

Stratified Simple Random Sampling Versus Volunteer Community-Wide Sampling for Estimates of COVID-19 Prevalence

Rachel J. Keith, PhD, Rochelle H. Holm, PhD, Alok R. Amraotkar, MD, Megan M. Bezold, J. Michael Brick, PhD, Adrienne M. Bushau-Sprinkle, PhD, Krystal T. Hamorsky, PhD, Kathleen T. Kitterman, MS, Kenneth E. Palmer, PhD, Ted Smith, PhD, Ray Yeager, PhD, and Aruni Bhatnagar, PhD

 See also Elliott, p. 721.

Objectives. To evaluate community-wide prevalence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection using stratified simple random sampling.

Methods. We obtained data for the prevalence of SARS-CoV-2 in Jefferson County, Kentucky, from adult random (n = 7296) and volunteer (n = 7919) sampling over 8 waves from June 2020 through August 2021. We compared results with administratively reported rates of COVID-19.

Results. Randomized and volunteer samples produced equivalent prevalence estimates ($P < .001$), which exceeded the administratively reported rates of prevalence. Differences between them decreased as time passed, likely because of seroprevalence temporal detection limitations.

Conclusions. Structured targeted sampling for seropositivity against SARS-CoV-2, randomized or voluntary, provided better estimates of prevalence than administrative estimates based on incident disease. A low response rate to stratified simple random sampling may produce quantified disease prevalence estimates similar to a volunteer sample.

Public Health Implications. Randomized targeted and invited sampling approaches provided better estimates of disease prevalence than administratively reported data. Cost and time permitting, targeted sampling is a superior modality for estimating community-wide prevalence of infectious disease, especially among Black individuals and those living in disadvantaged neighborhoods. (*Am J Public Health.* 2023;113(7):768–777. <https://doi.org/10.2105/AJPH.2023.307303>)

Accurate estimates of disease prevalence are a prerequisite for evaluating the spread of infectious diseases and for mounting appropriately scaled and targeted public health responses. Since John Snow's tracking to prove his hypothesis, appropriate approaches to the surveillance of community rates of infection have been a matter of ongoing public health debate.¹ More recently, many approaches have been proposed, the most rigorous of which

involve probability and nonprobability sampling.^{2–7} In the United States, estimates of the prevalence of local and imported infectious diseases are based on the National Notifiable Diseases Surveillance System, while many chronic conditions are estimated from a self-reported phone survey.^{8,9} There have been few, if any, attempts to estimate ongoing disease prevalence using stratified simple random sampling, and the utility and biases of this approach

remain unclear, vis-à-vis other modes of sampling.

During the recent COVID-19 pandemic, large-scale spread of infections necessitated rapid and timely estimates of prevalence. However, only administratively reported data were available. Early in the pandemic, testing was limited, but even when it became widely available, it relied on nonprobability sampling, which was disproportionately inaccessible to marginalized

communities¹⁰ and likely to be biased because of a higher probability of voluntary testing by individuals suspecting infections or, in contrast, not seeking testing to avoid quarantine. Most health surveys had to modify their collection programs because of the pandemic.¹¹ Although randomized sampling was used to estimate the prevalence of infection in California, Georgia, Indiana, Oregon, and Rhode Island,^{12–15} these surveys have limitations in sample size or spatial and temporal resolution and did not provide estimates of the reliability and the biases inherent to this approach. Hence, for improved public health response, we developed evidence-based estimates of community-wide prevalence based on stratified simple randomized sampling and compared this with convenience sampling and administratively reported cases.

METHODS

Our study took place in 8 waves between June 2020 and August 2021.

Probability Sampling

To estimate the prevalence of SARS-CoV-2 infections in Jefferson County, Kentucky, we conducted stratified simple random sampling (Table 1). Participants were residents aged 18 years or older. Study procedures included self-report electronic surveys to collect information on demographics, occupation, contact and risk, health history, lifestyle, COVID-19 vaccination (as applicable), and wastewater monitoring awareness (for wave 8 only) as well as professional collection of nasal swabs and blood samples.

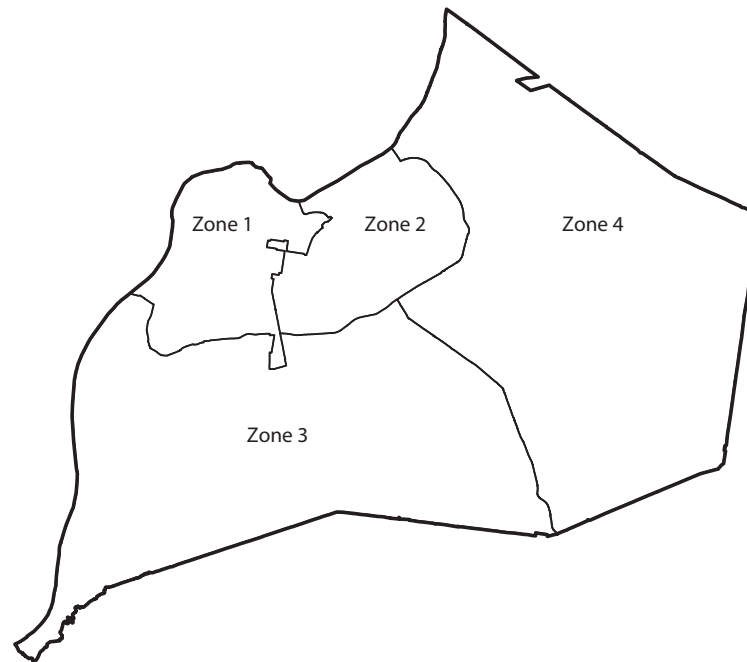
For recruitment, we divided the county into 4 geographic zones (Figure 1) guided by census tract lines to reflect distinctly different demographics (age, race, income, education, and population density). We integrated local knowledge to keep intact macro-neighborhoods encompassing similar cultural identity often delineated along physical geographic boundaries. Zones included at least 100 000 residents.

For the recruitment of the probability sample, we selected households using the address-based sampling frame derived from US Postal Service delivery files.¹⁶ For each wave, we mailed 18 000 to 36 000 invitations across the county (Appendix A, Figure A1, available as a supplement to the online version of this article at <http://ajph.org>). We divided the county into 8 sampling strata, where each of the 4 zones was split in half. We allocated the sample approximately equally to each strata to provide reliable estimates by stratum.

We determined the total number sampled for each wave based on the time and funds required to send out invitations between waves and response rates from previous waves. For the samples from the first 4 waves, we deduplicated them so each address could only be sampled once during these 4 waves. For the fifth wave, we selected an independent sample, and for the subsequent waves, we deduplicated the samples so each household could only be sampled once in waves 5 through 8. It is possible a household

TABLE 1— Demographic Characteristics and Percentage of Antibody-Positive Participants in the Probability and Volunteer Samples for Waves 1 Through 8: Jefferson County, KY, June 2020–August 2021

Characteristic	Probability (Waves 1–8)		Volunteer (Waves 1–8)	
	No. (Weighted %)	Positive Antibody (Weighted %)	No. (Weighted %)	Positive Antibody (Weighted %)
Total	7296 (100.0)	13.2	7919 (100.0)	14.1
Female	4363 (59.8)	12.9	4981 (62.9)	12.6
White	6271 (86.0)	12.7	6489 (81.9)	12.2
Age, y				
18–34	891 (12.2)	14.2	1463 (18.5)	11.6
35–59	2706 (37.1)	12.4	3655 (46.2)	12.6
≥ 60	3699 (50.7)	13.4	2801 (35.4)	13.5
Single family home	6293 (86.3)	13.4	6813 (86.0)	12.3
Smoker	574 (7.9)	8.5	595 (7.5)	8.3
E-cigarette user	181 (2.5)	14.1	219 (2.8)	12.5
Chronic conditions	3924 (53.8)	13.9	3607 (45.5)	13.2
COVID-19 symptoms	438 (6.0)	17.5	725 (9.2)	12.4



Zone	Adults, No.	Age, Years, %				Sex, %		Race, %	
		18–24	25–44	45–64	≥ 65	Male	Female	White	People of Color
Total	594 943	11.2	35.1	34.0	19.8	47.5	52.5	74.8	25.2
1	97 055	15.8	34.3	36.0	13.8	46.2	53.8	39.2	60.8
2	100 029	13.7	39.5	29.4	17.4	47.9	52.1	89.5	10.5
3	206 589	12.2	37.5	33.4	16.8	47.9	52.1	76.3	23.7
4	191 270	8.7	34.7	36.3	20.3	47.6	52.4	83.8	16.2

FIGURE 1— Sampling Zones and Demographic Characteristics Within Jefferson County, KY

could have been sampled once in waves 1 to 4 and sampled again once in waves 5 to 8. When we sampled a household, one adult was randomly selected to participate. For waves 7 and 8, we redefined the strata by wastewater treatment plant sewershed, but we aggregated the results back to the original strata. An invitation to participate in the study was mailed to the address, and we asked the sampled adult to go to a Web site to complete an English screening interview and schedule a testing appointment. Community partners provided incentives in waves 5 and 6.

Our weighting steps included adjustments for (1) sampling the household

from the stratum, (2) sampling 1 adult within the household, and (3) adjusting the estimated totals (raking) to the number of adults in the county from the American Community Survey¹⁷ tabulations (the 2018 5-year data file). The first 2 weighting steps accounted for the probability of selection. The primary goal of the raking was to adjust for nonresponse. The 3 raking dimensions were sex by age, race, and zone. There were no other data on the sampling frame that could reduce nonresponse bias. If any weights were too large, they were trimmed, and the weights were raked again to match the control totals. For variance estimation, we created 50 jackknife replicate

weights for each wave.¹⁸ We used these replicate weights to estimate the standard errors and 95% confidence intervals. The raking may reduce the nonresponse in the estimates but cannot eliminate it.

Volunteer Sampling

For the volunteer sample, social media, community outreach, press conferences, and news outlets, as well as personal contacts with influential community members, were used to invite community participants to sign up online and come to a testing facility. Those who walked up without previously signing up online were also accommodated.

Individuals who consented for testing but were not part of the probability sample were designated “volunteers.” To produce estimates from this sample, we assigned respondents an initial weight of 1 and then raked using the same procedures used for the probability sample. We also created replicate weights to estimate precision, assuming the volunteer sample was equivalent to a simple random sample within the 4 zones. The raking may reduce nonresponse bias but relies on model assumptions that rarely hold in practice.

Testing

Convenient testing dates and times were available for participants to come to drive-up collection sites spread throughout the 4 zones over about 5 days during each wave of testing. Probability and volunteer participants both had trained staff collect nasopharyngeal swabs, which were analyzed for current infection by reverse transcription polymerase chain reaction (RT-PCR) and blood finger-prick samples for the presence of SARS-CoV-2-specific antibodies by serological assessment (enzyme-linked immunoabsorbent assay [ELISA]). Previous exposure to SARS-CoV-2 was assessed via ELISA by measuring immunoglobulin G (IgG) responses to SARS-CoV-2 antigens spike (S), spike receptor binding domain, and nucleocapsid (N) proteins.¹⁹ The presence of IgG to N is highly indicative of natural infection, whereas the presence of IgG to S could be attributable to natural infection or vaccination. The SARS-CoV-2 vaccines were made widely available to the public in Kentucky starting in April 2021. For waves 1 through 4 (June 2020 through February 2021) and for unvaccinated participants in waves 5 through 8 (April through August 2021), previous

exposure to SARS-CoV-2 was determined by presence versus absence of S IgG antibodies in peripheral blood samples. From April through August 2021, for vaccinated participants, previous exposure was determined via presence of N IgG and absence of any self-reported disease or COVID-19-related symptoms before sampling.

Prevalence

To produce prevalence estimates, we made adjustments for sensitivity and specificity depending on the participant's vaccination status. If the participant reported they were not vaccinated, then we made the same adjustments as for the initial period based on the highly reliable S-protein ELISA. The sensitivity of the S-protein ELISA is 100.0 and the specificity is 98.8. If the participant was vaccinated, then we adjusted the estimates for the sensitivity (65.0) and specificity (85.0) of the N-protein test. The test for the N-protein was not available in wave 4; thus, we excluded the 120 vaccinated participants in that wave from our analysis.

Because of the very different measurement properties of the tests, the adjustments should be noted.²⁰ For example, if the observed prevalence rate was 10.0%, then the adjusted estimate would be 10.18% if it were based on the S-protein but 20.0% if based on the N-protein. All estimates of prevalence reported were adjusted unless stated otherwise (Appendix B, available as a supplement to the online version of this article at <http://ajph.org>).

Administratively Reported Data

Administrative data from July 2020 to August 2021 from the Jefferson County

health authority, Louisville Metro Public Health and Wellness, are publicly available. Administrative data for June 2020 are set to zero. We conducted geocoding to the study zones using ArcGIS Pro version 2.8.0 (Esri, Redlands, CA; Appendix C, available as a supplement to the online version of this article at <http://ajph.org>).

Decay

Respondents infected early in 2020 might not still show a positive antibody response if tested in mid-2021 because of seropositivity decay. One approach to make the estimates more comparable is to adjust the administrative statistics to account for the decay in seropositivity at the point of time of the testing and whether the participant was vaccinated or not.²¹ In essence, this approach transforms the administrative estimates of prevalence into estimates of seropositivity, and we refer to the regression estimate of positivity as the “decayed” administrative estimate.

Statistical Analysis

Study data were collected and managed using Research Electronic Data Capture (REDCap) tools hosted at the University of Louisville^{22,23} before being transferred to SAS version 9.4 (SAS Institute Inc, Cary, NC) for data curation and analysis.¹⁸ We computed the estimates and their standard errors using the final and replicate weights using PROC SurveyFreq to account for the complex sample design. The weighting was important in producing the prevalence estimates (before any measurement error adjustments) as the median absolute difference in the estimates for the probability sample was 1.8 percentage points. The analysis used either the *t* test or χ^2 test of significance.

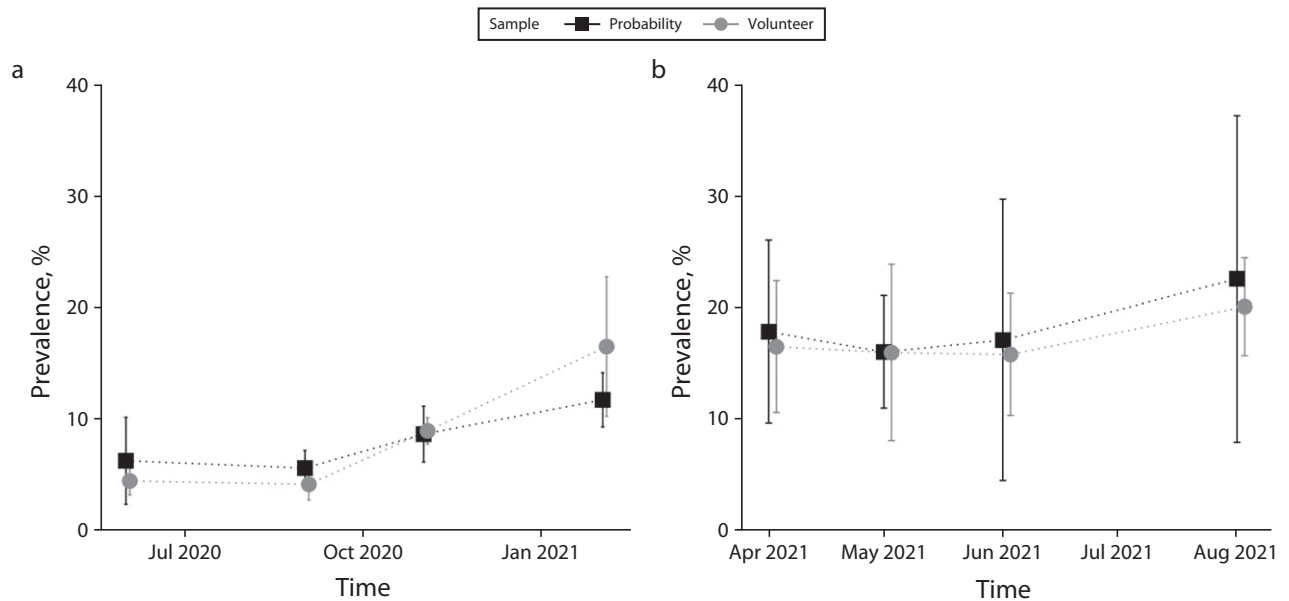


FIGURE 2— Prevalence Estimates for Probability and Volunteer Participants Who Tested Positive for SARS-CoV-2 Infections for (a) Waves 1–4 and (b) Waves 5–8: Jefferson County, KY, June 2020–August 2021

Note. IgG = immunoglobulin G; SARS-CoV-2 = severe acute respiratory syndrome coronavirus 2. Vertical lines represent 95% confidence intervals. Waves 1–4 in panel a present participants positive for SARS-CoV-2 spike (S) protein-specific IgG antibodies. Waves 5–8 in panel b present unvaccinated participants positive for SARS-CoV-2 S protein-specific IgG antibodies and vaccinated participants positive for SARS-CoV-2 nucleocapsid (N)-specific IgG antibodies and absence of any self-reported previous infection or related symptoms before sampling.

RESULTS

Over the 8 waves, most individuals in the probability sample were White (86%) and aged older than 35 years (88%), while 60% were female; the volunteer sample had a similar demographic distribution (Table 1). Distribution was highly clustered to the urban core of the county (Appendix A, Figure A2, available as a supplement to the online version of this article at <http://ajph.org>). Weighting attempted to compensate for these differences. Although we collected nasopharyngeal samples to detect active infections, the low positivity rates observed precluded their use in the analysis. Prevalence estimates in the following sections are, thus, based on positive antibodies. The adult participants in the probability and volunteer samples reported being vaccinated

against COVID-19 at a high rate, 90% by August 2021, which is substantially higher than the nearly 62% total county-wide residents who had received a first vaccine dose by that time.

The response rate for the probability sample (percentage of the sampled cases tested) ranged from 2.4% to 5.5% over the 8 waves. Recruitment of a representative study sample posed a significant challenge throughout the project. The exception was wave 8 (August 2021) with higher prevailing public concern about infection levels during the B.1.617.2 (Delta) surge compared with the original (wild-type) SARS-CoV-2 and its subsequent variants. We also scheduled wave 8 by using prediction modeling from community wastewater sampling,²⁴ which allowed approximation of the Delta variant surge dates. The response rate was higher for both

invited and volunteer study participants during wave 8.

Probability and Volunteer Prevalence

The estimated prevalence estimates from the probability and volunteer samples were comparable ($P > .05$) for each of the first 4 waves (Figure 2a; Appendix A, Figure A3, available as a supplement to the online version of this article at <http://ajph.org>). We examined prevalence estimates by other characteristics including age, sex, race, smoker status, e-cigarette use, chronic conditions, symptoms, and county zone, and the probability and volunteer estimates did not differ substantially ($P > .05$). For waves 5 through 8, the probability and volunteer prevalence estimates were also similar ($P > .05$;

Figure 2b); estimates by other characteristics again did not differ substantially during this period ($P > .05$), with only 1 of 10 differences being statistically significant. Because these 2 sources produced similar estimates, we combined them to produce composite estimates that were based on larger sample sizes and had greater precision. We created the composites separately for each of the 8 waves by weighting the estimates for each source by the number of observations and dividing by the total number of observations. We computed the variances by weighting in the proportion of the observations in the source.

Composite and Administrative Prevalence

The composite estimates of prevalence for waves 1 through 4 showed little variation for men (7.0%) versus women (7.1%), for those with (6.9%) or without (7.2%) chronic conditions, and for e-cigarette users (7.0%) versus non-users (7.0%). However, White (6.3%) versus Black (9.2%) persons with a difference of -3.0% (-4.8% to -1.2%) and smokers (2.4%) versus nonsmokers (7.4%) with a difference of -5.1% (-7.2% to -2.9%) showed diversity.

Early in the pandemic, the estimated prevalence estimates based on compositing the estimate were higher than from administrative sources ($P < .001$ for waves 1 through 4). The composite estimate was, on average, 11 percentage points higher than administrative sources across these 4 waves (Appendix A, Figure A4, available as a supplement to the online version of this article at <http://ajph.org>), but there were some differences in the populations covered. One difference is the administrative sources included all

persons regardless of age. Restricting the administrative source estimate to adults would shift the estimate up only approximately 0.5 percentage points and does not alter the conclusion that the administrative sources underestimated the prevalence rate. When we compared the composite estimates of prevalence for waves 1 through 4 with county administrative sources, the difference for men versus women was small (-0.2 compared with 0.5) but the difference for White versus Black (-3.0 compared with -0.1) was substantial. The administrative data indicated a lower rate of infection for Black individuals than our composite estimate.

For waves 5 through 8, the administrative estimates were closer to the composite estimates, and some of the differences were no longer statistically significant. The administrative statistics exhibited the expected monotone increasing pattern for prevalence, while the composite estimates dipped after February before rising again in August 2021. Because prevalence should be monotonically increasing over time, this dip may be attributable to a decay in the seropositivity of those infected earlier.²¹

When further analyzed for the 4 geographic zones (Figure 3), composite estimates were often higher than the administrative estimates. The sample precision to estimate population differences at this level of disaggregation was low, and the differences were not statistically significant at the .05 level.

Temporal Estimates

The decayed administrative estimate differed only slightly from the unadjusted administrative estimate until wave 5 (April 2021) when the difference was more substantial (Appendix A,

Figure A4, available as a supplement to the online version of this article at <http://ajph.org>). When we compared the decayed administrative estimate to the composite estimate for April through June 2021, we observed the earlier pattern with the composite estimate being higher for every wave, except wave 7, when the difference was not statistically significant (for the other waves all the differences had $P < .05$). The decayed administrative estimate was sharply lower than the composite for August 2021 when the peak percentage of participants' SARS-CoV-2 prevalence was 21%, almost 5 percentage points higher than the June 2021 estimate. The administrative data had a lower increase in reported cases during this period (1.9%) and was less than increases in both February 2021 (4.5%) and November 2020 (2.2%).

DISCUSSION

These unique data from a large study, unduplicated elsewhere, beginning within months of the pandemic emergence, enabled us to faithfully follow the pandemic trajectory and to assess the relative efficacy of different sampling approaches for estimating the community-wide prevalence of infection. We also recognized that probability sampling only would not provide accurate modeling opportunities and that not allowing volunteer participation when COVID-19 testing was limited would have been unethical in an emergency public health crisis. Still, the response rate varied from wave to wave. The limitations with low-prevalence diseases were also corroborated when probability-based sampling of active infections was used for households for HIV.²⁵

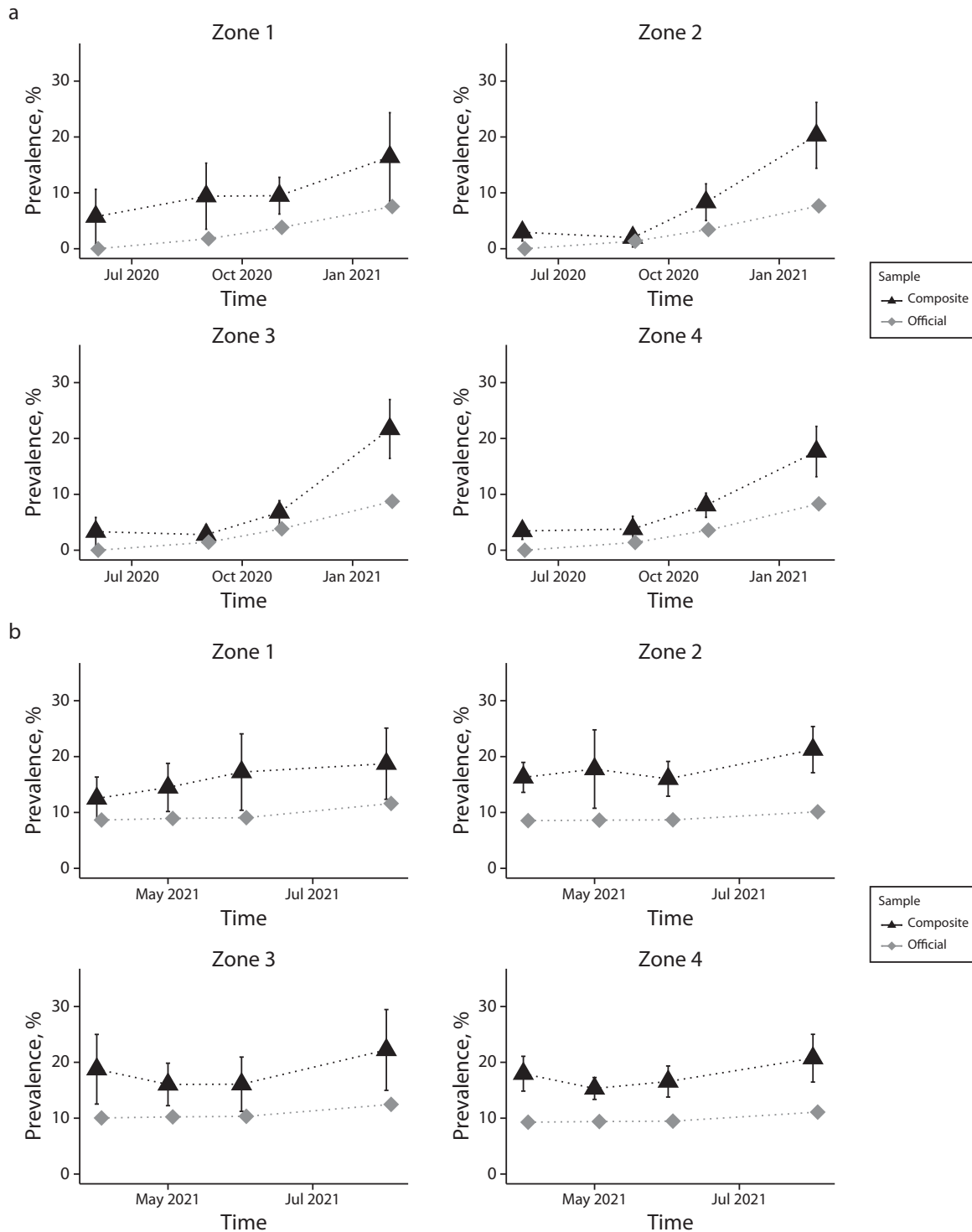


FIGURE 3— Prevalence Estimates for a Composite of Probability and Volunteer Participants Who Tested Positive for SARS-CoV-2 Infections and Administratively Reported Official Rates by Geographic Zones for (a) Waves 1–4 and (b) Waves 5–8: Jefferson County, KY, June 2020–August 2021

Note. IgG = immunoglobulin G; SARS-CoV-2 = severe acute respiratory syndrome coronavirus 2. Administratively reported data are from the Jefferson County health authority, Louisville Metro Public Health and Wellness. Vertical lines represent 95% confidence intervals. Waves 1–4 in panel a present participants positive for SARS-CoV-2 spike (S) protein-specific IgG antibodies and administratively reported official rates. Waves 5–8 in panel b present unvaccinated participants positive for SARS-CoV-2 spike (S) protein-specific IgG antibodies and vaccinated participants positive for SARS-CoV-2 nucleocapsid (N)-specific IgG antibodies and absence of any self-reported previous infection or related symptoms before sampling and administratively reported official rates.

We did not find substantive differences between the prevalence estimates from the volunteer and probability samples by wave. This might be somewhat surprising because self-selection bias in volunteer samples in social and behavioral sciences research has long been considered a serious problem,⁵ although the contribution of these biases may vary.^{2,7} We found even within serological convenience versus random sampling studies there was no clear concurrence, with no substantive differences in vaccine-preventable diseases in schoolchildren,³ while nontyphoid *Salmonella* infections had large differences.⁴ Unique to our study, during the pandemic, large amounts of unverified information being shared through social networks²⁶ might have contributed to selection bias. For example, an invited or volunteer participant would need to be interested in testing to seek out participation, and those who did not believe in COVID-19 or its adverse consequences would not sample regardless of invitation type.

Our low response rate for the probability samples (less than 6%) may result in a nonresponse bias similar to the self-selection bias in the volunteer sample estimates. Virtually all face-to-face surveys collecting physical specimens suspended operations during this time,²⁷ and at the beginning of this pandemic, there were few to no treatment options. So, our study is important in the context of how to recruit participants in a pandemic to support community-related public health and not optimize individual health. Probability sample response rates less than 30% are susceptible to substantial bias.²⁸ Our experience suggests that the probability sample is not likely to produce estimates with lower bias

unless methods for increasing the response rate can be implemented.

Although our volunteer and probability samples gave similar estimates of prevalence, it does not follow they were both unbiased. In both, the percentage who were vaccinated was much higher than county-level estimates. Moreover, even though our sample sizes were too small to detect differences in the 4 county zones, we did see variance in the administrative data per wave, which would suggest that other cities that utilized convenience sampling¹⁰ may also have underreported spatial inequities. Finally, the similarity between the probability and nonprobability sampling results may be because of the high prevalence of the disease and the large number of individuals tested (several thousand), which may have led to similar estimates of disease prevalence, regardless of the sampling approach.

In our analysis, we also aimed to understand the relationship between the tested samples and the administrative statistics as a comparator for sampling methods. Our composite sample has some precedent; it is similar to use of other probability and volunteer hybrid prevalence estimators for HIV²⁹ and in substance use and mental health outcomes.⁶ In our initial period before vaccinations, the composite estimates were consistently between 2- and 4-fold higher than estimates from administrative records. These results suggest that, on a national level, the rates of SARS-CoV-2 infections may have been at least 2-fold higher than previously estimated. We also found statistical differences of the composite estimates of prevalence for waves 1 through 4 for White and minority respondents, and both our probability and volunteer samples had a significantly higher number of minority participants than in administrative

records. This is consistent with the disproportionate impact of the pandemic on minorities.³⁰ Our study suggests that effective outreach improves participation of underserved or marginalized populations.

Waves 5 through 8 were more complicated, relying largely on data from the less reliable N-protein test. This meant that administrative seroprevalence (decayed) estimates were lower than the composite prevalence estimates. However, these adjustments reduce data comparability. If our conclusion that both the probability and volunteer samples are underestimates of prevalence holds, then the administrative statistics underestimate the prevalence to a greater extent than our comparisons show.

Limitations

While the major strength of our study was the longitudinal testing, results are limited by low participation. The low response rates for the probability sample could have introduced biases that could not be eliminated by the weighting. Moreover, because of the availability of vaccines, the antibody testing procedures had to be modified, and the sensitivity and specificity of the tests were vastly different. To help address these issues, we first analyzed the data from June 2020 to February 2021, before most vaccinations and substantial decrease in seropositivity over time, then we analyzed at the later period (through August 2021) as an attempt to compensate for the vaccination and time differences. We assumed that a positive antibody test result could be from 1 or more times of infection. Lastly, while volunteers may have been altruistic for public health research, they may have

enrolled because they thought they had COVID-19.

Public Health Implications

In our study, we found that administrative reports of positive cases during the COVID-19 pandemic underestimated the actual prevalence, which in the early stages of the pandemic may have been 2- to 4-fold higher than reported. Although stratified simple random sampling was superior to administrative record keeping, its efficacy was similar to convenient, invited sampling. Both approaches were limited by a low response rate. Nonetheless, targeted sampling by invitation led to greater participation of the Black population. These findings underscore the importance of community outreach in improving participation in testing and other public health interventions. *AJPH*

ABOUT THE AUTHORS

Rachel J. Keith, Rochelle H. Holm, Alok R. Amraotkar, Ted Smith, Ray Yeager, and Aruni Bhatnagar are with the Christina Lee Brown Envirome Institute, School of Medicine, University of Louisville, Louisville, KY. Megan M. Bezold, Adrienne M. Bushau-Sprinkle, Krystal T. Hamorsky, Kathleen T. Kitterman, and Kenneth E. Palmer are with the Center for Predictive Medicine for Biodefense and Emerging Infectious Diseases, University of Louisville. J. Michael Brick is with Westat Inc, Rockville, MD.

CORRESPONDENCE

Correspondence should be sent to Aruni Bhatnagar, Christina Lee Brown Envirome Institute, School of Medicine, University of Louisville, 302 E Muhammad Ali Blvd, Louisville, KY 40202 (e-mail: aruni.bhatnagar@louisville.edu). Reprints can be ordered at <http://ajph.org> by clicking the "Reprints" link.

PUBLICATION INFORMATION

Full Citation: Keith RJ, Holm RH, Amraotkar AR, et al. Stratified simple random sampling versus volunteer community-wide sampling for estimates of COVID-19 prevalence. *Am J Public Health*. 2023; 113(7):768–777.

Acceptance Date: March 20, 2023.

DOI: <https://doi.org/10.2105/AJPH.2023.307303>

CONTRIBUTORS

R. J. Keith and R. H. Holm are joint first authors. All authors participated in the design and study investigation. J. M. Brick conducted data analyses, and all authors reviewed analysis results. R. H. Holm drafted the article. All authors edited and approved the final article.

ACKNOWLEDGMENTS

This work was supported by contracts from the Centers for Disease Control and Prevention (GB210585) and the Louisville-Jefferson County Metro Government as a component of the Coronavirus Aid, Relief, and Economic Security Act, as well as grants from the James Graham Brown Foundation, the Owsley Brown II Family Foundation, and the Welch Family. Reverse transcription polymerase chain reaction diagnostics for waves 1 to 4 and serology work were also supported in part by the Jewish Heritage Fund and the Center for Predictive Medicine for Biodefense and Emerging Infectious Disease.

Note. The funding sources had no role in the study design; data collection; administration of the interventions; analysis, interpretation, or reporting of data; or decision to submit the findings for publication.

CONFLICTS OF INTEREST

The authors report no relevant competing interests.

HUMAN PARTICIPANT PROTECTION

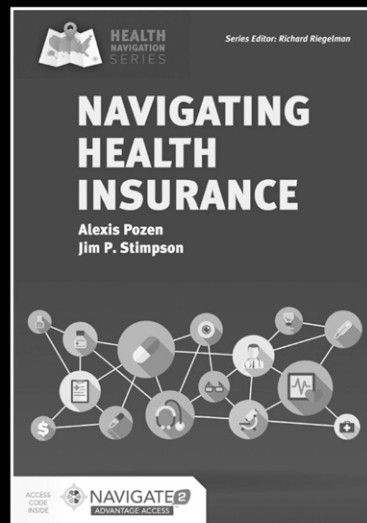
For the seroprevalence and data on severe acute respiratory syndrome coronavirus 2–infected individuals provided by Louisville Metro Public Health and Wellness under a data transfer agreement, the University of Louisville institutional review board (IRB) approved this as human participant research (IRB: 20.0393). Both participant groups consented to agree to participate in the research and provided electronic or written informed consent.

REFERENCES

1. Snow J. *Mode of Communication of Cholera*. 2nd ed. London, England: John Churchill; 1855.
2. Hultsch DF, MacDonald SW, Hunter MA, Maitland SB, Dixon RA. Sampling and generalisability in developmental research: comparison of random and convenience samples of older adults. *Int J Behav Dev*. 2002;26(4):345–359. <https://doi.org/10.1080/01650250143000247>
3. Kelly H, Riddell MA, Gidding HF, Nolan T, Gilbert GL. A random cluster survey and a convenience sample give comparable estimates of immunity to vaccine preventable diseases in children of school age in Victoria, Australia. *Vaccine*. 2002; 20(25–26):3130–3136. [https://doi.org/10.1016/S0264-410X\(02\)00255-4](https://doi.org/10.1016/S0264-410X(02)00255-4)
4. Emborg HD, Simonsen J, Jørgensen CS, et al. Seroincidence of non-typhoid *Salmonella* infections: convenience vs. random community-based sampling. *Epidemiol Infect*. 2016;144(2):257–264. <https://doi.org/10.1017/S0950268815001417>
5. Cheung KL, Peter M, Smit C, de Vries H, Pieterse ME. The impact of non-response bias due to sampling in public health studies: a comparison of voluntary versus mandatory recruitment in a Dutch national survey on adolescent health. *BMC Public Health*. 2017;17(1):276. <https://doi.org/10.1186/s12889-017-4189-8>
6. Drabble LA, Trocki KF, Korcha RA, Klinger JL, Veldhuis CB, Hughes TL. Comparing substance use and mental health outcomes among sexual minority and heterosexual women in probability and non-probability samples. *Drug Alcohol Depend*. 2018;185:285–292. <https://doi.org/10.1016/j.drugalcdep.2017.12.036>
7. Krueger EA, Fish JN, Hammack PL, Lightfoot M, Bishop MD, Russell ST. Comparing national probability and community-based samples of sexual minority adults: implications and recommendations for sampling and measurement. *Arch Sex Behav*. 2020;49(5):1463–1475. <https://doi.org/10.1007/s10508-020-01724-9>
8. Centers for Disease Control and Prevention. National Notifiable Diseases Surveillance System (NNDSS). Available at: <https://www.cdc.gov/nndss/index.html>. Accessed May 25, 2022.
9. Pickens CM, Pierannunzi C, Garvin W, Town M. Surveillance for certain health behaviors and conditions among states and selected local areas—Behavioral Risk Factor Surveillance System, United States, 2015. *MMWR Surveill Summ*. 2018;67(9): 1–90. <https://doi.org/10.15585/mmwr.ss6709a1>
10. Bilal U, Tabb LP, Barber S, et al. Spatial inequities in COVID-19 testing, positivity, confirmed cases, and mortality in 3 US cities: an ecological study. *Ann Intern Med*. 2021;174(7):936–944. <https://doi.org/10.7326/M20-3936>
11. Ward BW, Sengupta M, DeFrances CJ, Lau DT. COVID-19 pandemic impact on the national health care surveys. *Am J Public Health*. 2021; 111(12):2141–2148. <https://doi.org/10.2105/AJPH.2021.306514>
12. Menachemi N, Yiannoutsos CT, Dixon BE, et al. Population point prevalence of SARS-CoV-2 infection based on a statewide random sample—Indiana, April 25–29, 2020. *MMWR Morbid Mortal Wkly Rep*. 2020;69(29):960–964. <https://doi.org/10.15585/mmwr.mm6929e1>
13. Chan PA, King E, Xu Y, et al. Seroprevalence of SARS-CoV-2 antibodies in Rhode Island from a statewide random sample. *Am J Public Health*. 2021;111(4):700–703. <https://doi.org/10.2105/AJPH.2020.306115>
14. Layton BA, Kaya D, Kelly C, et al. Evaluation of a wastewater-based epidemiological approach to estimate the prevalence of SARS-CoV-2 infections and the detection of viral variants in disparate Oregon communities at city and neighborhood scales. *Environ Health Perspect*. 2022;130(6):67010. <https://doi.org/10.1289/EHP10289>
15. Sullivan PS, Siegler AJ, Shioda K, et al. Severe acute respiratory syndrome coronavirus 2 cumulative incidence, United States, August 2020–December 2020. *Clin Infect Dis*. 2022;74(7):1141–1150. <https://doi.org/10.1093/cid/ciab626>
16. Iannacchione VG. The changing role of address-based sampling in survey research. *Public Opin Q*. 2011;75(3):556–575. <https://doi.org/10.1093/poq/nfr017>
17. US Census Bureau. 2018 American Community Survey (ACS). US Department of Commerce. Available at: <https://www.census.gov/data/developers/data-sets/acs-5year.html>. Accessed June 21, 2021.

18. Rust KF, Rao JN. Variance estimation for complex surveys using replication techniques. *Stat Methods Med Res.* 1996;5(3):283–310. <https://doi.org/10.1177/096228029600500305>
19. Hamorsky KT, Bushau-Sprinkle AM, Kitterman K, et al. Serological assessment of SARS-CoV-2 infection during the first wave of the pandemic in Louisville Kentucky. *Sci Rep.* 2021;11(1):18285. <https://doi.org/10.1038/s41598-021-97423-z>
20. Sempos CT, Tian L. Adjusting coronavirus prevalence estimates for laboratory test kit error. *Am J Epidemiol.* 2021;190(1):109–115. <https://doi.org/10.1093/aje/kwaa174>
21. Alfego D, Sullivan A, Poirier B, et al. A population-based analysis of the longevity of SARS-CoV-2 antibody seropositivity in the United States. *EclinicalMedicine.* 2021;36:100902. <https://doi.org/10.1016/j.eclinm.2021.100902>
22. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42(2):377–381. <https://doi.org/10.1016/j.jbi.2008.08.010>
23. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium. Building an international community of software platform partners. *J Biomed Inform.* 2019;95:103208. <https://doi.org/10.1016/j.jbi.2019.103208>
24. Smith T, Holm RH, Keith RJ, et al. Quantifying the relationship between sub-population wastewater samples and community-wide SARS-CoV-2 seroprevalence. *Sci Total Environ.* 2022;853:158567. <https://doi.org/10.1016/j.scitotenv.2022.158567>
25. Shapiro MF, Berk ML, Berry SH, et al. National probability samples in studies of low-prevalence diseases. Part I: Perspectives and lessons from the HIV cost and services utilization study. *Health Serv Res.* 1999;34(5 pt 1):951–968.
26. Yu H, Yang CC, Yu P, Liu K. Emotion diffusion effect: negative sentiment COVID-19 tweets of public organizations attract more responses from followers. *PLoS One.* 2022;17(3):e0264794. <https://doi.org/10.1371/journal.pone.0264794>
27. Paulose-Ram R, Graber JE, Woodwell D, Ahluwalia N. The National Health and Nutrition Examination Survey (NHANES), 2021–2022: adapting data collection in a COVID-19 environment. *Am J Public Health.* 2021;111(12):2149–2156. <https://doi.org/10.2105/AJPH.2021.306517>
28. Hedlin D. Is there a “safe area” where the nonresponse rate has only a modest effect on bias despite non-ignorable nonresponse? *Int Stat Rev.* 2020;88(3):642–657. <https://doi.org/10.1111/insr.12359>
29. Hedt BL, Pagano M. Health indicators: eliminating bias from convenience sampling estimators. *Stat Med.* 2011;30(5):560–568. <https://doi.org/10.1002/sim.3920>
30. Tai DBG, Shah A, Doubeni CA, et al. The disproportionate impact of COVID-19 on racial and ethnic minorities in the United States. *Clin Infect Dis.* 2021;72(4):703–706. <https://doi.org/10.1093/cid/ciaa815>

A Practical, Balanced Guide to Understanding Health Insurance



Written from the perspective of the consumer, this new text is a comprehensive yet accessible examination of health insurance in the United States.

Instructor exam copies available at go.jblearning.com/Pozen

APHA PRESS
AN IMPRINT OF AMERICAN PUBLIC HEALTH ASSOCIATION

JONES & BARTLETT LEARNING
An Ascend Learning Company