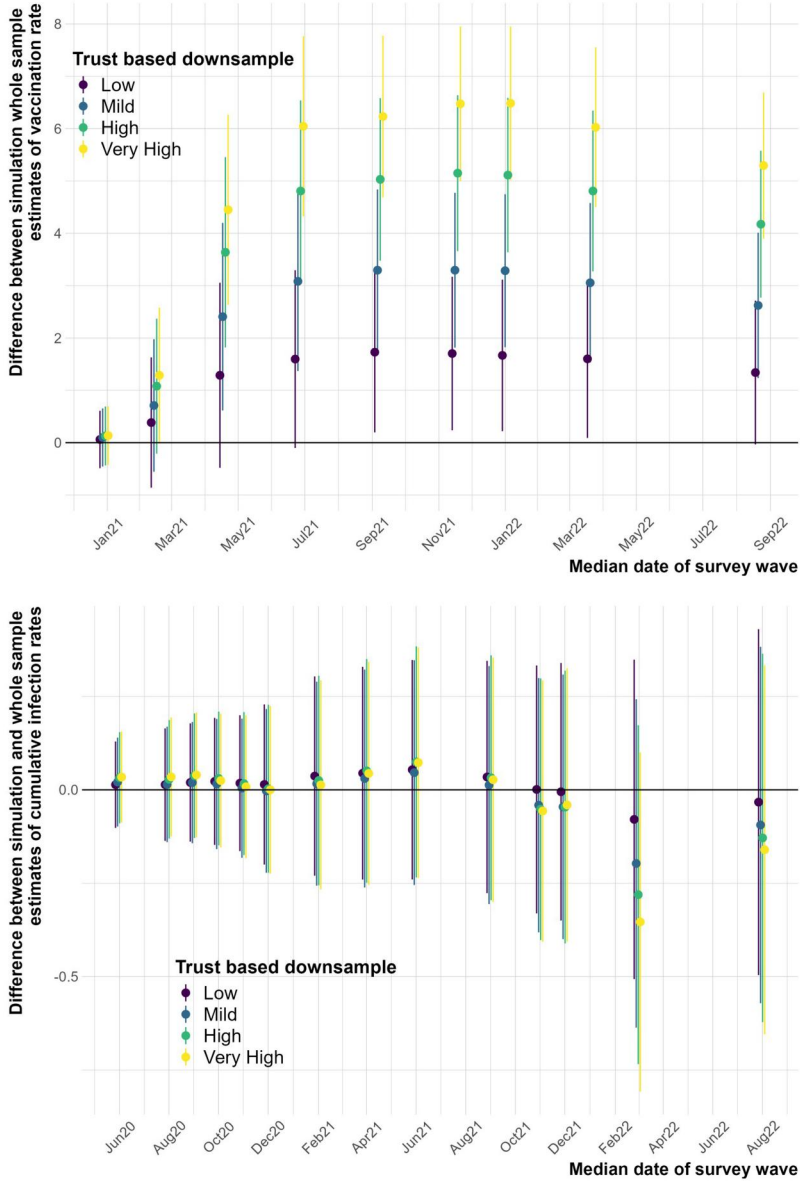

Figure 3. Results of Simulations Downsampling CSP Waves Following Trust in CDC Levels. The *Y*-axis displays the difference between vaccination rate (upper panel) and infection rate (bottom panel) after downsampling, compared to the original whole sample estimate. Error bars represent the sum of the average 95 percent MOE of the simulations and the whole sample MOE.

In contrast, the trust down-sampling has almost no influence on the cumulative infection rates. This is sensible insofar as trust constituted a crucial correlate of vaccine uptake (Adhikari et al. 2022), but *not* infection rates that depended more on sociodemographic and contextual factors (Karmakar et al. 2021). Indeed, in the CSP data, trust in scientists and in the CDC to handle the COVID-19 pandemic strongly correlates with vaccination rates but not infection rates (Figure S3). Notably, Bradley et al. (2021) evaluate the large unrepresentative samples regarding vaccination, but not infection. This simulation provides some evidence that unit non-response based on trust can skew vaccination estimates, a finding that aligns with speculation that other polls have yielded inaccurate estimates because of difficulty to calibrate unit non-response (Clinton et al. 2021).

We suspect the CSP may have done relatively well recruiting low-trusting respondents due to the recruitment strategy. As mentioned, recruitment for the bulk of CSP respondents (via PureSpectrum) made no mention of a sponsoring institution (e.g., a university) or a survey topic (e.g., health, COVID-19). This may have vitiated non-response by those with low trust in scientific and health institutions who otherwise may have been wary of responding to such inquiries (see Appendix A in the SOM for a discussion on how this compares to the surveys in Bradley et al. 2021). We explore this hypothesis by leveraging the aforementioned fact that a small number of CSP respondents were recruited differently (not through PureSpectrum), via Facebook. Unlike PureSpectrum recruitment, the Facebook recruitment made the sponsoring institutions and topic clear. We compare vaccination rates between those who were recruited via PureSpectrum and Facebook. The surveys and timing were identical; the only differences were in the recruitment message and source. We focus strictly on vaccination given the prior results that infection rates are unaffected by trust variation.

We plot, in Figure 4, the vaccination rate in each sample, compared to the CDC and Axios-Ipsos estimates. The vaccination rate in the Facebook sample is 15 percentage points higher than in the PureSpectrum sample on average across 11 survey waves. In Appendix A in the SOM, we further probe our hypotheses by showing that the Facebook sample exhibited significantly more trust in science relative to the PureSpectrum sample, and that trust may have played a role in mediating the impact of the recruitment source on vaccination rates. While we are not able to disentangle the role of the source (Facebook) from the role of the recruitment message in this experiment, consuming COVID-19 information on Facebook is associated with vaccine resistance (Green et al. 2023), pointing to the importance of the recruitment message.

While speculative, the results highlight a potential role of recruitment methods in affecting unit non-response (e.g., Groves et al. 2012) in an era where trust in knowledge-generating institutions has declined and polarized (Brady and Kent 2022; Schulman et al. 2025). While some work suggests no or modest survey sponsorship effects (Tourangeau et al. 2014; Crabtree et al. 2022),

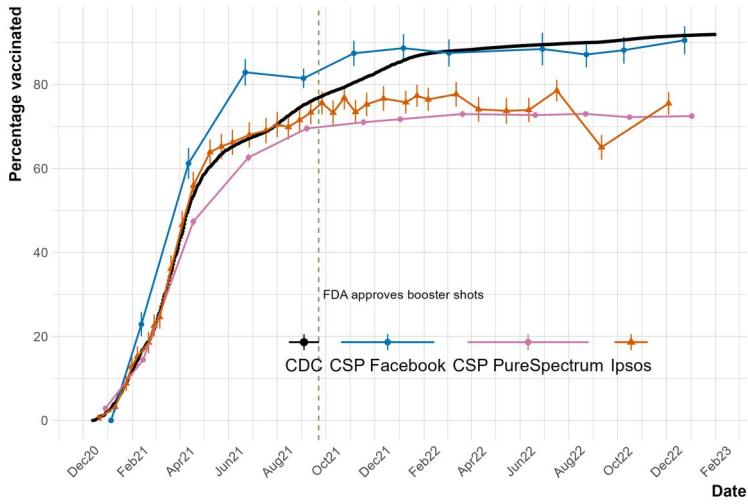

Figure 4. Vaccination rate in the subsamples of the CSP respondents recruited through Facebook, compared to those recruited through PureSpectrum, Axios-Ipsos, and the CDC data.

it is important to explicitly explore the role of trust (Silber et al. 2022) and consider its overtime relationship with various institutions.

4. DISCUSSION

What lessons do the Covid States Project provide? First, given the relatively poor performance of opt-in non-probability samples (MacInnis et al. 2018; Cornesse et al. 2020) in estimating descriptive population parameters, one cannot presume accuracy (none of our results question the, all else constant, superiority of probability samples). We presented our comparisons between the CSP, Axios-Ipsos, and the administrative data as a validation exercise. In real time, though, one could use the latter sources as "guardrails"—that is, quality benchmarks against which the non-probability estimates are regularly compared. These typically will be assessed at either a less granular spatial level and/or more limited period (as if not, then there are fewer reasons to use the non-probability data in the first place). Non-probability surveys should transparently report deviations from other sources. Documenting consistent accuracy as well as internal consistency (e.g., cumulative rates should not decline) can be thought of as a necessary condition for using opt-in non-probability data for surveillance. Along these lines, we emphasize that our results should not be taken as a "general" validation of the inherent challenges with non-probability samples (e.g., selection biases, missing not at random). However,

our work delineates an approach, based on regular checks with guardrail benchmarks, that can be applied to other non-probability online surveys.

A second lesson is that multiple validation metrics (or guardrails) can be essential. In our case, a probability survey provided a reliable benchmark at the national level, while the administrative data provided a reliable benchmark at the state level only across some periods of time. Having a state-level benchmark allowed us to confirm that the CSP state estimates were valid, facilitated identifying problematic states (often ones with smaller sample sizes) and allowed the calculation of data-driven confidence intervals that account for both statistical uncertainty and bias. Using two guardrails allowed the detection of periods where one of the original guardrails (administrative data) was unreliable. This is consistent with public health practices triangulating as many seemingly reliable signals as possible (Smolinski et al. 2015; Baltrusaitis et al. 2018; Lipsitch and Santillana 2019; Lu et al. 2021).

Third, researchers need to be deliberate when assessing or using administrative data, and refrain from characterizing them as a "ground truth" benchmark especially when third party sources (e.g., probability surveys) disagree with the official data. The administrative biases mean that for state-level estimates, the CSP provided the most reliable data on vaccination and cumulative infection rates of which we are aware across long periods of the COVID-19 pandemic. Bluntly, careful use of an opt-in non-probability data source offered otherwise unavailable information vital to surveilling public health.

Fourth, the CSP offers guidance on survey design. The CSP worked well due to various design decisions; among these, we suspect that ensuring response validity was key and the CSP employed extensive filtering. In addition, the samples were well balanced on the key correlates of the outcomes (infections and vaccination) thanks to the use of quotas that reduced the reliance on weighting. CSP's use of PureSpectrum platform, which aggregates diverse survey providers, may have alleviated the biases from these specific channels. Given the inevitable threat of unmeasured confounding in non-probability samples, ensuring balance on correlates of the target parameter is vital to accurate estimation; it also cannot be assumed and requires an understanding of the targeted outcome and validation of the approach.

In particular, trust in science has polarized more than any other institution in the United States (Schulman et al. 2025). Nonresponse among low-trust individuals may lead to inaccuracies in survey estimates. We recommend avoiding sponsor or content disclosure during recruitment to prevent this effect, or weighting by correcting variables if present (e.g., partisanship, see Appendix A in the SOM). That is the case, of course, if the target estimand correlates with trust, as with vaccination (but not infection).

All of this suggests that opt-in non-probability samples can play a vital role in situations that demand longitudinal, across space data. Examples of such scenarios include reactions to climate change, mass protests, political campaigns, gun purchasing, mental health, racism, and more. This can be done

without substantial financial cost. For instance, a 25,000-person CSP sample, with representative state-level coverage, costs around $40,000. While we identify potential biases for some states, our results indicate that a large portion of the inaccuracies of the CSP state estimates were due to sampling error, especially regarding cumulative infection rates. Therefore, larger samples would likely yield substantially more accuracy for a marginal increase in costs (Jerit and Barabas 2023). Probability surveys may produce more accurate data in general, but there are often logistical and financial tradeoffs. Finally, the CSP large samples provide an opportunity to look at small subpopulations of interest (Simonson et al. 2024). We did not do this here, but the samples are large enough, at least nationally, to zero in on very small groups.

More generally, we recognize our evaluation of the Covid States Project is limited. We looked at limited outcomes, employed statistical methods developed for probability samples, and did not offer a precise metric for suitable accuracy matches. We also did not explore the use of MRP with our data which may have helped in our poorer performing smaller sample states (i.e., not as a full substitute to the CSP but as a complement for these states). Studying how MRP can complement non-probability data collection for small geographical units is an interesting avenue for future work. Additionally, the CSP's approach to data collection was far from straightforward and required a host of decisions, many of which were made quickly to provide data as the COVID-19 pandemic unfolded. The field would benefit from building on task force reports on non-probability sampling and online sampling (e.g., McPhee et al. 2022) to further develop infrastructure that would be available for researchers to use when probability samples are unable to meet the demands and when fit for purpose.

SUPPLEMENTARY MATERIALS

Supplementary materials are available online at academic.oup.com/jssam.

DATA AVAILABILITY

The Axios-Ipsos and the CDC data used in this paper are available online: the CDC data can be downloaded from the CDC website at this link[1], and the Axios-Ipsos data is available at https://ropercenter.cornell.edu/. The individual level CSP data is accessible at the Harvard Dataverse (https://doi.org/10.7910/DVN/ZUJU6H) upon publication of the article (or if requested by reviewers). To avoid deidentification, only the variables used for the main text analysis

1. https://data.cdc.gov/Vaccinations/COVID-19-Vaccination-Age-and-Sex-Trends-in-the-Uni/5i5k-6cmh/

will be included, without demographics. In particular, the questions on vaccination, testing, and COVID-19 infections will be included, together with the state of the respondent and the weights used. The data with all variables used for weighting (enabling a full replication of the results, including the weights calculation) will be provided upon request to the authors after going through a data sharing agreement. All code used in this paper is available at https://github.com/LazerLab/COVID19-surveys.

REFERENCES

Adebisi, Y. A. (2025), "Utilizing Administrative Data to Inform Health Promotion, Policy, and Practice," in *Handbook of Concepts in Health, Health Behavior and Environmental Health*, Singapore: Springer, pp. 1–24. https://doi.org/10.1007/978-981-97-0821-5_134-1.

Adhikari, B., Yeong Cheah, P., and von Seidlein, L. (2022), "Trust is the Common Denominator for COVID-19 Vaccine Acceptance: A Literature Review," *Vaccine: X*, 12, 100213. https://doi.org/10.1016/j.jvacx.2022.100213.

Baghani, M., Fathalizade, F., Loghman, A. H., Samieefar, N., Ghobadinezhad, F., Rashedi, R., Baghsheikhi, H., Sodeifian, F., Rahimzadegan, M., and Akhlaghdoust, M. (2023), "COVID-19 Vaccine Hesitancy Worldwide and Its Associated Factors: A Systematic Review and Meta-Analysis," *Science in One Health*, 2, 100048. https://doi.org/10.1016/j.soh.2023.100048.

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., Gile, K. J., and Tourangeau, R. (2013), "Summary Report of the AAPOR Task Force on Non-Probability Sampling," *Journal of Survey Statistics and Methodology*, 1, 90–143. https://doi.org/10.1093/jssam/smt008.

Baltrusaitis, K., Brownstein, J. S., Scarpino, S. V., Bakota, E., Crawley, A. W., Conidi, G., Gunn, J., Gray, J., Zink, A., and Santillana, M. (2018), "Comparison of Crowd-Sourced, Electronic Health Records Based, and Traditional Health-Care Based Influenza-Tracking Systems at Multiple Spatial Resolutions in the United States of America," *BMC Infectious Diseases*, 18, 403. https://doi.org/10.1186/s12879-018-3322-3.

Beaumont, J. F., and Rao, J. N. K. (2021), "Pitfalls of Making Inferences from Non-Probability Samples: Can Sata Integration Through Probability Samples Provide Remedies," *The Survey Statistician*, 83, 11–22.

Berinsky, A. J., Frydman, A., Margolis, M. F., Sances, M. W., and Valerio, D. C. (2024), "Measuring Attentiveness in Self-Administered Surveys," *Public Opinion Quarterly, Oxford University Press (OUP)*, 88, 214–241. https://doi.org/10.1093/poq/nfae004.

Bradley, V. C., Kuriwaki, S., Isakov, M., Sejdinovic, D., Meng, X.-L., and Flaxman, S. (2021), "Unrepresentative Big Surveys Significantly Overestimated US Vaccine Uptake," *Nature, Nature Publishing Group*, 600, 695–700. https://doi.org/10.1038/s41586-021-04198-4.

Brady, H. E., and Kent, T. B. (2022), "Fifty Years of Declining Confidence & Increasing Polarization in Trust in American Institutions," *Daedalus*, 151, 43–66. https://doi.org/10.1162/daed_a_01943.

Callegaro, M., Villar, A., Yeager, D., and Krosnick, J. A. (2014), "A Critical Review of Studies Investigating the Quality of Data Obtained with Online Panels Based on Probability and Nonprobability samples," in *Online Panel Research*, eds. M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, and P. J. Lavrakas, John Wiley & Sons, Ltd, pp. 23–53. https://doi.org/10.1002/9781118763520.ch2.

CDC (2023), "Data Definitions for COVID-19 Vaccinations in the United States," Available at https://archive.cdc.gov/www_cdc_gov/coronavirus/2019-ncov/vaccines/reporting-vaccinations.html.

Clinton, J., Agiesta, J., Brenan, M., Burge, C., Connelly, M., Edwards-Levy, A., Fraga, B., Guskin, E., Hillygus, D., Jackson, C., Jones, J., Keeter, S., Khanna, K., Lapinski, J., Saad, L., Shaw, D., Smith, A., Wilson, D., and Wlezien, C. (2021), *Task Force on 2020 Pre-Election*

Polling: An Evaluation of the 2020 General Election Polls, Alexandria, VA, USA: American Association for Public Opinion Research (AAPOR).
Cohen, J. (2020), "The United States Badly Bungled Coronavirus Testing—But Things May Soon Improve," *Science*, 367, 1282–1283. https://doi.org/10.1126/science.abb5152.
Cornesse, C., Blom, A. G., Dutwin, D., Krosnick, J. A., De Leeuw, E. D., Legleye, S., Pasek, J., Pennay, D., Phillips, B., Sakshaug, J. W., Struminskaya, B., and Wenz, A. (2020), "A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research," *Journal of Survey Statistics and Methodology*, 8, 4–36. https://doi.org/10.1093/jssam/smz041.
Crabtree, C., Kern, H. L., and Pietryka, M. T. (2022), "Sponsorship Effects in Online Surveys," *Political Behavior*, 44, 257–270. https://doi.org/10.1007/s11109-020-09620-7.
Daikeler, J., Roßmann, J., Bretschi, D., Gummer, T., and Silber, H. (2024), "Instructed Response Item Attention Checks in Mixed-Mode Questionnaires: Evidence from a Probability-Based Mail and Online Panel," *Field Methods*, 37, 228–243. https://doi.org/10.1177/1525822X241274103.
DeSalvo, K., Hughes, B., Bassett, M., Benjamin, G., Fraser, M., Galea, S., and Gracia, J. N. (2021), "Public Health COVID-19 Impact Assessment: Lessons Learned and Compelling Needs," *NAM Perspectives*, 1–29, Discussion Paper, Washington, DC: National Academy of Medicine. https://doi.org/10.31478/202104c.
Donovan, D. (2022), "U.S. Surpasses 100 Million Reported COVID-19 Cases," *Johns Hopkins Coronavirus Resource Center*, Available at https://coronavirus.jhu.edu/pandemic-data-initiative/news/u-s-surpasses-100-million-reported-covid-19-cases.
Downes, M., and Carlin, J. B. (2020), "Multilevel Regression and Poststratification Versus Survey Sample Weighting for Estimating Population Quantities in Large Population Health Studies," *American Journal of Epidemiology*, 189, 717–725. https://doi.org/10.1093/aje/kwaa053.
Dutwin, D., and Buskirk, T. D. (2017), "Apples to Oranges or Gala versus Golden Delicious?: Comparing Data Quality of Nonprobability Internet Samples to Low Response Rate Probability Samples," *Public Opinion Quarterly*, 81, 213–239. https://doi.org/10.1093/poq/nfw061.
Fields, J., Hunter-Childs, J., Tersine, A., Sisson, J., Parker, E., Velkoff, V., Logan, C., and Shin, H. (2020), *Design and Operation of the 2020 Household Pulse Survey*, Washington, DC, USA: U.S. Census Bureau, Available at https://www2.census.gov/programs-surveys/demo/technical-documentation/hhp/2020_HPS_Background.pdf
Freese, J., and Jin, O. (2025), "Online Nonprobability Samples," *Annual Review of Sociology*, 51, 109–128. https://doi.org/10.1146/annurev-soc-090524-043117
Ghitza, Y., and Gelman, A. (2013), "Deep Interactions with MRP: Election Turnout and Voting Patterns Among Small Electoral Subgroups," *American Journal of Political Science*, 57, 762–776. https://doi.org/10.1111/ajps.12004.
Green, J., Druckman, J. N., Baum, M. A., Ognyanova, K., Simonson, M. D., Perlis, R. H., and Lazer, D. (2023), "Media Use and Vaccine Resistance," *PNAS Nexus*, 2, pgad146. https://doi.org/10.1093/pnasnexus/pgad146.
Groves, R. M., Presser, S., Tourangeau, R., West, B. T., Couper, M. P., Singer, E., and Toppe, C. (2012), "Support for the Survey Sponsor and Nonresponse Bias," *Public Opinion Quarterly*, 76, 512–524. https://doi.org/10.1093/poq/nfs034.
Haffajee, R. L., and Mello, M. M. (2020), "Thinking Globally, Acting Locally–The U.S. Response to Covid-19," *New England Journal of Medicine, Massachusetts Medical Society*, 382, e75. https://doi.org/10.1056/nejmp2006740.
Jackson, C., Newall, M., Duran, J., Rollason, C., and Golden, J. (2022), "Most Americans not worrying about COVID going into 2022 Holidays," *Ipsos*.
Jerit, J., and Barabas, J. (2023), "Are Nonprobability Surveys Fit for Purpose?," *Public Opinion Quarterly*, 87, 816–840. https://doi.org/10.1093/poq/nfad037.
Karmakar, M., Lantz, P. M., and Tipirneni, R. (2021), "Association of Social and Demographic Factors With COVID-19 Incidence and Death Rates in the US," *JAMA Network Open*, 4, e2036462. https://doi.org/10.1001/jamanetworkopen.2020.36462.
Kawabata, E., Major-Smith, D., Clayton, G. L., Shapland, C. Y., Morris, T. P., Carter, A. R., Fernández-Sanlés, A., Borges, M. C., Tilling, K., Griffith, G. J., Millard, L. A. C., Smith, G. D.,

Lawlor, D. A., and Hughes, R. A. (2024), "Accounting for Bias Due to Outcome Data Missing Not at Random: Comparison and Illustration of Two Approaches to Probabilistic Bias Analysis: A Simulation Study," *BMC Medical Research Methodology*, Springer Science and Business Media LLC, 24, 278. https://doi.org/10.1186/s12874-024-02382-4.

Kim, J.-K., and Tam, S.-M. (2021), "Data Integration by Combining Big Data and Survey Sample Data for Finite Population Inference," *International Statistical Review*, 89, 382–401. https://doi.org/10.1111/insr.12434.

Lavrakas, P. J., Pennay, D., Neiger, D., and Phillips, B. (2022), "Comparing Probability-Based Surveys and Nonprobability Online Panel Surveys in Australia: A Total Survey Error Perspective," *Survey Research Methods*, 16, 241–266. https://doi.org/10.18148/srm/2022.v16i2.7907.

Lipsitch, M., and Santillana, M. (2019), "Enhancing Situational Awareness to Prevent Infectious Disease Outbreaks from Becoming Catastrophic," in *Global Catastrophic Biological Risks*, eds. T. V. Inglesby and A. A. Adalja, Cham: Springer International Publishing, pp. 59–74. https://doi.org/10.1007/82_2019_172.

Lu, F. S., Nguyen, A. T., Link, N. B., Molina, M., Davis, J. T., Chinazzi, M., Xiong, X., Vespignani, A., Lipsitch, M., and Santillana, M. (2021), "Estimating the Cumulative Incidence of COVID-19 in the United States Using Influenza Surveillance, Virologic Testing, and Mortality Data: Four Complementary Approaches," *PLOS Computational Biology, Public Library of Science*, 17, e1008994. https://doi.org/10.1371/journal.pcbi.1008994.

MacInnis, B., Krosnick, J. A., Ho, A. S., and Cho, M.-J. (2018), "The Accuracy of Measurements with Probability and Nonprobability Survey Samples: Replication and Extension," *Public Opinion Quarterly*, 82, 707–744. https://doi.org/10.1093/poq/nfy038.

Marker, D. A., Hilton, C., Zelko, J., Duke, J., Rolka, D., Kaufmann, R., and Boyd, R. (2024), "Real-World Data Versus Probability Surveys for Estimating Health Conditions at the State Level," *Journal of Survey Statistics and Methodology*, 12, 1515–1530. https://doi.org/10.1093/jssam/smae036.

McPhee, C., Barlas, F., Brigham, N., Darling, J., Dutwin, D., Jackson, C., Jackson, M., Kirzinger, A., Little, R., Lorenz, E., Marlar, J., Mercer, A., Scanlon, P. J., Weiss, S., and Wronski, L. (2022), *Data Quality Metrics for Online Samples: Considerations for Study Design and Analysis*, Alexandria, VA, USA: American Association for Public Opinion Research (AAPOR).

Mercer, A., and Lau, A. (2023), "Comparing Two Types of Online Survey Samples," *Pew Research Center*.

Molina, I., and Rao, J. N. K. (2023), "Historical Overview of Small Area Estimation in the 50th Birthday of the IASS," *The Survey Statistician*, 88, 23–25.

National Academies of Sciences, Engineering, and Medicine (2020), *Evaluating Data Types: A Guide for Decision Makers Using Data to Understand the Extent and Spread of COVID-19*, Washington, D.C.: National Academies Press. https://doi.org/10.17226/25826.

Nuzzo, J. B., and Ledesma, J. R. (2023), "Why Did the Best Prepared Country in the World Fare So Poorly during COVID?," *Journal of Economic Perspectives, American Economic Association*, 37, 3–22. https://doi.org/10.1257/jep.37.4.3.

Patel, N. V. (2020), "Why the CDC botched its coronavirus testing," *MIT Technology Review*, Available at https://www.technologyreview.com/2020/03/05/905484/why-the-cdc-botched-its-coronavirus-testing/.

Platoff, E., Walters, E., and Champagne, S. (2020), "Why Texas' coronavirus data comes with caveats," *The Texas Tribune*, Available at https://www.texastribune.org/2020/08/04/texas-coronavirus-data/.

Rader, B., Gertz, A., Iuliano, A. D., Gilmer, M., Wronski, L., Astley, C. M., Sewalk, K., Varrelman, T. J., Cohen, J., Parikh, R., Reese, H. E., Reed, C., and Brownstein, J. S. (2022), "Use of At-Home COVID-19 Tests–United States, August 23, 2021–March 12, 2022," *MMWR. Morbidity and Mortality Weekly Report*, 71, 489–494. https://doi.org/10.15585/mmwr.mm7113e1.

Rao, J. N. K. (2021), "On Making Valid Inferences by Integrating Data from Surveys and Other Sources," *Sankhya B*, 83, 242–272. https://doi.org/10.1007/s13571-020-00227-w.

Rohr, B., Silber, H., and Felderer, B. (2024), "Comparing the Accuracy of Univariate, Bivariate, and Multivariate Estimates across Probability and Nonprobability Surveys with Population Benchmarks," *Sociological Methodology*, 55, 121–154. https://doi.org/10.1177/00811750241280963.

Salomon, J. A., Reinhart, A., Bilinski, A., Chua, E. J., La Motte-Kerr, W., Rönn, M. M., Reitsma, M. B., Morris, K. A., LaRocca, S., Farag, T. H., Kreuter, F., Rosenfeld, R., and Tibshirani, R. J. (2021), "The US COVID-19 Trends and Impact Survey: Continuous Real-Time Measurement of COVID-19 Symptoms, Risks, Protective Behaviors, Testing, and Vaccination," *Proceedings of the National Academy of Sciences, Proceedings of the National Academy of Sciences*, 118, e2111454118. https://doi.org/10.1073/pnas.2111454118.

Salvatore, C. (2023), "Inference with Non-Probability Samples and Survey Data Integration: A Science Mapping Study," *METRON*, 81, 83–107. https://doi.org/10.1007/s40300-023-00243-6.

Salvatore, C., Biffignandi, S., Sakshaug, J. W., Wiśniowski, A., and Struminskaya, B. (2024), "Bayesian Integration of Probability and Nonprobability Samples for Logistic Regression," *Journal of Survey Statistics and Methodology*, 12, 458–492. https://doi.org/10.1093/jssam/smad041.

Santillana, M., Uslu, A. A., Urmi, T., Quintana-Mathé, A., Druckman, J. N., Ognyanova, K., Baum, M., Perlis, R. H., and Lazer, D. (2024), "Tracking COVID-19 Infections Using Survey Data on Rapid At-Home Tests," *JAMA Network Open*, 7, e2435442. https://doi.org/10.1001/jamanetworkopen.2024.35442.

Schulman, J., Druckman, J. N., Safarpour, A. C., Baum, M., Ognyanova, K., Lunz Trujillo, K., Quintana, A., Qu, H., Uslu, A., Perlis, R., and Lazer, D. (2025), "Continuity and Change in Trust in Scientists in the United States: Demographic Stability and Partisan Polarization," *Public Opinion Quarterly*, 89, 1200–1238. https://doi.org/10.2139/ssrn.4929030.

Silber, H., Moy, P., Johnson, T. P., Neumann, R., Stadtmüller, S., and Repke, L. (2022), "Survey Participation as a Function of Democratic Engagement, Trust in Institutions, and Perceptions of Surveys," *Social Science Quarterly*, 103, 1619–1632. https://doi.org/10.1111/ssqu.13218.

Simonson, Block Jr, R., Druckman, J. N., Lazer, D. M., and Ognyanova, K. (2024), *Black Networks Matter: The Role of Interracial Contact and Social Media in the 2020 Black Lives Matter Protests*, Cambridge University Press.

Smolinski, M. S., Crawley, A. W., Baltrusaitis, K., Chunara, R., Olsen, J. M., Wójcik, O., Santillana, M., Nguyen, A., and Brownstein, J. S. (2015), "Flu Near You: Crowdsourced Symptom Reporting Spanning 2 Influenza Seasons," *American Journal of Public Health, American Public Health Association*, 105, 2124–2130. https://doi.org/10.2105/AJPH.2015.302696.

Szilagyi, P. G., Thomas, K., Shah, M. D., Vizueta, N., Cui, Y., Vangala, S., Fox, C., and Kapteyn, A. (2021), "The Role of Trust in the Likelihood of Receiving a COVID-19 Vaccine: Results from a National Survey," *Preventive Medicine*, Elsevier BV, 153, 106727. https://doi.org/10.1016/j.ypmed.2021.106727.

Tourangeau, R., Presser, S., and Sun, H. (2014), "The Impact of Partisan Sponsorship on Political Surveys," *Public Opinion Quarterly*, 78, 510–522. https://doi.org/10.1093/poq/nfu020.

Virnig, B. A., and McBean, M. (2001), "Administrative Data for Public Health Surveillance and Planning," *Annual Review of Public Health*, 22, 213–230. https://doi.org/10.1146/annurev.publhealth.22.1.213.

Wiśniowski, A., Sakshaug, J. W., Pérez Ruiz, D. A., and Blom, A. G. (2020), "Integrating Probability and Nonprobability Samples for Survey Inference," *Journal of Survey Statistics and Methodology*, 8, 120–147. https://doi.org/10.1093/jssam/smz051.

Woolf, S. H. (2022), "The Growing Influence of State Governments on Population Health in the United States," *JAMA*, 327, 1331–1332. https://doi.org/10.1001/jama.2022.3785.

Yasmin, F., Najeeb, H., Moeed, A., Naeem, U., Asghar, M. S., Chughtai, N. U., Yousaf, Z., Seboka, B. T., Ullah, I., Lin, C.-Y., and Pakpour, A. H. (2021), "COVID-19 Vaccine Hesitancy in the United States: A Systematic Review," *Frontiers in Public Health, Frontiers*, 9, 770985. https://doi.org/10.3389/fpubh.2021.770985.

Journal of Survey Statistics and Methodology, 2026, 00, 1–24
https://doi.org/10.1093/jssam/smaf041
Applications