Survey Research Methods (2020)
Vol. 14, No. 2, pp. 131-139
doi:10.18148/srm/2020.v14i2.7757

ISSN 1864-3361
http://www.surveymethods.org

Design for a Mail Survey to Determine Prevalence of SARS-CoV-2 Antibodies in the United States

Alicia M. Frasier
RTI International

Heidi Guyer
RTI International

Laura DiGrande
RTI International

Rose Domanico
RTI International

Darryl Cooney
RTI International

Stephanie Eckman
RTI International

Although counts of the novel Coronavirus (SARS-CoV-2) infections and deaths are reported by several sources online, precise estimation of the exposed proportion of the population is not possible in most areas of the world. Estimates of other disease prevalence in the United States are often obtained through in-person seroprevalence surveys. The availability of testing only for individuals with symptoms, combined with stay-at-home and social distancing mandates to stem the spread of the disease, limit in-person data collection options. A probability-based mail survey with at-home, self-administered testing is a feasible method to safely estimate SARS-CoV-2 antibody prevalence within the United States while also easing burden on the U.S. public and health care system. This mail survey could be a one-time, cross-sectional design, or a repeated cross-sectional or longitudinal survey. We discuss several options for designing and conducting this survey.

Keywords: COVID-19; prevalence survey; mailed data collection; self-administered at-home testing

1 Introduction

Although daily counts of SARS-CoV-2 infections and deaths are reported by many sources online, precise estimation of a prevalence rate is not possible in most areas of the world. Due to the shortage of tests in many countries, those who show symptoms of the SARS-CoV-2 illness (COVID-19) are more likely to be tested (Spinelli & Pellino, 2020). However, some who get infected show no symptoms, even though they are able to spread the virus to others (Day, 2020). To estimate how widespread the infection has been in a country such as the United States, random testing for antibodies to the virus is needed. Antibody testing can reveal who has been infected, not just who is currently infected or who has shown common COVID-19 symptoms. A population-based SARS-CoV-2 antibody prevalence estimate would allow public health officials and policymakers to allocate resources where immunity prevalence is low, assess current cloth face covering, stay-at-home, or other orders, and guide vaccination program implementation planning.

Contact information: Alicia M. Frasier, 30 W. Monroe Street, Suite 2100 Chicago, IL, United States 60606-4901 (E-mail: afrasier@rti.org)

The design and execution of a mail survey that uses self-administered, at-home antibody testing with a population-based sample of non-institutionalized residents is a feasible and safe method to assess the prevalence of SARS-CoV-2 antibodies in the U.S. population. Mail surveys tend to be less expensive than in-person surveys and also provide population-based data faster than in-person field surveys. This survey could be tailored to represent the whole country or a specific state or city. Although testing is currently occurring in hospitals, clinics, and commercial sites, implementing a population-based design would mitigate issues of coverage bias from unequal distribution of sites and selection bias from guidelines that prioritize tests for symptomatic individuals. The survey we discuss in this paper would provide a more accurate estimate of SARS-CoV-2 antibody prevalence in the population. Although it is still unclear whether positive immunological markers confer immunity, this alternative design captures the prevalence of those markers and the virus's symptom profile from a population-based sample; this is an added value not possible at testing sites, clinics, or hospitals where asymptomatic and mildly symptomatic individuals are unlikely to be present. Obtaining questionnaire data on demographics, essential-worker status, and symptoms through this design also allows for stratification of disease prevalence and immunity marker types across subpopulations.

Serosurveys for SARS-CoV-2 are being planned or have begun in some local jurisdictions and countries around the world. Some have used non-probability samples (Bendavid et al., 2020; "NIH begins study to quantify undetected cases of coronavirus infection," 2020; The Boston Globe, 2020). These studies have been criticized as potentially providing higher seropositive rates than truly exist, because individuals with symptoms of COVID-19 who had not yet been tested may have been more motivated to participate. The prior studies also vary across sample type (e.g., residents, health care workers), testing locations (e.g., neighborhood site, public health department), diagnostic test brand/type, and sensitivity/specificity. This lack of standardization leads to variability in the results unrelated to true geographic variation in prevalence. The best way to estimate the prevalence in the U.S. population is to conduct a general population probability-based study, such as the one proposed here.

All countries need nationally representative estimates of the prevalence of antibodies to SARS-CoV-2, by demographic and policy-relevant subgroups, such as age, race, ethnicity, region, and essential-worker status. Below we provide our thoughts on how such a survey could be conducted in the United States.

2 Sample Design

Although the United States does not have a national population register, there is a frame of residential mailing addresses, maintained and updated by the U.S. Postal Service. A representative sample of household addresses with high coverage can be selected from this frame (Amaya, Zimmer, Morton, & Harter, 2018; Battaglia et al., 2016; Eckman & English, 2012; Peytchev, Ridenhour, & Krotki, 2010).

2.1 Sample Size

To be useful to public health researchers, the confidence interval on the resulting prevalence estimates should be narrow enough that it provides meaningful information. Figure 1 displays the sample sizes, on the y-axis, needed to detect a given significant difference, on the x-axis. The three lines refer to three null hypotheses about the prevalence rate (1%, 5%, and 10%). If the prevalence rate is near 1%, then a sample size of 5,000 cases would result in a 95% confidence interval of 0.63%-1.42%. Note that this number refers to a national sample size. If we wish to estimate prevalence rates for subgroups such as major urban centers, age groups, or race/ethnicity groups, we will need 5,000 in each group. Some groups of interest can be targeted geographically by oversampling areas with high concentrations of racial and ethnic minorities. Other groups of interest, such as the elderly, cannot be targeted geographically, which makes it harder to meet a target sample size.

Researchers will likely also want to compare prevalence rates across subgroups. The required sample sizes for such

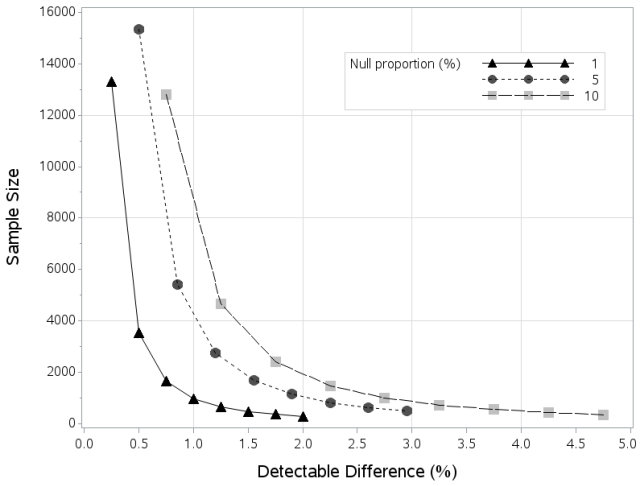

Figure 1. Necessary sample size to detect a difference for different hypothesized prevalence rates

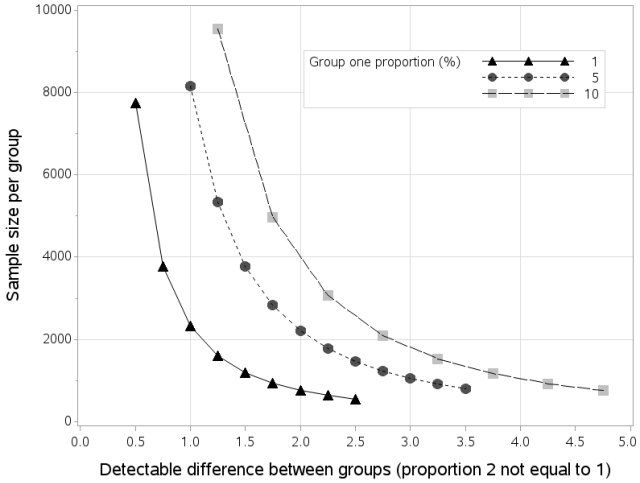

Figure 2. Necessary sample size per group to detect a difference between two groups assuming different hypothesized prevalence rates

comparison are shown in Figure 2. If the prevalence rate in Group A is 5% and we wish to detect a difference of 2 percentage points in the rates of Groups A and B (with 80% power and 5% type 1 error), we will need at least 2,213 completed cases in each group. Adjustment to the sample sizes shown in the two figures may be necessary to account for the false positive and false negative rates of the test used.

The desired number of completed cases must be inflated to account for returned mail, survey nonresponse, and nonviable samples. For example, if we were to assume 25% of surveys are returned with tests and 95% of samples are viable, then we need to sample and mail to approximately 21,052 addresses to collect 5,000 viable tests. The sample of addresses from the address frame should be stratified by zip code, since infection is geographically clustered.

We recommend selecting many more addresses and dividing them into replicates of 2,000 or so. Each replicate would be a random sample of the population. Releasing the sample in replicates would allow for adjustment if the assumed rates are too low or too high. We might start by releasing eight replicates (16,000 addresses) to monitor the response and test viability rates. Then additional replicates could be released to meet the target number of completes. Replicates would also allow the survey to expand if additional data collection resources are found, which would increase the power of the study to detect differences between subgroups.

2.2 Selection of Respondent within Household

For cost and statistical reasons, the best approach is to randomly sample one adult in each selected household. Each test involves costs to purchase, mail, and analyze. In addition, there is likely little statistical value in administering the test to more than one household member. Because the virus is highly contagious, the likelihood of multiple household members sharing the same antibody status is high.

However, random selection of one adult in a household in a self-administered survey is challenging. We foresee three options for the selection of one adult. Option 1 is to use the youngest male/oldest female method or the last (or next) birthday method. The household could then select the respondent and give the survey and test to him or her. However, households do not always follow these instructions and in practice, the selected respondents do not match the population (Lynn, 2019; Olson & Smyth, 2017, 2014). Option 2 is to give no specific instructions and allow any adult in the household to participate. This approach may in practice be very similar to Option 1 if few households follow the selection procedure. Some households will give the test and survey to those likely have been infected. If all adults in the household have the same antibody status, then it would not matter (for the test result) who participates. Option 3 is to select all adults in the household. Given the cost of the tests and the likely high correlation in results, this approach seems wasteful. However, it would likely be attractive to respondents who want to know if any household member has been exposed. The approach might also be valuable for longitudinal studies.

Unfortunately, there is no ideal method for selecting respondent(s) within households without interviewer involvement. Below we presume that Option 1 will be used.

3 Data Collection Methods

A self-administered mail or web survey offers low costs and high coverage. Below we discuss how the survey could be conducted. We do not discuss details on which test the survey should use because the testing landscape is continually changing. We presume that a test exists with reasonable sensitivity and specificity.

3.1 Initial Mailing

Data collection begins by mailing a packet to all addresses containing: a letter, a paper survey, one self-administered test kit, and a prepaid addressed return envelope. The introductory letter should provide a website and phone number for additional information and offer the option to complete the survey by web or phone. It should also contain instructions for selecting one adult in the household to complete both the survey and specimen collection.

The self-administered test kit should include the specimen collection kit (nasal or oral/buccal swab; finger-prick; saliva), latex gloves, self-collection instructions (with a link to a video demonstration and an option to video chat with a collection expert), specimen packaging materials, and a postage-paid return envelope. All materials, including the survey and self-collection instructions, should be available in English and Spanish. Follow-up mailings with nonrespondents should also be sent, though we do not recommend sending additional test kits, due to their cost.

3.2 Questionnaire

The questionnaire should include succinct questions that the public is accustomed to answering. We suggest a one-page paper survey with demographic questions, an item about whether the respondent is an essential worker, a checklist of symptoms, and previous SARS-CoV-2 testing results. Use of categories for each variable that correspond to the CDC's Case Report Form ("Information for Health Departments on Reporting Cases of COVID-19," 2020) allows for comparison to data reported in tables by the CDC's weekly surveillance summary ("Key updates for week 17, ending April 25, 2020. COVIDView: A weekly surveillance summary of U.S. COVID-19 activity," 2020). To maximize response among the Hispanic population, the questionnaire should be printed in English on one side and Spanish on the other.

The questionnaire would ask whether the respondent had symptoms associated with COVID-19, had been tested, and the result of the test (if tested). It would ask for participants' permission to recontact to allow for any follow-up surveys/contact. Complete and accurate follow-up information will be especially important if results are relayed to respondents or if longitudinal testing is conducted to detect seroprevalence changes over time among individuals who previously tested negative. Additionally, there may be some benefit to re-contacting positive individuals if a future wave of SARS-CoV-2 takes place to determine the degree of protection immunological markers may confer.

3.3 Testing Methods

Using self-administered kits to collect samples would relieve the burden on the health care system to provide this

information. It would also reduce risk and burden on respondents: no interviewer would visit their home and they would not have to visit a testing site. As the pandemic continues, the health care system's focus will continue to be on diagnosing and treating those who are symptomatic or have been exposed. This type of testing is crucial but will not allow public health researchers to estimate the true prevalence rate among the U.S. adult population, for the reasons discussed previously. In addition, pooled testing has recently been proposed as an efficient strategy to detect SARS-CoV-2 infections and determine return-to-work status. While pooled testing is useful for screening purposes of select groups of individuals, essential personnel, or travelers, it may not be ideal for the purposes of generating population-based prevalence estimates nor to determine the prevalence by race, ethnicity, or other factors of interest.

Serology tests detect the presence of immunological antibodies, referred to as immunoglobulin, currently present. Immunoglobulin M, or IgM, antibodies are produced early in the infection period, typically within a week, and decline within two weeks after reaching a peak. IgM antibodies are not detectable in the first week of infection, during which an infected individual is contagious, and are no longer detectable approximately three weeks after the initial infection. Immunoglobulin G, or IgG, is produced later in the infection window than IgM antibodies and remain detectable for a longer period of time, potentially indefinitely. Detection of IgG occurs approximately fourteen days after infection occurs and is an indicator that an individual has been exposed to the virus and developed antibodies. IgG remains in the blood and may provide long-term immunity to future SARS-CoV-2 infections or severity of symptoms.

There are several viable methodologies to collect specimens to detect SARS-CoV-2 antibodies. Although the U.S. Food and Drug Administration (FDA) is slowly approving nasal swabs, finger pricks, saliva collection, and buccal swabs for at-home COVID-19 testing, wider approval is anticipated in the near future. As of May 11, 2020, twelve tests have received Emergency Use Authorization approval for use by the FDA. Not all of these are viable for in-home or self-administered specimen collection. Research or public health responses may be exempt from this approval. Testing supplies are limited in the United States but are expected to become more readily available.

The validity of a test is determined by the extent to which the test accurately measures the disease of interest. The validity of a screening test is measured by the specificity and sensitivity rates which can be considered forms of testing error. To reduce the error, the study could administer more than one test to each subject. Testing using more than one method would increase the positive predictive value, the likelihood that a positive test is a true indication of antibody presence. If testing with two methods is feasible, the test with the higher sensitivity should be used first, followed by a test with higher specificity. This approach increases costs but also provides a more complete picture of the true positives. Such methods are especially important for rare diseases, or those with a prevalence rate of 1% or lower (Pottinger & Sia, 2020). In the case of this pandemic, although we are still determining the prevalence of the virus itself, it is clear that the rate varies geographically and demographically.

Numerous surveys have had positive results with self-collection of biological specimens including nasal (Lunny et al., 2015), buccal (Walter, Dole, Siega-Riz, & Entwisle, 2011; Woody, Hamilton, Livitz, Figueroa, & Zoccola, 2017), or vaginal swabs (Jaszczak, Lundeen, & Smith, 2009; Lindau et al., 2009; Suzman, 2009); saliva (Crimmins et al., 2013; Dykema, DiLoreto, Croes, Garbarski, & Beach, 2017; Sastry, Fomby, & McGonagle, 2017); finger-pricks with blood (Health and Retirement Study, 2007; Sakhi et al., 2015) collected on filter paper, capillary tubes, or on a test strip such as diabetics use to test glucose. Participants have also provided self-collected samples of biological fluids including saliva, urine, feces (Herd et al., 2017), blood, and other samples such as hair and fingernails or toenails. Consent rates are typically higher on longitudinal studies where participants have some degree of familiarity, association, and loyalty to the study (Gatny, Couper, & Axinn, 2013; Sakshaug, Ofstedal, Guyer, & Beebe, 2014). However, samples have been successfully collected from cross-sectional surveys or in a baseline year of a longitudinal survey as well (Jaszczak et al., 2009). Specimen collections can also be successful in web, phone, and mail surveys (Dykema et al., 2017; Sastry et al., 2017), as well as in person (Crimmins et al., 2013; Jaszczak et al., 2009; Pramanik et al., 2012). Prevalence rates of various health conditions have been successfully estimated using these self-administered biospecimen collection methods. While self-administered SARS-CoV-2 serology tests are still under development and review for use, self-collection tests are used to test for antibodies to other viruses, such as HIV or HPV (Jaszczak et al., 2009; Lindau et al., 2009; Suzman, 2009).

Although collection of biologic specimens by respondents themselves is challenging, we believe that the high relevance of COVID-19 will lead to a high consent rate. Respondents will likely be motivated to take part by the opportunity to learn about their own antibody status but also to help others and contribute to knowledge about the disease.

3.4 Specimen Collection

Special care should be taken at each step of the process to ensure that the sample is collected properly and is viable for analysis purposes. Collection kits must contain all the collection materials, collection instructions, and packaging and shipping materials and instructions clearly described for the participant. Ideally these materials should be sent in both En-

glish and Spanish, with instructions in additional languages available online.

The sample collection environment must be as clean as possible to ensure that the specimen is not contaminated. For example, imagine SARS-CoV-2 is present on a counter or other surface and a swab is set on that surface, either before or after specimen collection occurs. The swab could then give a positive result, but not an accurate result for the sample member. Instructions should also be provided to participants to thoroughly wash their hands before and after sample collection, again to reduce the likelihood of any cross-contamination. Respondents will need additional information both on how to obtain a sufficient sample and examples of what is considered insufficient. For example, how far to extend the nasal swab into the nasal cavity, the areas of the mouth and areas of the swab to coat with a buccal swab, or how to obtain an adequate finger stick and sufficient quantity of blood. Clear instructions must also be provided if the participant should do any processing of the sample, such as inserting the swab in a vial with a reagent, mixing a reagent with the blood collected in a capillary tube, or sufficiently shaking the container to disperse the reagent and coat the specimen. The respondent will also need directions on how to package the sample to ensure that it remains intact during shipping (i.e., enclose in a second container or bubble wrap). Instructions should also be provided as to how to appropriately clean and sanitize the specimen collection area after collection is complete and how to prevent cross-contamination if the individual thinks that they may have an active SARS-CoV-2 infection.

Additional resources can be provided to the respondent to increase the rate of viable samples among those collected:

- A website the respondent can access with additional information, pictures, and videos of proper collection.
- A virtual chat feature to collect the sample while an interviewer or study team member is providing guidance and reassurance via a video connection. This method may lead to improved collection results and may be more acceptable to respondents given the current increase in Telehealth visits because of the current pandemic.
- A toll-free phone number to call for assistance, to receive answers to questions related to collection, packaging and shipping, and to request any additional collection supplies (if additional lancets are needed or the capillary tube breaks, for example).
- A follow-up phone call from a knowledgeable interviewer or study staff member to ensure collection and answer any questions related to the collection, packaging, or shipping could be made if telephone numbers are available.

3.5 Return Packaging and Mailing

Most collection kits currently available require that the specimen be mailed to a laboratory for processing. One of the most important considerations regardless of the type of specimen collected is how long the specimen can be kept at room temperature without degrading, or if the sample must be stored at lower temperatures from the time of collection. A testing mechanism that requires samples to be maintained at -30C from the time of collection would not be viable for this study, because dry ice would be required. However, samples that must be kept cold could be mailed with gel-activated ice packs that are provided to the respondent. All samples should be mailed in a container that offers protection from breakage and leakage because of handling, temperature, or pressure changes. Samples should be shipped using the most expedient method to ensure that they arrive at the lab as quickly as possible, and are logged and processed, or frozen, while still intact. The U.S. postal service requires that packages containing biological fluids carry a biohazard sticker visible on the front. Time and day of shipping the sample and receipt at the lab or storage facility should be considered as well to ensure the timely logging of receipt and processing of the sample. If the specimen is stored until processing occurs, storage requirements will need to be taken into account, to ensure adequate space, temperature, and lighting to maintain the integrity of the samples.

3.6 Reporting of Results

Results would be communicated to the respondent or any other required parties (CDC), as required by Institutional Review Boards and aligned with best ethical practices. Reporting to the local or state health department may also be required. Delivery of results should also include resources for further information, follow-up treatment, or counseling, as necessary.

Another option is to use tests that report results quickly and directly to the respondent. These are referred to as Rapid Tests and are similar to a home pregnancy test or a glucose check conducted by diabetics. Rapid tests provide results in 15 minutes or less. These tests are currently the least reliable, but they may increase in precision over time and when accompanied by a second, non-rapid test. Several self-collection kits currently in use, and a number that are under development, allow the respondent to collect the sample and view the results. The sample is collected, typically a drop of blood, and placed on a sheet of treated filter paper that is enclosed in a plastic cassette. Often, a reagent or buffer solution is applied, and colored lines appear on the filter paper within 15 minutes to indicate the presence of IgG or IgM antibodies. Respondents would then be responsible for relaying the results to the study team, by calling a phone number, taking a picture and uploading it to a website or sending it to

an email address, collecting the sample while video chatting with an interviewer and showing the collection device to the interviewer, or even mailing the collection device back to the study team.

3.7 Costs

The data collection costs for the survey we propose would include labor for creation and mailing of the materials; test kits; printing and postage for mailed materials, including test kits; return postage for test kits and questionnaires; nonresponse follow-up in the form of mailings or telephone calls; analysis of test results; and communication of test results. If this survey were executed as an in-person data collection, there would still be the need for creation and printing of written materials, mailing lead letters, purchase of test kits, and analysis of the tests. The additional cost for an in-person survey would come from the training, travel, and labor costs associated with sending interviewers to respondents' homes. This cost would be significant when compared to printing and postage costs. In addition, because the pandemic continues, in-person data collection would add additional risk for project staff to train interviewers (if an in-person training were conducted), interviewers to be in respondents' homes, and respondents allowing interviewers into their homes.

4 Conclusion

The study outlined in this paper supports the estimation of the prevalence of antibodies in the U.S. adult population and in important subgroups. Development and authorization of new testing methods are proceeding rapidly, but for the next several months, the observed prevalence of SARS-CoV-2 in the population will still be inaccurate. A probability-based, self-administered mail survey containing a test kit and questionnaire is a cost-effective, fast, and viable method to collect prevalence data. Using survey research methods to estimate the burden of SARS-CoV-2 in the United States population circumvents several issues related to hospital- and clinic-based testing.

Strengths of this approach include known probabilities of selection for participants; broad inclusion criteria that extend beyond symptomatic individuals; scalability of the sample size via replicates; and ability to obtain relatively fast results. The mail survey could be a one-time, cross-sectional design, or a repeated cross-sectional or longitudinal approach. However, the protocol needs to include an adequate sample size that allows for stable and reliable population estimates, particularly among demographic subgroups of interest. Other challenges include the need to ensure representativeness to the population of interest whether it be the United States as a whole, or a state or city, and the current limited availability of testing supplies.

The study design is flexible and could accommodate several additional options to increase the statistical power of the results to detect policy-relevant phenomena, including sample expansion to include more cases for more accurate subgroup comparisons; administering a second, different test to improve the positive predictive value of the results; and periodically re-contacting participants to collect additional samples at a pre-specified frequency (every 2-4 weeks) or during a future outbreak.

The SARS-CoV-2 prevalence mail survey should share testing results with each participant. Although questions may exist over the ability to ensure that private information with biological samples is secure across the U.S. Postal System, the increased use of such methods for HIV and genetic testing indicate that the public is risk aware. However, the survey must provide follow-up that addresses the psychosocial aspects of SARS-CoV-2 infection and educates the respondent that detection of antibodies through serological tests may not mean that he or she is fully immune to reinfection. A SARS-CoV-2 prevalence mail survey is not the terminal point in public health surveillance for the disease, but it is a critical step to increase the understanding of the pandemic and to set the stage for discourse with policymakers regarding the programs and initiatives necessary to mitigate its burden.

Acknowledgments

The authors thank Paul Pulliam for discussion of the ideas and Jerry Rench for comments on an earlier version of the paper.

References

Amaya, A., Zimmer, S., Morton, K., & Harter, R. (2018). Does undercoverage on the U.S. address-based sampling frame translate to coverage bias? *Sociological Methods & Research*. doi:10.1177/0049124118782539

Battaglia, M. P., Dillman, D. A., Frankel, M. R., Harter, R., Buskirk, T. D., McPhee, C. B., ... Yancey, T. (2016). Sampling, data collection, and weighting procedures for address-based sample surveys. *Journal of Survey Statistics and Methodology*, *4*(4), 476–500. doi:10.1093/jssam/smw025

Bendavid, E., Mulaney, B., Sood, N., Shah, S., Ling, E., Bromley-Dulfano, R., ... Bhattacharya, J. (2020). COVID-19 antibody seroprevalence in Santa Clara county, California. doi:10.1101/2020.04.14.20062463

Crimmins, E. M., Faul, J. D., Kim, J. K., Guyer, H., Langa, K. M., Ofstedal, M. B., ... Weir, D. R. (2013). Documentation of biomarkers in the 2006 and 2008 Health and Retirement Study. Institute for Social Research, University of Michigan. Retrieved from https://hrs.isr.umich.edu/sites/default/files/biblio/Biomarker2006and2008.pdf

Day, M. (2020). COVID-19: Four fifths of cases are asymptomatic, China figures indicate. *BMJ*, m1375. doi:10.1136/bmj.m1375

Dykema, J., DiLoreto, K., Croes, K. D., Garbarski, D., & Beach, J. (2017). Factors associated with participation in the collection of saliva samples by mail in a survey of older adults. *Public Opinion Quarterly*, 57–85. doi:10.1093/poq/nfw045

Eckman, S., & English, N. (2012). Creating housing unit frames from address databases. *Field Methods*, *24*(4), 399–408. doi:10.1177/1525822x12445141

Gatny, H. H., Couper, M. P., & Axinn, W. G. (2013). New strategies for biosample collection in population-based social research. *Social Science Research*, *42*(5), 1402–1409. doi:10.1016/j.ssresearch.2013.03.004

Health and Retirement Study. (2007). Diabetes study version 2.0, april 2007 (sensitive health data) data description and usage. Accessed on May 11, 2020. Retrieved from http://hrsonline.isr.umich.edu/modules/meta/diabetes/desc/diab2003dd.pdf

Herd, P., Schaeffer, N. C., DiLoreto, K., Jacques, K., Stevenson, J., Rey, F., & Roan, C. (2017). The influence of social conditions across the life course on the human gut microbiota: A pilot project with the Wisconsin Longitudinal Study. *The Journals of Gerontology: Series B*, *73*(1), 124–133. doi:10.1093/geronb/gbx029

Iannacchione, V. G. (2011). The changing role of address-based sampling in survey research. *Public Opinion Quarterly*, *3*, 556–575.

Information for Health Departments on Reporting Cases of COVID-19. (2020). Accessed on May 11, 2020. Centers for Disease Control and Prevention. Retrieved from https://www.cdc.gov/coronavirus/2019-ncov/php/reporting-pui.html

Jaszczak, A., Lundeen, K., & Smith, S. (2009). Using non-medically trained interviewers to collect biomeasures in a national in-home survey. *Field Methods*, *21*(1), 26–48. doi:10.1177/1525822x08323988

Key updates for week 17, ending April 25, 2020. COVIDView: A weekly surveillance summary of U.S. COVID-19 activity. (2020). Accessed on May 11, 2020. Centers for Disease Control and Prevention. Retrieved from https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html

Lindau, S. T., Hoffmann, J. N., Lundeen, K., Jaszczak, A., McClintock, M. K., & Jordan, J. A. (2009). Vaginal self-swab specimen collection in a home-based survey of older women: Methods and applications. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *64B*(Supplement 1), i106–i118. doi:10.1093/geronb/gbn021

Lunny, C., Taylor, D., Hoang, L., Wong, T., Gilbert, M., Lester, R., … Ogilvie, G. (2015). Self-collected versus clinician-collected sampling for Chlamydia and Gonorrhea screening: A systemic review and meta-analysis. *PLOS ONE*, *10*(7), e0132776. doi:10.1371/journal.pone.0132776

Lynn, P. (2019). Evaluating push-to-web methodology for mixed-mode surveys using address-based samples. *14*(1), 19–30. doi:10.18148/SRM/2020.v14i1.7591

NIH begins study to quantify undetected cases of coronavirus infection. (2020). National Institutes of Health. Retrieved from https://www.nih.gov/news-events/news-releases/nih-begins-study-quantify-undetected-cases-coronavirus-infection

Olson, K., & Smyth, J. D. [Jolene D]. (2017). Within-household selection in mail surveys. *Public Opinion Quarterly*, *81*(3), 688–713. doi:10.1093/poq/nfx025

Olson, K., & Smyth, J. D. [Jolene D.]. (2014). Accuracy of within-household selection in web and mail surveys of the general population. *Field Methods*, *26*(1), 56–69. doi:10.1177/1525822x13507865

Peytchev, A., Ridenhour, J., & Krotki, K. (2010). Differences between RDD telephone and ABS mail survey design: Coverage, unit nonresponse, and measurement error. *Journal of Health Communication*, *15*(sup3), 117–134. doi:10.1080/10810730.2010.525297

Pottinger, Y., & Sia, S. (2020). A rapidly evolving landscape: Coronavirus tests for diagnosis, surveillance and vaccine efficacy. ICAP Grand Rounds Webinar. Retrieved from https://icap.columbia.edu/icap-grand-rounds-webinar-a-rapidly-evolving-landscape-coronavirus-tests-for-diagnosis-surveillance-and-vaccine-efficacy/

Pramanik, R., Thompson, H., Kistler, J. O., Wade, W. G., Galloway, J., Peakman, T., & Proctor, G. B. (2012). Effects of the UK biobank collection protocol on potential biomarkers in saliva. *International Journal of Epidemiology*, *41*(6), 1786–1797. doi:10.1093/ije/dys166

Sakhi, A. K., Bastani, N. E., Ellingjord-Dale, M., Gundersen, T. E., Blomhoff, R., & Ursin, G. (2015). Feasibility of self-sampled dried blood spot and saliva samples sent by mail in a population-based study. *BMC Cancer*, *15*(1). doi:10.1186/s12885-015-1275-0

Sakshaug, J., Ofstedal, M. B., Guyer, H., & Beebe, T. (2014). 201. In T. P. Johnson (Ed.), *Handbook of health survey methods* (pp. 383–419). Wiley & Sons.

Sastry, N., Fomby, P., & McGonagle, K. (2017). Using the Panel Study of Income Dynamics (PSID) to conduct life course health development analysis. In *Handbook of life course health development* (pp. 579–599). doi:10.1007/978-3-319-47143-3_24

Spinelli, A., & Pellino, G. (2020). COVID-19 pandemic: Perspectives on an unfolding crisis. *British Journal of Surgery*. doi:10.1002/bjs.11627

Suzman, R. (2009). The national social life, health, and aging project: An introduction. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *64B*(Supplement 1), i5–i11. doi:10.1093/geronb/gbp078

The Boston Globe. (2020). Nearly a third of 200 blood samples taken in Chelsea show exposure to coronavirus. Boston Globe. Retrieved from https://www.bostonglobe.com/2020/04/17/business/nearly-third-200-blood-samples-taken-chelsea-show-exposure-coronavirus/

Walter, E. B., Dole, N., Siega-Riz, A. M., & Entwisle, B. (2011). The national children's study in North Carolina: A study of the effect of the environment on children's health, growth, and development. *North Carolina medical journal*, *72*(2), 160–164. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21721507

Woody, A., Hamilton, K., Livitz, I. E., Figueroa, W. S., & Zoccola, P. M. (2017). Buccal telomere length and its associations with cortisol, heart rate variability, heart rate, and blood pressure responses to an acute social evaluative stressor in college students. *Stress*, *20*(3), 249–257. doi:10.1080/10253890.2017.1328494

Commentary

The authors propose a mail survey design for conducting a probability-based assessment of COVID-19 antibody prevalence in the United States (US). According to the authors, current antibody prevalence estimates often rely on non-random samples collected in testing sites, clinics, or hospitals, where it is unlikely that those who show no or mild symptoms will be tested. The resulting estimates are therefore likely to be biased. In their view, a mail survey, in which self-administered antibody tests and an accompanying questionnaire are sent to a random sample of US residents, could address this issue. As a sampling plan, the authors propose the list of residential mailing addresses maintained by the US Postal Service. According to the authors, this design is more accurate than non-probability based approaches, and more cost-effective, more timely, and less risky for participants (as no interviewers would have to visit them at home) than comparable designs that involve in-person tests and interviews.

We agree that antibody testing based on large-scale, representative samples is essential to obtain reliable prevalence estimates. However, in its current form, the paper is lacking crucial information that is necessary to assess whether the proposed approach is a viable and cost-effective alternative to in-person designs.

First, address-based sampling has become increasingly popular in recent years, but the authors do not discuss its coverage problems and the potential bias that this might create. For example, existing research suggests that the coverage of urban areas based on residential mail addresses is better than that of more rural areas. Furthermore, depending on whether or not P.O. boxes are considered, under coverage or over coverage may occur in areas in which P.O. boxes are principal mode of mail delivery or are a frequent addition to regular mail addresses. Finally, people who live in non-institutional group accommodation (e.g., students who live in college dormitories) may be more difficult to sample with a mailing register than people who live on their own or with their family (Iannacchione, 2011)[1]. Hence, even though address-based sampling has many advantages, it is not completely bias free and researchers will need to address this bias in their study design.

Second, the authors assume that self-administered tests are generally available, without focusing on any specific test. This is a strong assumption, but even if we accept it, there are potential problems. The authors highlight that the accuracy of a correctly administered test (the rates at which false-positives and false-negatives occur) needs to be considered when calculating the required sample size to detect different prevalence rates, next to other factors, such survey return rates and the share of returned specimen that are viable. They also highlight that the survey package needs to contain detailed instructions to ensure that the sample is collected correctly. Yet, one important point that the authors do not consider is that even with such instructions, there might be systematic variation in the correct application across demographic groups, e.g., because of differences in language skills, or the ability to follow complex instructions. To the extent that such differences affect rates of false-positives or false-negatives, this might bias prevalence estimates across demographic groups.

Third, and most importantly, the authors do not provide a cost and risk calculation related to their approach. This makes it difficult to evaluate its benefits compared to other approaches. From a cost perspective, the fact that mail surveys can dispense of interviewers arguably reduces their costs compared to designs that employ interviewers, ceteris paribus. Yet, some of the challenges that the authors highlight themselves might increase certain cost aspects of mail surveys. For example, (i) special modes of transport might be necessary to ensure that the collected specimen remain viable in the mail; (ii) there might be a need for multiple at-home tests, to increase test accuracy; (iii) non-response may lead to a loss of unused tests. Depending on the costs of individual testing kits, non-response might be of particular concern, if we consider that response rates in general surveys are typically well below 50%. These issues do not exist, or are of lesser concern, when in-person interviews are conducted. Interviewers can more easily ensure the proper storage and transport of samples, ensure that the tests are applied properly, and can use tests that have not been used in one household in a different household. From a risk perspec-

[1]The references are listed among the references for the main article

tive, we agree that minimizing the need to interact with other people who could carry an infectious disease is desirable, but the authors do not discus how much bigger the infection risk would be in a carefully conducted in-person seroprevalence survey. Such surveys are currently being planned, or have taken place already, in several countries, and a discussion of the risks (and costs) that such an approach involves (assuming, e.g., a similar population-based approach as the authors suggest) would have been instructive to researchers who are pondering which design to choose.

André Grow, Daniela Perrotta, Emanuele Del Fava, and Jorge Cimentada
Max Planck Institute for Demographic Research, Germany

Reply to Grow, Perotta, Del Fava and Cimentada

The reviewers from the Max Planck Institute for Demographic Research in Germany provided us with very thoughtful consideration and critique of the manuscript. We briefly address their points below.

The reviewers pointed out "...[ABS] is not completely bias free and researchers will need to address this bias in their study design." The reviewers are correct that the ABS frame excludes those who live in institutions, such as nursing homes or jails, or in hidden apartments, as well as those who are homeless. These populations are vulnerable to SARS-CoV-2 infection and their exclusion likely biases the study's estimate of the prevalence downward. We have designed this study to estimate the prevalence of antibodies in the U.S. household population and any results published would highlight this limitation.

The reviewers also indicated that the paper lacked discussion of a specific antibody test. We purposefully did not address this point in the original or revised manuscript because the universe of available and approved tests for antibodies is continually changing. Additionally, the tests currently available have lower than optimal sensitivity and specificity rates. The development of adequate serology tests is an area of focus internationally, in addition to developing vaccines and treatments. Suggesting a specific test would date the information provided in this manuscript.

The comments also raised the possibility of bias resulting from respondents not correctly carrying out the at-home testing procedure. This is a valid point and would be a risk in the self-administered test. The risk could be minimized, as suggested, by providing clear and direct instructions, supplying help via a toll-free help line, and providing the option to collect the sample while a trained interviewer or staff person is on a video call with the respondent. These tactics would all help minimize error in collection.

In the revised manuscript, we have added some text to address the reviewers' concern about the lack of discussion of the costs of the study.

The reviewers' thoughtful observations point to the complexity of such a study. These points should be carefully considered before such a study is undertaken.

Alicia M. Frasier, Heidi Guyer, Laura DiGrande, Rose Domanico, Darryl Cooney, and Stephanie Eckman