



# Estimating seroprevalence of SARS-CoV-2 in Ohio: A Bayesian multilevel poststratification approach with multiple diagnostic tests

David Kline<sup>a,1,2</sup>, Zehang Li<sup>b,1</sup>, Yue Chu<sup>c</sup>, Jon Wakefield<sup>d,e,1</sup>, William C. Miller<sup>f</sup>, Abigail Norris Turner<sup>g</sup>, and Samuel J. Clark<sup>c,1,2</sup>

<sup>a</sup>Center for Biostatistics, Department of Biomedical Informatics, The Ohio State University, Columbus, OH 43210; <sup>b</sup>Department of Statistics, University of California, Santa Cruz, CA 95064; <sup>c</sup>Department of Sociology, The Ohio State University, Columbus, OH 43210; <sup>d</sup>Department of Statistics, University of Washington, Seattle, WA 98195; <sup>e</sup>Department of Biostatistics, University of Washington, Seattle, WA 98195; <sup>f</sup>Division of Epidemiology, College of Public Health, The Ohio State University, Columbus, OH 43210; and <sup>g</sup>Division of Infectious Diseases, College of Medicine, The Ohio State University, Columbus, OH 43210

Edited by Adrian E. Raftery, University of Washington, Seattle, WA, and approved May 18, 2021 (received for review December 10, 2020)

Globally, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has infected more than 59 million people and killed more than 1.39 million. Designing and monitoring interventions to slow and stop the spread of the virus require knowledge of how many people have been and are currently infected, where they live, and how they interact. The first step is an accurate assessment of the population prevalence of past infections. There are very few population-representative prevalence studies of SARS-CoV-2 infections, and only two states in the United States—Indiana and Connecticut—have reported probability-based sample surveys that characterize statewide prevalence of SARS-CoV-2. One of the difficulties is the fact that tests to detect and characterize SARS-CoV-2 coronavirus antibodies are new, are not well characterized, and generally function poorly. During July 2020, a survey representing all adults in the state of Ohio in the United States collected serum samples and information on protective behavior related to SARS-CoV-2 and coronavirus disease 2019 (COVID-19). Several features of the survey make it difficult to estimate past prevalence: 1) a low response rate; 2) a very low number of positive cases; and 3) the fact that multiple poor-quality serological tests were used to detect SARS-CoV-2 antibodies. We describe a Bayesian approach for analyzing the biomarker data that simultaneously addresses these challenges and characterizes the potential effect of selective response. The model does not require survey sample weights; accounts for multiple imperfect antibody test results; and characterizes uncertainty related to the sample survey and the multiple imperfect, potentially correlated tests.

coronavirus | COVID-19 | imperfect diagnostic tests | SARS-CoV-2 | seroprevalence survey

Slowing or stopping the spread of a new virus for which a vaccine does not exist starts with two key pieces of information. One is what fraction of the population has been infected and is thereby potentially less susceptible or even immune to future infection; two is what fraction of the population is currently infected and potentially infectious to others. Together with a basic understanding of the infection process, this information roughly characterizes the potential for the epidemic to grow. However, in the early months of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic, there was great uncertainty about the number of people who had been infected due to challenges with adequate testing and asymptomatic and mild cases. Due to this uncertainty, debate occurred over whether the past infection prevalence was far higher than recorded cases suggested, even potentially approaching herd immunity thresholds, believed to be 70 to 85%. A population-based estimate of past infection prevalence was of critical importance to public health officials and policy makers who have the

responsibility to manage the epidemic, make policy, and protect the public.

As of this writing (late November 2020), the global SARS-CoV-2 pandemic has infected more than 59 million people, and coronavirus disease 2019 has killed more than 1.39 million (1). Basic epidemiological information to describe the pandemic is scarce because the virus is new and the pandemic exploded rapidly. In its place is a wide variety of indicators based on convenient, mostly nonrepresentative, or indirectly related data—counts of all-cause deaths (e.g., refs. 2–4), facility-based testing results for symptomatic patients, nonrepresentative samples, and in many situations, results from inadequately characterized tests that perform poorly (5). Franceschi et al. (6) identify 37 SARS-CoV-2 prevalence studies from 19 countries. Most present results are from nonrepresentative, otherwise special, or very small study populations. Just 14 represent large-enough populations to be of policy interest—national or state level—and use a probability-based sample from a credible sampling frame to

## Significance

In July 2020, there was great uncertainty around the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Despite its vital importance for public health policy, knowledge about the cumulative incidence of past infections was limited by challenges with diagnostic testing and the presence of mild or asymptomatic cases. Within this environment, competing narratives emerged around the prevalence of past SARS-CoV-2 infections, which would have had differing policy implications. To address this, in July 2020 a population-representative household survey collected serum for SARS-CoV-2 antibody detection in Ohio in the United States. This study describes a Bayesian statistical method developed to estimate the population prevalence of past infections accounting for the low positive rate; multiple imperfect diagnostic tests; and nonignorable nonresponse.

Author contributions: W.C.M., A.N.T., and S.J.C. designed research; D.K., A.N.T., and S.J.C. performed research; D.K., Z.L., Y.C., J.W., and S.J.C. contributed new reagents/analytic tools; D.K. and Y.C. analyzed data; and D.K., Z.L., Y.C., J.W., W.C.M., A.N.T., and S.J.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

<sup>1</sup>D.K., Z.L., J.W., and S.J.C. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: david.kline@osumc.edu or clark.2962@osu.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2023947118/-/DCSupplemental>.

Published June 25, 2021.

produce results that could represent the population of interest—Asia: one (7); Europe: seven (8–14); North America: two (15, 16); and South America: four (17–20). The two in North America are from the state of Indiana (15) and the state of Connecticut (16) in the United States.

Conducting population-representative biomarker surveys is difficult—particularly in the United States. Good sampling frames exist in a variety of forms (tax rolls, telephone numbers, etc.), but recruiting willing respondents is exceptionally difficult and likely affected by selection relative to the outcome of interest. Both of the studies in the United States had low response rates for the full interview with valid test results for SARS-CoV-2—Indiana: 23.4% and Connecticut: 7.8%. Further complicating analysis, there were few positive tests among those who did respond—Indiana:  $47/3,658 = 1.3\%$  for the PCR test of current infection and  $38/3,658 = 1.0\%$  for antibody test of ever infected and Connecticut:  $23/567 = 4.1\%$  for the antibody test of ever infected. Both studies described concern that the nonresponding participants were likely to be at higher risk of infection with SARS-CoV-2. Finally, like all SARS-CoV-2 immunology investigations to date, both studies struggled with poor-quality antibody tests whose unfavorable performance characteristics were not well understood; ref. 7 has an overview of these issues.

Statistical analysis of data like these is difficult. First, the low response rate requires extensive recalibration of the sampling weights, and in the worst case, there may be sampling units with no respondents at all. Second, the very small number of positive cases pushes the asymptotic (large-sample) assumptions of frequentist methods to their limits and can break them. Third, the imperfect and poorly characterized antibody tests potentially add a lot of uncertainty that must be reflected in the results, particularly in low-prevalence settings (21). Fourth, when results from multiple tests with different performance characteristics are combined, the joint result must be accurately described and its uncertainty propagated to the final estimate of prevalence—importantly, including the possibility that results from individual tests are correlated. Finally, if there is selection on the outcome, then the effect of this must be understood. In our review of the literature, we did not find an existing method that addresses all of these challenges in a unified way.

Here, we describe an analytical approach developed to produce estimates of past infection with SARS-CoV-2 using data from a probability-based household survey representing adults in the state of Ohio in the United States. Like the SARS-CoV-2 prevalence studies in Indiana and Connecticut, the Ohio survey had a low response rate, few positive cases, and the possibility of selective response. Additionally, the Ohio survey used multiple imperfect antibody tests for the same antibodies, resulting in the need to quantify uncertainty in the joint result and account for possible dependence among results.

To overcome these challenges, we weave together two well-established modeling frameworks into a single coherent approach. We utilize the literature on modeling multiple imperfect diagnostic tests through the use of a Bayesian latent class model (e.g., refs. 22 and 23). This enables us to combine information across tests to infer the true latent infection status of a participant while incorporating uncertainty about the characteristics of the tests. We use the latent infection status to generate model-based estimates of the population prevalence using multilevel regression and poststratification (21, 24). These approaches are integrated into a single Bayesian model that allows for the full propagation of uncertainty, exact inferences, and the ability to specify informative priors using external information. By doing so, we produce estimates that reflect all available information and uncertainty.

## Methods

The purpose of this study is to estimate the prevalence of past SARS-CoV-2 infections in the state of Ohio using three separate antibody tests given to randomly selected adult participants. We know that each antibody test is imperfect, and there is no gold standard for detecting prior SARS-CoV-2 infection. Prevalence estimates based on a single imperfect test are always biased but particularly in the case of SARS-CoV-2 infection rates, which are low (21). To mitigate that bias and incorporate variability due to error in the testing results, we will take a Bayesian latent class approach for modeling multiple diagnostic tests. Our approach will be based on combining a fixed effects framework for modeling conditional dependence across multiple diagnostic tests (22, 23) with a model-based analysis using multilevel regression and poststratification (24) to acknowledge the complex design aspects of the survey.

**Survey Design.** The survey was designed to provide policy makers with a “quick” overall snapshot of the prevalence of prior infection at the state level. The survey sampling scheme was designed as a stratified two-stage cluster sample. Strata were defined by eight administrative regions used by the state. Within each region, 30 census tracts were randomly selected with probability proportional to size (PPS) based on total population size. Then, within a selected tract, five households were randomly selected, and one adult (at least 18 y of age) within each household was randomly selected to participate in the study from all eligible adults in the household using a random number generator. Adults were eligible for the study if they were aged 18 y or older, slept in the home at least 4 d of the last 7 d, were proficient in English or Spanish, were willing to provide blood and nasopharyngeal swab samples, and were able to provide consent. If the selected individual declined to participate, the team moved to the next selected household. Thus, the planned target sample size was 1,200 participants. The study was conducted from 9 to 28 July 2020. The Ohio Department of Health (ODH) Institutional Review Board (IRB) reviewed and approved the research. The Ohio State University IRB ceded review to the ODH IRB. All participants provided written informed consent.

Each participant in the study provided biological samples, which would be put through a series of diagnostic tests. In this paper, we will focus on prior infection as determined by the presence of SARS-CoV-2 antibodies. Since antibody tests for SARS-CoV-2 are new to the market and of varying quality, we processed participant samples using three different antibody tests. Specifically, we used the Abbott Immunoglobulin G (IgG), Liaison IgG, and Epitepe Immunoglobulin M (IgM) tests.

**Model.** For each participant in the study, indexed by  $i = 1, \dots, n$ , let  $T_i = (T_{i1}, T_{i2}, T_{i3})$  be indicators of a positive test result of the Abbott IgG, Liaison IgG, and Epitepe IgM, respectively. Let  $D_i$  be the unobserved indicator of whether participant  $i$  had a prior infection of SARS-CoV-2. This latent indicator of prior infection is our primary outcome of interest. Analysis methods for multiple diagnostic tests without a gold standard hinge on assumptions related to conditional independence (23). We will assume that  $(T_{i1}, T_{i2})$  and  $T_{i3}$  are independent given the true infection status. This implies that conditional on the true presence of prior infection, we assume tests for the same antibody are dependent and tests for different antibodies are independent. Based on the underlying design of the tests and what they target, we believe these assumptions are reasonable for this specific application.

Given the assumptions stated above, we can consider this a problem with two conditionally independent sets of tests: the two IgG tests and the IgM test. Thus, we can decompose the joint probability as

$$\Pr(T_{i1}, T_{i2}, T_{i3}|D_i) = \Pr(T_{i1}, T_{i2}|D_i) \Pr(T_{i3}|D_i).$$

By doing so, each conditional probability on the right-hand side can be estimated following a fixed effects approach (22). Since each test result is binary, this leads to eight potential combinations of results, shown in Table 1, and suggests the following distribution:

$$Y_i|D_i \sim \text{Multinomial}(1, \mathbf{p}_i),$$

where  $Y_i$  is an indicator of participant  $i$ 's result pattern and  $\mathbf{p}_i = \mathbf{p}_i(D_i)$  is a vector of length eight where each element is the probability of a result pattern.

To construct the probability vector  $\mathbf{p}_i$ , we first need to define the conditional probabilities within each antibody. Let  $S_j$  be the sensitivity and  $C_j$  be the specificity of test  $j$ . Since we have two IgG tests ( $j = 1, 2$ ), we allow for

**Table 1. Mapping from the three binary test results to the multinomial response vector and corresponding probabilities**

$Y_j$	$T_{i1}$	$T_{i2}$	$T_{i3}$	$p_i$
1	1	1	1	$p_{i1}$
2	1	1	0	$p_{i2}$
3	1	0	1	$p_{i3}$
4	0	1	1	$p_{i4}$
5	1	0	0	$p_{i5}$
6	0	1	0	$p_{i6}$
7	0	0	1	$p_{i7}$
8	0	0	0	$p_{i8}$

their results to be correlated. Let  $R_{12}^1$  be the covariance between the results of tests 1 and 2 given the infection status is positive and  $R_{12}^0$  be the covariance when the infection status is negative. Then, for the joint probabilities of the IgG test results (22), we have

$$\begin{aligned} \Pr(T_{i1} = t_{i1}, T_{i2} = t_{i2} | D_i = 1) &= & [1] \\ S_1^{t_{i1}} S_2^{t_{i2}} (1 - S_1)^{1-t_{i1}} (1 - S_2)^{1-t_{i2}} + (-1)^{t_{i1}+t_{i2}} R_{12}^1 \\ \Pr(T_{i1} = t_{i1}, T_{i2} = t_{i2} | D_i = 0) &= \\ C_1^{1-t_{i1}} C_2^{1-t_{i2}} (1 - C_1)^{t_{i1}} (1 - C_2)^{t_{i2}} + (-1)^{t_{i1}+t_{i2}} R_{12}^0. \end{aligned}$$

Since we only have one IgM test ( $j = 3$ ), we have

$$\begin{aligned} \Pr(T_{i3} = t_{i3} | D_i = d_i) &= & [2] \\ d_i S_3^{t_{i3}} (1 - S_3)^{1-t_{i3}} + (1 - d_i) C_3^{1-t_{i3}} (1 - C_3)^{t_{i3}}. \end{aligned}$$

We then use these conditional probabilities to construct the probabilities in the vector  $p_i$  in the multinomial distribution above. Those eight probabilities can be calculated with the following general equation for each test result pattern:

$$\begin{aligned} \Pr(T_{i1} = t_{i1}, T_{i2} = t_{i2}, T_{i3} = t_{i3} | D_i = d_i) &= \\ \Pr(t_{i1}, t_{i2} | d_i) \Pr(t_{i3} | d_i). \end{aligned}$$

These probabilities are the individual elements of the vector  $p_i$ , which is a function of  $D_i, S, C, R$ , where  $S = (S_1, S_2, S_3)$ ,  $C = (C_1, C_2, C_3)$ , and  $R = (R_{12}^1, R_{12}^0)$ .

The latent true infection status is the primary process of interest as we are interested in estimating the prevalence of prior SARS-CoV-2 infections. We assume

$$D_i | \pi_i \sim \text{Bernoulli}(\pi_i),$$

where  $\pi_i$  is the probability of prior infection for participant  $i$ , which will be determined by region, strata, and census tract. Specifically, we let  $r[i]$ ,  $s[i]$ , and  $t[i]$  refer to the region, strata, and census tract for participant  $i$ , respectively, and assume

$$\text{logit}(\pi_i) = \alpha_{r[i]} + \mathbf{X}_{s[i]} \beta^s + \mathbf{X}_{t[i]} \beta^t + b_{t[i]}, \quad [3]$$

where  $\alpha_r$  is a region-specific random intercept,  $\mathbf{X}_s$  is a vector of stratum indicators with fixed effects vector  $\beta^s$ ,  $\mathbf{X}_t$  is a vector of census tract covariates with fixed effects vector  $\beta^t$ , and  $b_t$  is a random effect for census tract. In this analysis,  $\mathbf{X}_s$  includes indicators of participant age group and sex, and  $\mathbf{X}_t$  is a scalar and corresponds to the log of the total census tract population. We code the groups using sum to zero contrasts, and the log population was standardized to have zero mean and SD of one across the entire state. We assume  $\alpha_r \stackrel{iid}{\sim} N(\alpha, \sigma^2)$ . Since the fixed effects are coded to sum to zero,  $\alpha$  reflects the overall mean on the logit scale, and  $\alpha_r$  can be interpreted as regional means. We also assume  $b_t \stackrel{iid}{\sim} N(0, \tau^2)$ , which accounts for correlation between individuals in the same census tract.

At this point, we link the diagnostic test model to a multilevel regression and poststratification approach (21, 24) for estimating population prevalence. In Eq. 3, we specified a multilevel logistic regression model for the probability of prior infection. Since true infection was rare in our sample, we were unable to fit a saturated model. Instead, we chose to use a hierarchical model with fixed strata effects. By doing so, we assume that while regions may have different probabilities of infection, the relative ordering of the strata will be the same across regions (i.e., there is no

interaction between region and strata). Effects for region and census tract population are included to account for the characteristics of the underlying survey design (stratification and PPS sampling), making the selection process ignorable (25).

To obtain the population prevalence, we calculate

$$\pi = \frac{\sum_r \sum_s \sum_t \pi_{rst} P_{rst}}{\sum_r \sum_s \sum_t P_{rst}},$$

where  $\pi_{rst}$  is the prevalence and  $P_{rst}$  is the adult population in stratum  $s$  in census tract  $t$  in region  $r$ . The prevalence contribution for region  $r$ , strata  $s$ , and tract  $t$  is

$$\pi_{rst} = \text{expit}(\alpha_r + \mathbf{X}_s \beta^s + \mathbf{X}_t \beta^t + b_t),$$

where  $b_t \stackrel{iid}{\sim} N(0, \tau^2)$ .

Since we are fitting the model in the Bayesian paradigm, we must specify prior distributions on all unknown parameters. The main reason we chose a fixed effects model for the test results was because it is directly determined by test sensitivity and specificity. This allows us to transparently incorporate prior information using the validation data on the package insert (26–28) (as of October 2020) for each test. Each validation study examined test performance on a set of patients with known past infection status by the test result. Using  $\text{Beta}(\alpha, \beta)$  distributions, the mean is  $\alpha/(\alpha + \beta)$ , and the variance is scaled by  $(\alpha + \beta)$ . Commonly, Beta prior distributions are parameterized as hypothetical binomial experiments with  $\alpha$  successes of  $\alpha + \beta$  trials. Here, rather than using a hypothetical experiment, we can directly use the validation data from the  $2 \times 2$  tables on the package inserts. For sensitivity,  $\alpha$  is the number of true positives, and  $\beta$  is the number of false negatives. For specificity,  $\alpha$  is the number of true negatives, and  $\beta$  is the number of false positives. For the Epitope IgM test, there were no false positives, so  $\beta$  was set to 0.1 since it must be greater than 0. This sets the mean of each distribution at the observed value with variability according to the size of the validation study. Based on this information, we let

$$\begin{aligned} S_1 &\sim \text{Beta}(109, 13) & C_1 &\sim \text{Beta}(1066, 4) \\ S_2 &\sim \text{Beta}(96, 39) & C_2 &\sim \text{Beta}(1074, 16) \\ S_3 &\sim \text{Beta}(9, 11) & C_3 &\sim \text{Beta}(54, 0.1). \end{aligned}$$

This implies that the means (variances) of the prior distributions are  $S_1$ , 0.893 (0.0008);  $C_1$ , 0.996 (0.000003);  $S_2$ , 0.711 (0.002);  $C_2$ , 0.985 (0.00001);  $S_3$ , 0.450 (0.01); and  $C_3$ , 0.998 (0.00003). The prior densities are also shown in Fig. 1. The remaining parameters for the diagnostic testing part of the model are the covariance parameters. We enforce necessary constraints on these parameters by specifying independent uniform prior distributions for each parameter restricted to its allowable range (22). Assuming only positive dependence between tests, we have

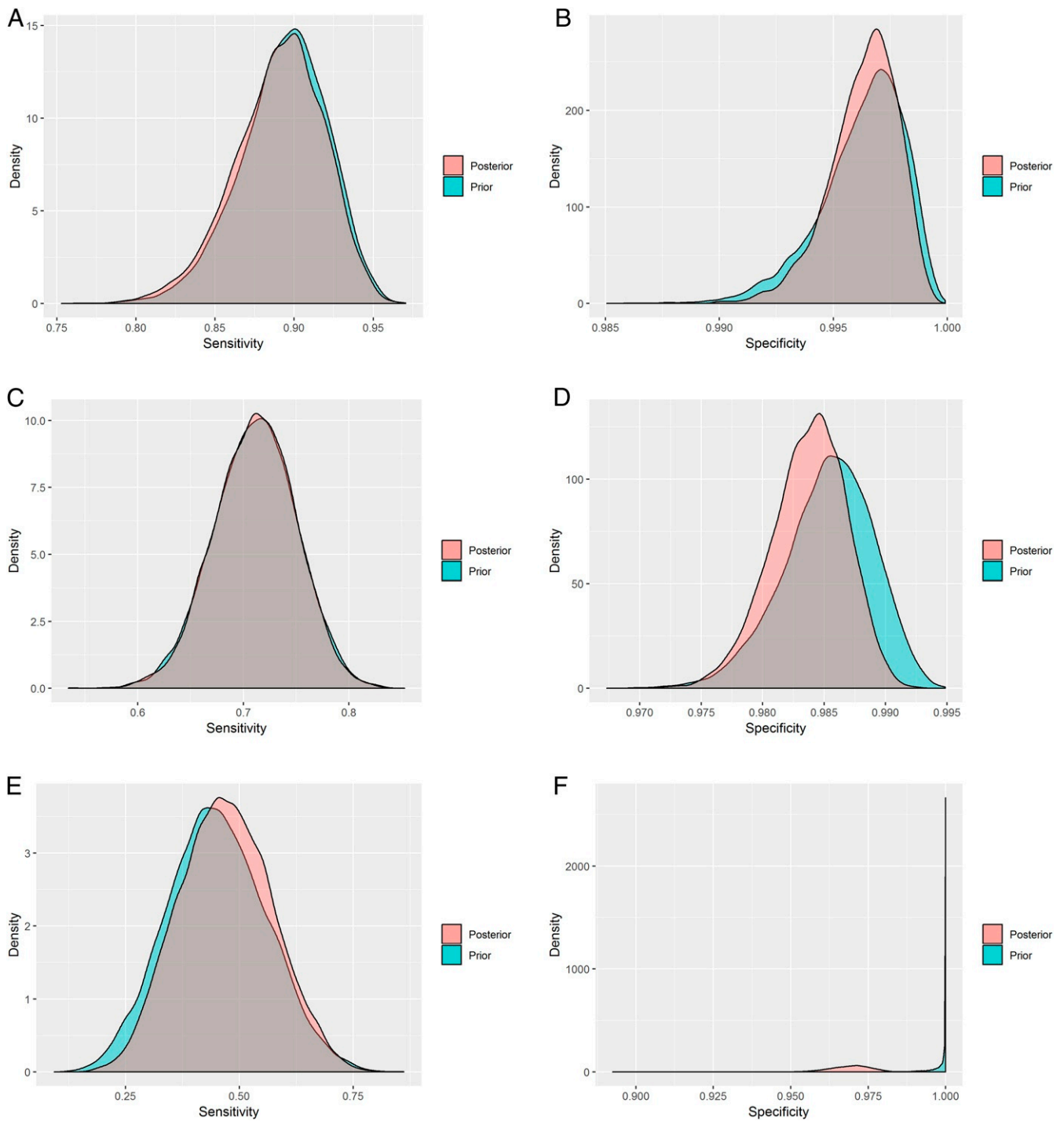
$$\begin{aligned} R_{12}^1 &\sim \text{Uniform}(0, \min(S_1, S_2) - S_1 S_2) \\ R_{12}^0 &\sim \text{Uniform}(0, \min(C_1, C_2) - C_1 C_2). \end{aligned}$$

For the multilevel regression part of the model, we assume each element of  $\beta$  is independently normally distributed with zero mean and variance of nine. We assume  $\tau$  and  $\sigma$  have independent uniform distributions on (0, 5). We assume  $\alpha \sim N(\text{logit}(0.03), 1)$  to reflect prior belief that the prevalence of prior infection is around 3% and puts 95% of the probability between 0.4 and 18.0%.

By putting everything together, we have the full model:

$$\begin{aligned} p(\mathbf{S}, \mathbf{C}, \mathbf{R}, \mathbf{D}, \alpha, \beta, \mathbf{b}, \tau^2, \alpha, \sigma^2 | \mathbf{T}) &\propto \\ \prod_{i=1}^n \Pr(T_i | \mathbf{S}, \mathbf{C}, \mathbf{R}, D_i) &\times p(\mathbf{R} | \mathbf{S}, \mathbf{C}) \times p(\mathbf{S}, \mathbf{C}) \\ &\times \prod_{i=1}^n \Pr(D_i | \alpha, \beta, \mathbf{b}) \times p(\mathbf{b} | \tau^2) \times p(\alpha | \alpha, \sigma^2) \\ &\times p(\beta) p(\alpha) p(\tau^2) p(\sigma^2), \end{aligned}$$

where  $\mathbf{T} = (T_1, \dots, T_n)$ ,  $\mathbf{S} = (S_1, S_2, S_3)$ ,  $\mathbf{C} = (C_1, C_2, C_3)$ ,  $\mathbf{R} = (R_{12}^1, R_{12}^0)$ ,  $\mathbf{D} = (D_1, \dots, D_n)$ ,  $\beta = (\beta^s, \beta^t)$ ,  $\alpha$  is the vector of region-specific intercepts, and  $\mathbf{b}$  is the vector of census tract random effects. To compute the posterior



**Fig. 1.** Density plots of the prior and posterior distributions for the sensitivity and specificity of each antibody test. (A) Sensitivity of Abbott IgG. (B) Specificity of Abbott IgG. (C) Sensitivity of Liaison IgG. (D) Specificity of Liaison IgG. (E) Sensitivity of Epitope IgM. (F) Specificity of Epitope IgM.

distribution, the model was fit using a Markov Chain Monte Carlo algorithm implemented in R (29) using NIMBLE (30). The algorithm was run for 500,000 iterations, discarding the first 250,000 as burn-in and thinning the remaining iterations by keeping every 20th draw. Convergence was assessed by visually inspecting trace plots. Posterior distributions are summarized by the posterior mean and 95% highest posterior density credible interval. Code is available online at <https://github.com/sinafala/bayes-prevalence>.

**Missing Test Results.** Some participants did not have results for all three tests which was primarily due to an insufficient amount of sample

to run the test. There were also 17 participants with an inconclusive result for the IgM test, which we considered equivalent to not having a result. However, since our primary interest is in the latent infection status of the participant, we can still gain some information about this from the test results that are available. Test result pattern probabilities would still be calculated as above corresponding to the test or pair of tests available for the participant. For example, if a participant only had the IgG test results, there would be four patterns of results that would be defined by the probabilities in Eq. 1. If IgG and IgM tests are observed, the four probabilities would be computed from products of

$$\Pr(T_{ij} = t_{ij} | D_i = d_i) = \tag{4}$$

$$d_i S_j^{t_{ij}} (1 - S_j)^{1-t_{ij}} + (1 - d_i) C_j^{1-t_{ij}} (1 - C_j)^{t_{ij}}$$

A single available test result would simply lead to a Bernoulli distribution of whether the result was positive ( $t_{ij} = 1$ ) with the probability calculated as in Eq. 4 with the appropriate parameters for the observed test.

**Sensitivity Analyses.** For the analysis of any survey, it is important to account for the potential of nonresponse bias. We note that our primary analysis is valid assuming a nonresponse mechanism that is ignorable given stratum. This is likely a strong assumption, and so, we will consider sensitivity analyses under several assumptions of potential nonignorable scenarios. To do so, assume

$$\pi_{rst} = \pi_{rst}^R (1 - \rho_{rst}^N) + \pi_{rst}^N \rho_{rst}^N \tag{5}$$

where  $\pi_{rst}^R$  and  $\pi_{rst}^N$  are the prevalence rates among the responders and non-responders, respectively, and  $\rho_{rst}^N$  is the probability of nonresponse. Note that when ignorability is satisfied, we have  $\pi_{rst} = \pi_{rst}^R = \pi_{rst}^N$ , which results in the estimates from the primary analysis. Due to the survey design, we can only obtain household nonresponse rates by region, and so, we assume  $\rho_{rst}^N = \rho_r^N$  for all strata and tracts. In order to carry out a sensitivity analysis, we will assume the prevalence among nonresponders is

$$\pi_{rst}^N = \lambda \pi_{rst}^R \tag{6}$$

where  $\lambda$  is the prevalence ratio in nonresponders compared with responders and  $\pi_{rst}^R$  is the prevalence estimated from the logistic regression described above. We will vary  $\lambda$  to explore different scenarios and assess the sensitivity of our estimates to changes in the prevalence in nonresponders.

We conduct a second sensitivity analysis to assess the impact of using informative prior distributions on the test sensitivity parameters. This is important because of the limited validation data in individuals with confirmed infection and questions surrounding the detectability of antibodies over time, which would impact the sensitivity of the test when used in a population survey such as this. For this analysis, we fit the model as described above but use Beta(1, 1) prior distributions for  $S_1$ ,  $S_2$ , and  $S_3$ . This is a uniform distribution on (0, 1) and serves as a less informative prior distribution. Finally, we conduct a third sensitivity analysis that analyzes the IgG and IgM tests separately. Again, since antibody presence varies with unknown time since infection, these estimates reflect the prevalence of individuals in the postinfection period when IgG or IgM antibodies are detectable and together, can serve as a reasonable upper bound for prior infection prevalence.

## Results

A total of 727 adults participated in the survey. To be included in the analysis, participants had to have at least one antibody test result and have age and sex recorded. Of the 727 participants, 667 (92%) were included in the analysis. Characteristics of

those included in the analysis are shown in Table 2. We observe that our included participants tend to be older and female. We also observe that 23.4% of included participants were from the northwest region of Ohio.

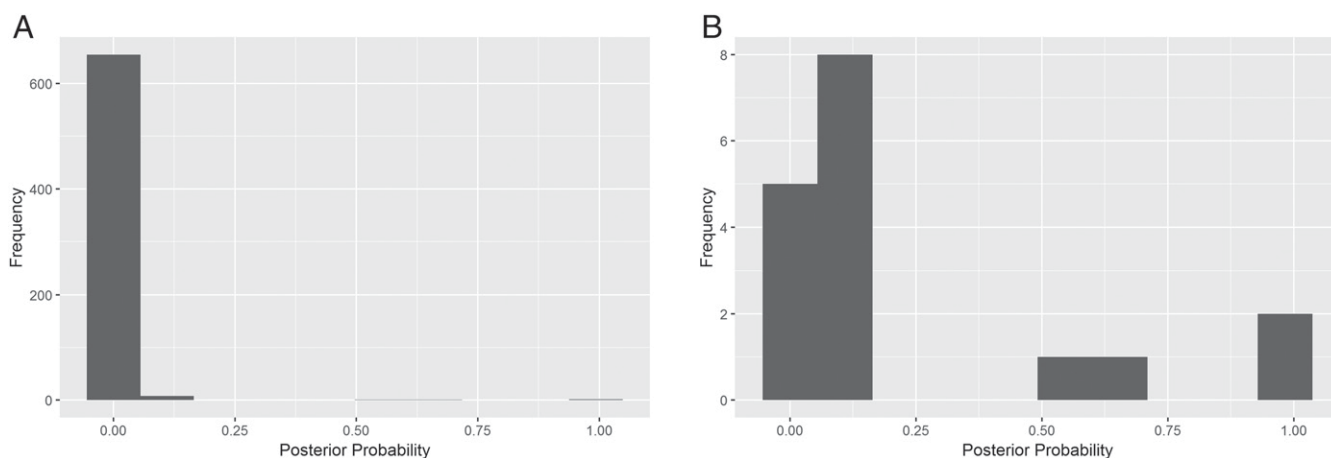
Our primary goal is to estimate the statewide prevalence of prior infection with SARS-CoV-2. Based on our model, the posterior mean prevalence is 1.3% with a 95% credible interval of (0.2, 2.7%). This corresponds to approximately 118,000 noninstitutionalized adults with a 95% credible interval of (22,000, 240,000). As noted in *Methods*, the prevalence estimates are based on the latent infection status estimates inferred from the antibody test results. In Fig. 2, we show the posterior mean probability of prior infection for all 667 participants included in the analysis. This illustrates that there were very few participants in the study who were estimated to actually have prior infection. In fact, only 17 (2.5%) participants had a posterior probability of past infection greater than 1%. As expected, those with the highest estimated probabilities were those for whom there was agreement across the tests. In general, there was little observed agreement across the three diagnostic tests. Of the 39 participants with at least one positive result, only 3 (7.7%) had a positive result on more than one test. For parameters that had informative prior distributions, Fig. 1 and *SI Appendix, Fig. S1* show prior and posterior densities to illustrate learning based on the observed data.

We now turn to the sensitivity analysis that accounted for potential nonignorable nonresponse. The household-level nonresponse rates by region are shown in Table 3. We believe that nonresponders were most likely to have a higher prevalence of past infection than responders. In Fig. 3, we show that if we assume that the prevalence of past infection in nonresponders was three times that of responders, the upper bound of the 95% credible interval is at a prevalence of 7%. Thus, across the range of reasonable scenarios that we considered, we observe rates of past infection that do not dramatically differ from the estimates in our primary analysis.

Finally, we consider the sensitivity analysis using less informative prior distributions for the test sensitivity parameters. In *SI Appendix, Fig. S2*, we show posterior mean and credible intervals across prevalence ratios of nonresponders to responders. For the primary analysis, the posterior mean prevalence is 5.2% with 95% credible interval of (0.2, 14.8%). In *SI Appendix, Figs. S3 and S4*, we show prior and posterior densities and see that this analysis estimates the test sensitivities to be much lower than what was suggested by the validation data, leading to higher estimates of prevalence. We show estimates using only the IgG test results in *SI Appendix, Fig. S5* and only the IgM test result in

**Table 2. Descriptive statistics and counts of positive antibody test results for participants with at least one antibody test result who were included in the analysis of prior infection prevalence**

Variable	Count	Proportion	All three tests	Abbott IgG and Liaison IgG	Abbott IgG and Epitope IgM	Liaison IgG and Epitope IgM	Abbott IgG only	Liaison IgG only	Epitope IgM only
Age 18–44	175	0.262	1	0	0	0	1	2	4
Age 45–64	239	0.358	0	1	0	0	0	2	7
Age 65 and over	253	0.379	1	0	0	0	2	7	11
Male	275	0.412	2	0	0	0	1	5	7
Female	392	0.588	0	1	0	0	2	6	15
Central	82	0.123	0	0	0	0	0	1	3
East central	90	0.135	1	1	0	0	1	1	4
Northeast	75	0.112	0	0	0	0	0	1	3
Northwest	156	0.234	0	0	0	0	2	2	4
Southeast	77	0.115	0	0	0	0	0	0	1
Southeast central	56	0.084	0	0	0	0	0	1	3
Southwest	65	0.097	0	0	0	0	0	5	1
West central	66	0.099	1	0	0	0	0	0	3
Total	667	1.000	2	1	0	0	3	11	22



**Fig. 2.** Posterior probability of past infection of SARS-CoV-2 for all participants and limited to those participants with a probability of greater than 1%. (A) Posterior probability of infection. (B) Posterior probability of infection greater than 1%.

*SI Appendix, Fig. S6.* The posterior mean prevalence is 1.3% with 95% credible interval of (0.2, 2.6%) using the IgG tests and is 7.8% with 95% credible interval of (1.0, 18.5%) using the IgM test. However, all of the estimates in the sensitivity analyses would be aligned with a similar public health and policy response as they are not compatible with having reached herd immunity.

## Discussion

In this paper, we present an approach to coherently integrate multiple imperfect diagnostic tests and a model-based analysis. By using this approach, we are able to estimate the past prevalence of SARS-CoV-2 infection in the state of Ohio in the United States while appropriately accounting for uncertainty in multiple antibody tests and leveraging the strengths of a designed survey. Through the Bayesian paradigm, we are also able to incorporate external information through informative prior distributions and provide exact inferences for the prevalence estimates. Our approach provides policy makers with the best possible estimates of prior infection given the survey results and our current knowledge of the quality of the antibody tests.

Through our study, we estimate the magnitude of prior SARS-CoV-2 infection to be low, roughly 118,000 noninstitutionalized adults. However, this still reflects a higher burden of infection than the reported number of cases during the 3-mo period when we would anticipate being able to detect antibodies (31–35)—approximately 62,000 cases from mid-April to mid-July 2020. Our sensitivity analysis for nonresponse shows that even with a prevalence ratio of three between nonresponders and responders, the upper bound of the credible interval is a prevalence of about 7%. Our sensitivity analysis for test sensitivity shows slightly increased estimates of prevalence but with considerable uncertainty. Similar results are also observed in sensitivity analyses that separately consider each type of antibody. Thus, even in hypothetical scenarios with large selective response and low estimated test sensitivity, we conclude that the majority of Ohio adults still remain susceptible to SARS-CoV-2 infection. This implies that the state must remain vigilant and continue to deploy nonpharmaceutical interventions like masks and social distancing to limit the spread of SARS-CoV-2.

Methodologically, we developed a coherent statistical framework for analyzing seroprevalence surveys with multiple imperfect diagnostic tests. Typically, one might rely on a single test or the creation of deterministic rules that combine test results to assess positive cases. In contrast, we utilized existing methodology for combining the results of multiple diagnostic tests

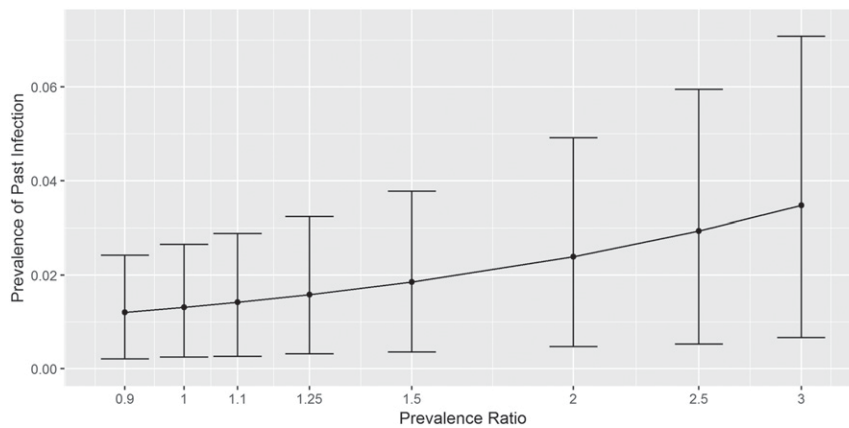
to fully incorporate information on infection status from each test. We then connected the estimated infection status to a multilevel regression and poststratification approach for generating population-based estimates. Through our fully Bayesian, model-based analysis, we are able to appropriately propagate uncertainty and provide exact inferences while synthesizing the information from each type of test administered. Our approach allows us to make full use of the data collected by the survey while also accounting for its quality. Future methodological challenges remain as test characteristics are likely a function of time since infection, which is unknown and not explicitly considered in our model.

Our study has significant strengths: a representative random sample, employing multiple diagnostic tests, and conducting a fully Bayesian analysis. There are also limitations. The survey was designed to estimate SARS-CoV-2 infection prevalence at the level of the whole state and was not designed for estimation within any subgroups or smaller geographies. A future study should consider oversampling high-risk subgroups or regions to enable inference specific to those important populations. Non-response was a major issue that was addressed in the field using all available resources. We do not have detailed information on nonresponders that could have been incorporated into a more sophisticated model for missing data. However, based on our sensitivity analysis, we do not believe the substantive policy implications of the study would change based on a more complicated missing data model. Finally, the informative prior distributions for antibody test characteristics are based on validation data presented in the test inserts—we are unable to verify the rigor and quality of those validation studies.

In conclusion, we have developed a statistical analysis framework for analyzing seroprevalence survey data with few positive cases and results from multiple imperfect diagnostic tests. Since

**Table 3.** Household-level nonresponse rates by region

Region	Rate
Central	0.783
East central	0.841
Northeast	0.840
Northwest	0.832
Southeast	0.752
Southeast central	0.790
Southwest	0.810
West central	0.818



**Fig. 3.** Posterior mean and 95% highest posterior density credible intervals for the prevalence of past infection under several scenarios of nonignorable nonresponse. The prevalence ratio is  $\lambda$  in Eq. 6.

estimates of prior and current prevalence are relevant to policy, seroprevalence studies of SARS-CoV-2 are becoming increasingly important and more frequently conducted. Our methodological approach is a critical component to ensuring that serology data are analyzed in a way that is consistent both with the design of the survey and with the inherent limitations in the accuracy of antibody tests. This enables policy makers to have access to the best available estimates that also fully and honestly account for all of the sources of uncertainty that contribute to the quantification of SARS-CoV-2 infection.

**Data Availability.** Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data are not publicly available. However, deidentified data is available upon reasonable request to the authors. Code is publicly available in GitHub at <https://github.com/sinafala/bayes-prevalence>.

**ACKNOWLEDGMENTS.** We acknowledge those who were involved in the conception, design, and implementation of the study, including Alison Norris, Stanley Lemeshow, Maria Gallo, Elisabeth Root, Gene Phillips, and Morgan Spahnne. We also thank all of the study staff for their effort collecting the data in the field. This study was funded by the ODH.

- E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).
- National Center for Health Statistics and Others, *Excess Deaths Associated with COVID-19* (Center for Disease Control and Prevention, Washington, DC, 2020).
- D. M. Weinberger *et al.*, Estimation of excess deaths associated with the COVID-19 pandemic in the United States, March to May 2020. *J. Am. Med. Ass. Intern. Med.* **180**, 1336–1344 (2020).
- L. M. Rossen, A. M. Branum, F. B. Ahmad, P. Sutton, R. N. Anderson, Excess deaths associated with COVID-19, by age and race and ethnicity—United States, January 26–October 3, 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 1522 (2020).
- M. Lisboa Bastos *et al.*, Diagnostic accuracy of serological tests for COVID-19: Systematic review and meta-analysis. *BMJ* **370**, m2516 (2020).
- V. B. Franceschi *et al.*, Population-based prevalence surveys during the COVID-19 pandemic: A systematic review. *Rev. Med. Virol.*, 10.1002/rmv.2200 (2020).
- M. Shakiba *et al.*, Seroprevalence of SARS-CoV-2 in Guilan province, Iran, April 2020. *Emerg. Infect. Dis.* **27**, 636–638 (2021).
- H. Ward *et al.*, Antibody prevalence for SARS-CoV-2 in England following first peak of the pandemic: REACT2 study in 100,000 adults. medRxiv [Preprint] (2020). <https://doi.org/10.1101/2020.08.12.20173690> (21 April 2020).
- S. Riley *et al.*, Community prevalence of SARS-CoV-2 virus in England during May 2020: REACT study. medRxiv [Preprint] (2020). <https://doi.org/10.1101/2020.07.10.20150524> (11 July 2020).
- P. M. Vodičar *et al.*, Low prevalence of active COVID-19 in Slovenia: A nationwide population study of a probability-based sample. *Clin. Microbiol. Infect.* **26**, 1514–1519 (2020).
- B. Merkely *et al.*, Novel coronavirus epidemic in the Hungarian population, a cross-sectional nationwide survey to support the exit policy in Hungary. *GeroScience* **42**, 1063–1074 (2020).
- M. Pollán *et al.*, Prevalence of SARS-CoV-2 in Spain (ENE-COVID): A nationwide, population-based seroepidemiological study. *Lancet* **396**, 535–544 (2020).
- C. J. Snoeck *et al.*, Prevalence of SARS-CoV-2 infection in the Luxembourgish population—the CON-VINCE study. medRxiv [Preprint] (2020). <https://doi.org/10.1101/2020.05.11.20092916> (18 May 2020).
- D. F. Gudbjartsson *et al.*, Spread of SARS-CoV-2 in the Icelandic population. *N. Engl. J. Med.* **383**, 2184–2185 (2020).
- N. Menachemi *et al.*, Population point prevalence of SARS-CoV-2 infection based on a statewide random sample—Indiana, April 25–29, 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 960–964 (2020).
- S. Mahajan *et al.*, Seroprevalence of SARS-CoV-2-specific IgG antibodies among adults living in Connecticut: Post-infection prevalence (PIP) study. *Am. J. Med.* **134**, 526–534.e11 (2021).
- C. Costa Gomes *et al.*, A population-based study of the prevalence of COVID-19 infection in Espírito Santo, Brazil: Methodology and results of the first stage. medRxiv [Preprint] (2020). <https://doi.org/10.1101/2020.06.13.20130559> (16 June 2020).
- A. A. M. Da Silva *et al.*, Population-based seroprevalence of SARS-CoV-2 is more than halfway through the herd immunity threshold in the State of Maranhao, Brazil. medRxiv [Preprint] (2020). <https://doi.org/10.1101/2020.08.28.20180463> (1 September 2020).
- M. F. Silveira *et al.*, Population-based surveys of antibodies against SARS-CoV-2 in Southern Brazil. *Nat. Med.* **26**, 1196–1199 (2020).
- P. Hallal *et al.*, Remarkable variability in SARS-CoV-2 antibodies across Brazilian regions: Nationwide serological household survey in 27 states. medRxiv [Preprint] (2020). <https://doi.org/10.1101/2020.05.30.20117531> (30 May 2020).
- A. Gelman, B. Carpenter, Bayesian analysis of tests with unknown specificity and sensitivity. *J. Roy. Stat. Soc. Ser. C Appl. Stat.* **69**, 1269–1283 (2020).
- N. Dendukuri, L. Joseph, Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* **57**, 158–167 (2001).
- Z. Wang, N. Dendukuri, H. J. Zar, L. Joseph, Modeling conditional dependence among multiple diagnostic tests. *Stat. Med.* **36**, 4843–4859 (2017).
- A. Gelman, Struggles with survey weighting and regression modeling. *Stat. Sci.* **22**, 153–164 (2007).
- S. Makela, Y. Si, A. Gelman, Bayesian inference under cluster sampling with probability proportional to size. *Stat. Med.* **37**, 3849–3868 (2018).
- Abbott, SARS-CoV-2 IgG package insert (2021). <https://www.fda.gov/media/137383/download>. Accessed 21 August 2020.
- Diasorin, Liaison SARS-CoV-2 S1/S2 IgG package insert (2020). <https://www.diasorin.com/sites/default/files/allegati/liaisonr-sars-cov-2-s1s2-igg-brochure.pdf.pdf>. Accessed 10 August 2020.
- Epitope Diagnostics, Inc., EDI Novel Coronavirus COVID-19 ELISA Kits package insert (2020). <http://www.epitopediagnostics.com/covid-19-elisa>. Accessed 30 July 2020.
- R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2019).
- P. de Valpine *et al.*, Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *J. Comput. Graph Stat.* **26**, 403–417 (2017).
- M. M. Patel *et al.*, Change in antibodies to SARS-CoV-2 over 60 days among health care personnel in Nashville, Tennessee. *J. Am. Med. Assoc.* **324**, 1781–1782 (2020).
- F. J. Ibarrondo *et al.*, Rapid decay of anti-SARS-CoV-2 antibodies in persons with mild COVID-19. *N. Engl. J. Med.* **383**, 1085–1087 (2020).
- D. Zhao *et al.*, Asymptomatic infection by SARS-CoV-2 in healthcare workers: A study in a large teaching hospital in Wuhan, China. *Int. J. Infect. Dis.* **99**, 219–225 (2020).
- A. S. Iyer *et al.*, Persistence and decay of human antibody responses to the receptor binding domain of SARS-CoV-2 spike protein in COVID-19 patients. *Sci. Immunol.* **5**, eabe0367 (2020).
- Q.-X. Long *et al.*, Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat. Med.* **26**, 1200–1204 (2020).