



Performance of Three Tests for SARS-CoV-2 on a University Campus Estimated Jointly with Bayesian Latent Class Modeling

 T. Alex Perkins,^a Melissa Stephens,^b Wendy Alvarez Barrios,^c Sean Cavany,^a Liz Rulli,^c Michael E. Pfrender^a

^aDepartment of Biological Sciences, University of Notre Dame, Notre Dame, Indiana, USA

^bGenomics and Bioinformatics Core Facility, University of Notre Dame, Notre Dame, Indiana, USA

^cNotre Dame Research, University of Notre Dame, Notre Dame, Indiana, USA

ABSTRACT Accurate tests for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have been critical in efforts to control its spread. The accuracy of tests for SARS-CoV-2 has been assessed numerous times, usually in reference to a gold standard diagnosis. One major disadvantage of that approach is the possibility of error due to inaccuracy of the gold standard, which is especially problematic for evaluating testing in a real-world surveillance context. We used an alternative approach known as Bayesian latent class modeling (BLCM), which circumvents the need to designate a gold standard by simultaneously estimating the accuracy of multiple tests. We applied this technique to a collection of 1,716 tests of three types applied to 853 individuals on a university campus during a 1-week period in October 2020. We found that reverse transcriptase PCR (RT-PCR) testing of saliva samples performed at a campus facility had higher sensitivity (median, 92.3%; 95% credible interval [CrI], 73.2 to 99.6%) than RT-PCR testing of nasal samples performed at a commercial facility (median, 85.9%; 95% CrI, 54.7 to 99.4%). The reverse was true for specificity, although the specificity of saliva testing was still very high (median, 99.3%; 95% CrI, 98.3 to 99.9%). An antigen test was less sensitive and specific than both of the RT-PCR tests, although the sample sizes with this test were small and the statistical uncertainty was high. These results suggest that RT-PCR testing of saliva samples at a campus facility can be an effective basis for surveillance screening to prevent SARS-CoV-2 transmission in a university setting.

IMPORTANCE Testing for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been vitally important during the COVID-19 pandemic. There are a variety of methods for testing for this virus, and it is important to understand their accuracy in choosing which one might be best suited for a given application. To estimate the accuracy of three different testing methods, we used a data set collected at a university that involved testing the same samples with multiple tests. Unlike most other estimates of test accuracy, we did not assume that one test was perfect but instead allowed for some degree of inaccuracy in all testing methods. We found that molecular tests performed on saliva samples at a university facility were similarly accurate as molecular tests performed on nasal samples at a commercial facility. An antigen test appeared somewhat less accurate than the molecular tests, but there was high uncertainty about that.

KEYWORDS Bayesian statistics, COVID-19, RT-PCR, SARS-CoV-2, epidemiology, molecular diagnostic, public health surveillance

Molecular testing has played a vital role in efforts to suppress the transmission of SARS-CoV-2. This applies in both community settings (1, 2) and in more specialized settings, such as hospitals (3, 4), workplaces (5, 6), schools (7, 8), and travel (9, 10). Although many contextual factors affect the success of testing (11, 12), the foundation

Editor Jennifer Dien Bard, Children's Hospital Los Angeles, University of Southern California

Copyright © 2022 Perkins et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to T. Alex Perkins, taperkins@nd.edu, or Michael E. Pfrender, mpfrende@nd.edu.

The authors declare no conflict of interest.

Received 9 August 2021

Accepted 12 December 2021

Published 19 January 2022

of any successful testing program is the availability of tests that are sufficiently sensitive and specific to achieve the program's objectives.

Most evaluations of the sensitivity and specificity of molecular tests for SARS-CoV-2 have been performed in reference to a diagnostic that was considered a gold standard (13). Designating a diagnostic as a gold standard makes the calculation of sensitivity and specificity straightforward, as true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) test outcomes can all be defined clearly in reference to the gold standard. Under this assumption, the sensitivity can be estimated as $TP/(TP + FN)$ and specificity as $TN/(TN + FP)$.

A key limitation of this approach is that the estimates it yields are only as reliable as the gold standard on which they are based. The most common gold standard is reverse transcriptase PCR (RT-PCR) testing (14). This standard is far from golden, however. Especially with respect to sensitivity, the performance of these tests for SARS-CoV-2 has been found to vary as a function of the method of sample extraction (15, 16), day of infection (17, 18), and disease severity of the subject (19). Furthermore, designation of one method as a gold standard makes it impossible to evaluate whether another test might actually have better sensitivity or specificity than the presumed gold standard (20).

One way to circumvent the limitations associated with relying on a gold standard is to use an alternative method for analysis, such as Bayesian latent class modeling (BLCM) (21). This method involves joint estimation of the sensitivity and specificity of each type of test used, by virtue of considering the possibility that any given test result could have been erroneous for some, all, or none of the tests used. While this approach does not make test results more accurate *per se*, it does reduce the risk of bias associated with erroneously assuming that a gold standard is without error. This approach has been applied in some cases for molecular tests for SARS-CoV-2, resulting in differences relative to estimates that relied on a gold standard (22–24). For example, in a meta-analysis comparing RT-PCR testing of nasopharyngeal and saliva samples, allowing for imperfections in both types of tests resulted in higher estimates of specificity and narrower uncertainty about sensitivity (23).

In this study, we applied BLCM to a data set from a SARS-CoV-2 testing program in a university setting during October 2020. A unique feature of this data set is that it includes both RT-PCR and antigen tests, which have not been compared in previous BLCM analyses for SARS-CoV-2 that we are aware of (22–24). Another unique feature of this data set is that the majority of subjects were tested during surveillance screening and were not suspected of being infected at the time of testing. This presents an opportunity to quantify the test performance in a context that is highly relevant for public health (11). Moreover, the fact that the majority of subjects were in the 18-to-25 age range presents an opportunity to quantify test performance in a population for which tests could potentially be less sensitive (19, 25, 26) yet are of high value for surveillance (7, 8).

RESULTS

Combining data across all tests performed, the test positivity was 2.5% (43/1,716). The positivity was lower among individuals tested for surveillance purposes (1.9%) than among individuals tested for other reasons (12.3%). The positivity was also lower for commercial tests (1.4%) than for saliva (3.1%) and antigen (13.5%) tests. The lower positivity among commercial tests held when controlling for the method by which individuals came to participate in the study (Table S1). Despite the differences in positivity, the very low positivity overall meant that the concordance was high: 99.3% between commercial and saliva tests, 96.3% between commercial and antigen tests, and 97.3% between saliva and antigen tests.

Using Bayesian analysis, we estimated a total of eight parameters (Table 1) by leveraging joint information about all observed combinations of testing outcomes across the three types of tests (Table 2). Application of this method to 100 simulated data sets showed good coverage of the true parameter values (Fig. S2 in the supplemental material).

TABLE 1 Parameter definitions and posterior estimates

Symbol	Definition	Posterior median (95% CrI) ^a
Se_{Comm}	Sensitivity of commercial PCR test	0.859 (0.547–0.994)
Se_{Saliva}	Sensitivity of saliva test	0.923 (0.732–0.996)
$Se_{Antigen}$	Sensitivity of antigen test	0.748 (0.373–0.969)
Sp_{Comm}	Specificity of commercial test	0.998 (0.992–0.999)
Sp_{Saliva}	Specificity of saliva test	0.993 (0.983–0.999)
$Sp_{Antigen}$	Specificity of antigen test	0.978 (0.888–0.999)
$Prev_{Non-surv}$	Prevalence among nonsurveillance samples	0.141 (0.058–0.258)
$Prev_{Surv}$	Prevalence among surveillance samples	0.018 (0.009–0.030)

^aDecimal values are provided, consistent with the definitions of these quantities as probabilities, rather than percentages, in Materials and Methods.

Applying the method to empirical data demonstrated good convergence (Fig. S3) and resulted in posterior samples with moderately low correlation (Fig. S4), suggesting that the data were reasonably informative about the parameters we sought to estimate.

The prevalence inferred by our Bayesian analysis was similar to the test positivity for surveillance testing (median, 1.8%; 95% credible interval [CrI], 0.9 to 3.0%) (Fig. 1A) and slightly higher for nonsurveillance testing (median, 14.1%; 95% CrI, 5.8 to 25.8%) (Fig. 1B). For the saliva tests, we estimated a sensitivity of 92.3% (95% CrI, 73.2 to 99.6%) (Fig. 1C, green) and a specificity of 99.3% (95% CrI, 98.3 to 99.9%) (Fig. 1D, green). Had we considered the commercial tests to be the gold standard, we would have instead estimated the sensitivity and specificity of the saliva tests to be 83.3% and 99.5%, respectively. Similarly, for the antigen tests, our estimate of the sensitivity (74.8%; 95% CrI, 37.3 to 96.9%) (Fig. 1C, blue) was greater than an estimate made in reference to the commercial tests (0.5), and our estimate of the specificity (97.8%; 95% CrI, 88.8 to 99.9%) (Fig. 1D, blue) was lower than an estimate made in reference to the commercial tests (100%). It is important to note the large uncertainty around these estimates due to the relatively low number of individuals who received an antigen test. These discrepancies were a result of the fact that we did not consider commercial tests to be the gold standard and estimated their sensitivity and specificity alongside those of the other two test types. Doing so resulted in estimates of the sensitivity of 85.9% (95% CrI, 54.7 to 99.4%) (Fig. 1C, red) and specificity of 99.8% (95% CrI, 99.2 to 99.9%) (Fig. 1D, red) for the commercial tests.

While comparison of the median sensitivities and specificities implies that some tests were more sensitive or specific than others, the wide uncertainty of our estimates must be considered when making such comparisons. We obtained more nuanced insight into the relative sensitivities and specificities of the three tests by calculating the proportion of samples in which the sensitivity of one test exceeded that of another,

TABLE 2 Testing data^a

Reason for testing	Commercial	Saliva	Antigen	No. of participants
Surveillance	-	-	-	3
	-	-	NA	778
	-	NA	NA	7
	NA	-	NA	3
	-	+	NA	4
	+	-	NA	1
	+	+	NA	9
Other	NA	+	NA	7
	-	-	-	22
	-	-	NA	7
	NA	-	-	5
	NA	+	+	4

^aEach of the 853 study participants fell into one of the categories represented by each row. These categories differed with respect to the reason for testing and the outcome of each test. NA, not applicable; this indicates that a given test was not performed for those individuals. This table constitutes the full information used in our analysis. -, symbol indicates a negative test; +, indicates a positive test.

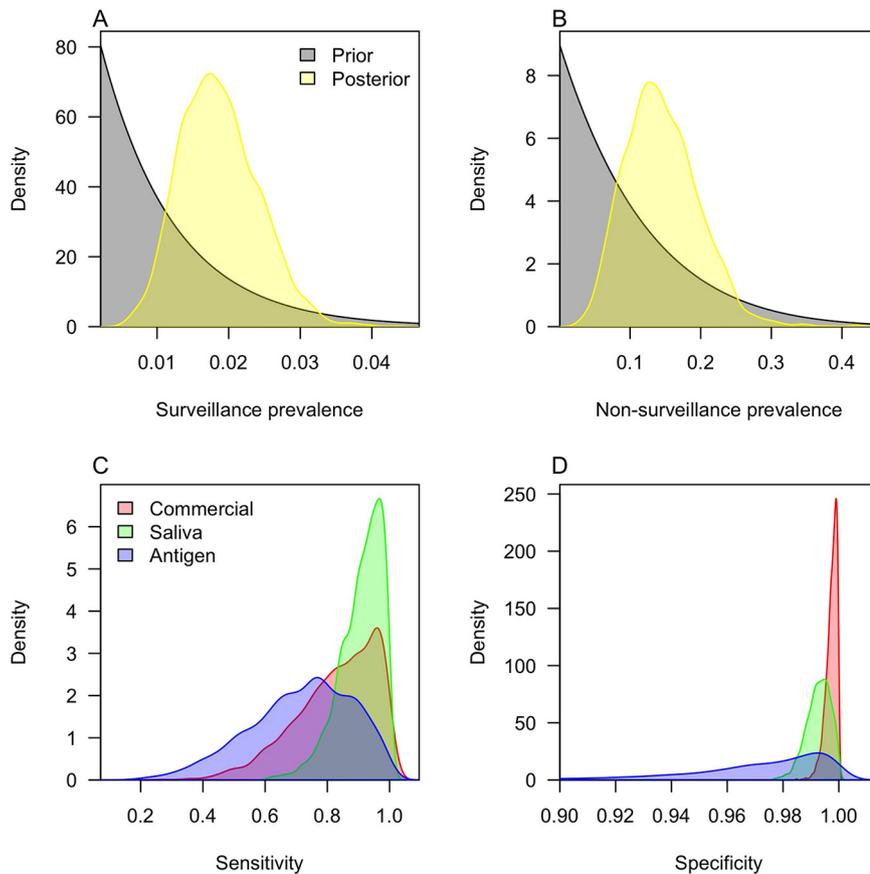


FIG 1 Posterior parameter estimates. (A) Prevalence among individuals participating in surveillance testing; (B) prevalence among individuals participating in testing for reasons other than surveillance; (C) test sensitivity; and (D) test specificity. The colors in A and B distinguish the prior from posterior distributions, and the colors in C and D distinguish the different types of tests. Values outside the range 0 to 1 occur only as a result of smoothing. Decimal values are shown along the x axes, consistent with the definitions of these quantities as probabilities, rather than percentages, in Materials and Methods.

and likewise for the specificity. On that basis, we found a probability of 0.69 that the saliva test was more sensitive than the commercial test (Table S2). The probabilities that the saliva and commercial tests were more sensitive than the antigen test were 0.88 and 0.71, respectively. The probabilities that the commercial test was more specific than the saliva and antigen tests were 0.86 and 0.92, respectively (Table S3). The saliva test was more specific than the antigen test with a probability of 0.81.

Joint inference of the test properties and the prevalence of infection allowed us to estimate the frequency of different outcomes from surveillance testing (Fig. 2). Due to its high sensitivity, the saliva test was predicted to yield the most true positives (16.1 per 1,000 tests; 95% CrI, 8.2 to 27.6) (Fig. 2A) and the fewest false negatives (1.3 per 1,000 tests; 95% CrI, 0.07 to 5.4) (Fig. 3C). At the other extreme, 1,000 antigen tests were predicted to yield 12.7 true positives (95% CrI, 5.4 to 23.9) and 4.4 false negatives (95% CrI, 0.5 to 13.1). The antigen tests also had the lowest specificity, resulting in the largest number of false positives (21.3 per 1,000 tests; 95% CrI, 0.8 to 110.3) (Fig. 2B). The commercial tests were estimated to perform best in this regard, yielding only 2.1 false positives per 1,000 tests (95% CrI, 0.1 to 7.5).

Because of their high specificity and the low prevalence of infection, the commercial tests were predicted to have the highest positive predictive value during the study period (87%; 95% CrI, 62 to 99%) (Fig. 3A). Given such low prevalence, all tests had high negative predictive values and were predicted to result in mostly true negatives (Fig. 3B) for the vast majority of tests. Under a scenario of surveillance screening at

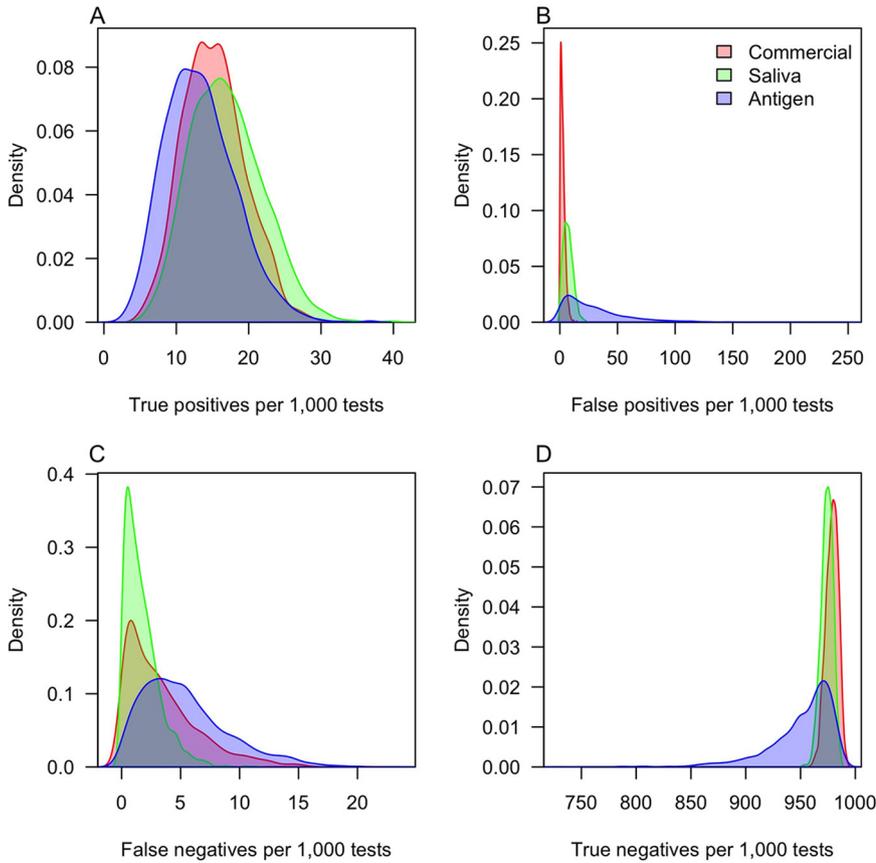


FIG 2 Estimates of the frequency of different testing outcomes. Out of 1,000 tests, the panels show the number of (A) true positives, (B) false positives, (C) false negatives, and (D) true negatives. The colors distinguish the different types of tests. Values outside the range 0 to 1,000 occur only as a result of smoothing.

random from the campus population during the fall 2020 semester, we estimated that saliva tests during the semester would be expected to have median positive predictive values as low as 0.2% (95% CrI, 0.07 to 2.4%) on 1 August and as high as 89% (95% CrI, 76 to 99%) on 22 August (Fig. 4B). The negative predictive values of the saliva test never would have been less than a median of 99.5% (95% CrI, 98.4 to 99.9%) under this scenario (Fig. 4E). The commercial tests would have had higher positive predictive values under this scenario (Fig. 4A), and both the commercial and antigen tests would have had lower negative predictive values (Fig. 4D and F).

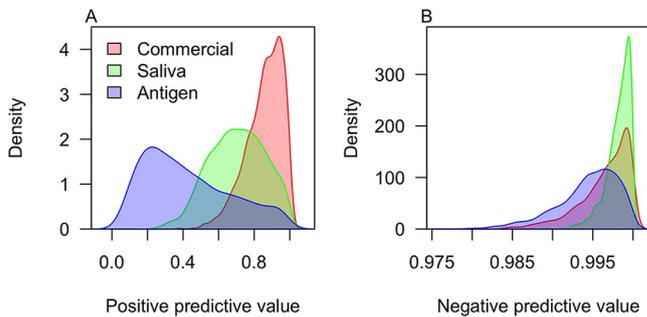


FIG 3 Estimates of the predictive values of each test during the study period. The panels show estimates of (A) the positive predictive value and (B) the negative predictive value. The colors distinguish the different types of tests. Values outside the range 0 to 1 occur only as a result of smoothing. Decimal values are shown along the x axes, consistent with the definitions of these quantities as probabilities, rather than percentages, in Materials and Methods.

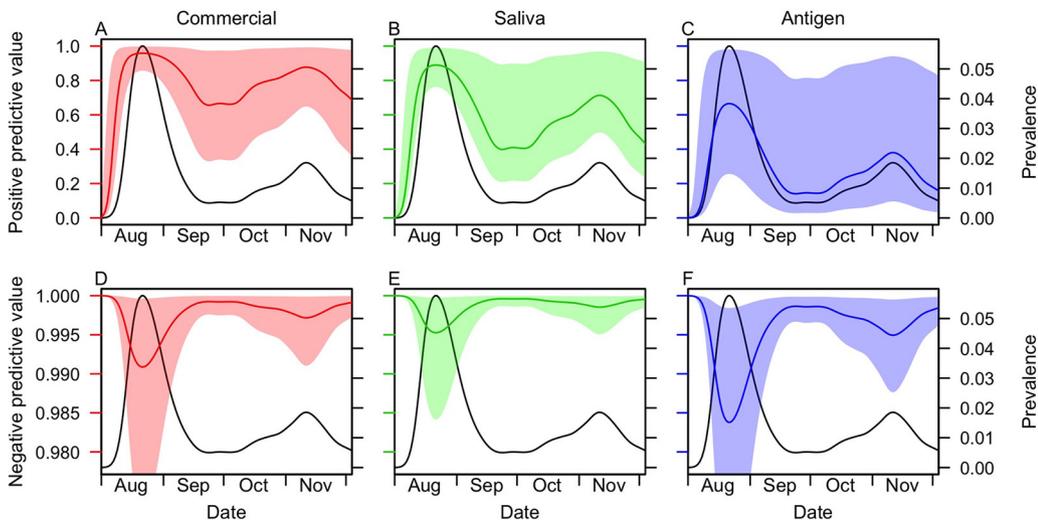


FIG 4 Positive predictive value (A to C) and negative predictive value (D to F) over the course of the entire semester. These values represent the probability that a positive or negative test result under random surveillance screening would have accurately relayed the true positive or negative status of the individual being tested. The change over time was a result of the time-varying prevalence of detectable infection (black lines, right axis). The uncertainty reflects uncertainty about the sensitivity and specificity of each type of test: commercial (red), saliva (green), and antigen (blue). Decimal values are shown along the y axes, consistent with the definitions of these quantities as probabilities, rather than percentages, in Materials and Methods.

DISCUSSION

With respect to the sensitivity, there were important differences between our modeled estimates and those based on the raw test data. The clearest difference was for the commercial test, for which we obtained a median estimate of 86% for the sensitivity. Had we considered that test to be a gold standard, we would have obtained a point estimate for the sensitivity of the saliva test 16% lower than our median estimate, and 33% lower for the antigen test. On the contrary, we found support for the saliva test likely being more sensitive than the commercial test, which would not have been possible to infer had we assumed a gold standard (20). Our finding that the saliva test was more sensitive than the commercial test, which was based on a nasal swab sample, is consistent with findings from other studies (27–30). We estimated that the antigen test had markedly lower sensitivity than either of the RT-PCR tests, which is also consistent with other studies (31, 32). Given that only 37 individuals received an antigen test, it is important to bear in mind that the uncertainty about its properties is high.

With respect to specificity, the medians of our modeled estimates were slightly lower than estimates based on the commercial test as a gold standard. Even so, our specificity estimate for the saliva test (median, 99.3%; 95% CrI, 98.3 to 99.9%) was strikingly similar to an independent estimate (median, 99.2%; 95% CrI, 98.2 to 99.8%) that also used a BLCM, but in the context of a meta-analysis (23). Our analysis generated high confidence in the commercial test being the most specific and the antigen test being the least specific, with a probability of 0.8. Our median estimate of 97.8% for the antigen test specificity was within the range of published estimates for seven different antigen tests (33), which had median values ranging from 98.5 to 100% for five tests and from 88.9 to 94.8% for two outliers. Uncertainty about the test specificity was relatively low in our estimates, given that the vast majority of the individuals we tested were likely true negatives.

These test sensitivities and specificities have implications for several metrics of public health importance. Given its high sensitivity, the saliva test was expected to detect the most true positives and produce the fewest false negatives, as indicated by a high negative predictive value. For the purpose of identifying infections in surveillance

screening so that they can be isolated and their transmission curtailed, this test was most ideal, especially at times of high prevalence. Given its high specificity, the commercial test was expected to result in the fewest false positives and the most true negatives, as indicated by a high positive predictive value. These properties are ideal from the perspective of minimizing unnecessary demand on resources for case isolation and contact tracing, more so at times of low prevalence when demand is already low. The antigen test performed least well in all regards. While this may make it seem like a less desirable option, it should be noted that sensitivity over the course of infection as a whole need not be paramount. In the event that a test has a high sensitivity around the time of peak infectiousness, its value for curtailing transmission could still be very high (34).

Given the implications of our estimates of sensitivity and specificity, it is important to understand their empirical basis. It is notable that, out of 853 individuals tested, only six had discrepant results. In four of those cases, the saliva test was positive and the commercial test was negative. In two, the commercial test was positive and one or both of the other tests were negative. A strength of our modeling approach is that it integrated across all of the available information to inform its estimates, rather than those six discrepancies alone. The model also took into account the higher positivity of the saliva and antigen tests, compared to the commercial test. Likewise, it was capable of balancing that with indications that the antigen test had a lower specificity, which could explain its higher observed test positivity in part. Additionally, the model was able to account for higher positivity among the nonsurveillance tests due to a higher prevalence in that group, which is important given that the three types of tests were not applied evenly across the two groups. These competing influences on our estimates underscore the value of the BLCM approach we used, which was able to balance them appropriately and express that balance in the form of quantitative descriptions of uncertainty.

Although our analysis was able to provide insight into the properties of these three tests, there were some uncertainties that we were unable to resolve. Correlations among three parameters— Se_{Comm} , $Sp_{Salivar}$, and $Prev_{Surv}$ —were indicative of uncertainty about whether the four individuals with positive saliva tests and negative commercial tests resulted from false-positive saliva tests or false-negative commercial tests. More data would be helpful for resolving this uncertainty, although doing so would require relatively rare discrepant results. Another limitation of our study is that the only information we used to resolve the uncertainty about the true infection status was whether individuals were tested for surveillance purposes. Given that the prevalence differed by an order of magnitude between these groups, this was quite beneficial. However, additional information—such as recent contacts or status as student, faculty, or staff—could have potentially helped narrow this type of uncertainty further. Doing so would have required estimating more parameters and could have made the analysis more susceptible to bias if an increasingly complex model were not specified properly. A related limitation is that we were unable to examine the possibility of differing sensitivities as a function of the symptom status. Had we incorporated this into our model, it would have been impossible to distinguish between differing sensitivities and differing prevalences among the individuals with differing symptom statuses. Similar to our study, a large-scale study (35) of over 1 million people in the United Kingdom found that the prevalence among individuals presenting with symptoms was several-fold higher than among those not reporting symptoms. Accordingly, we felt that it was appropriate to place more emphasis on estimating the differences in prevalence than the differences in sensitivity between these groups.

In addition to limitations of our analysis, there were also limitations of our data set. First, a more balanced testing effort across different test types and groups of subjects could have helped reduce the uncertainty about certain parameters, especially those relating to the antigen tests. Second, for individuals who truly were positive, we have no information about how many days elapsed between their initial exposure and

when they were tested. Given how variable the test sensitivity is over the course of an infection (11, 17), this factor alone could be a major driver of the sensitivities we estimated. For example, individuals tested for surveillance screening had presumably not displayed any symptoms up to the time of testing, making it possible that many of our positive surveillance tests came from presymptomatic individuals with high viral loads (36). Even so, the balance of saliva and commercial tests across the individuals tested for surveillance versus nonsurveillance purposes was similar, meaning that any differences in the timing of testing between these two groups should not have affected our inferences about the relative sensitivities of these two types of tests.

In conclusion, our analysis leveraged all the data collected from this study to estimate the sensitivities and specificities of three types of tests, without the need to consider any of those tests as a gold standard. These estimates are pertinent to a setting for which surveillance testing has been (37, 38), and remains (39), a major emphasis of COVID-19 prevention. Although there is appreciable uncertainty associated with our estimates, this uncertainty was quantified carefully and could be reduced in the future by updating our estimates with additional data. Bayesian analyses lend themselves to this naturally, given that posterior estimates from one study can serve as prior estimates for another.

MATERIALS AND METHODS

Sample collection. All samples for this study were collected during a 5-day period from Monday, 12 October, through Friday, 16 October 2020. In total, 1,716 tests were performed on samples collected from 853 individuals, with multiple tests for a single individual applied to separate specimens collected on the same day. Most individuals (811) participated in response to a request for surveillance testing, while others (37) participated either as a result of reporting symptoms associated with COVID-19 (40), because of suspected exposure through contact (10), or because they had previously tested positive and were undergoing a second test 4 days later (3). The participants consisted of 846 students and 7 faculty and staff. The majority (87.6%) of the students were between the ages of 18 and 22 inclusive, with a range of 18 to 39, a median age of 20, and a mean age of 21.2. The median age of the staff was 40, and two members of the staff were over the age of 65, with a range of 29 to 72 (Fig. S1). These individuals received a total of 833 commercial RT-PCR tests on nasal swab specimens, 846 in-house RT-PCR tests on saliva specimens, and 37 antigen tests on nasal swab specimens. We refer to these tests here as *commercial*, *saliva*, and *antigen tests*, respectively. A majority of individuals (799) received commercial and saliva tests but not an antigen test, a subset (41) received all three tests, and the remainder (41) received either one or two tests in other combinations. The low number of antigen tests limited the precision of estimates of its performance.

Laboratory testing. (i) SARS-CoV-2 detection in saliva samples. Following the University of Notre Dame Institutional Biosafety Committee (IBC)-approved protocol (20-08-6161), fresh saliva samples were obtained from study participants and tested for the presence of SARS-CoV-2 within 17 h after collection. Steps for the detection of SARS-CoV-2 through RT-qPCR in the saliva samples were adapted from Ranoa et al. (41). A minimum of 200 μL saliva was collected from each participant in a barcoded nuclease-free 50-mL conical tube. Following collection, the samples were heat inactivated by incubating them in a 95°C circulating water bath for 30 min. After cooling to room temperature, the inactive specimen was diluted at a 1:1 ratio (vol/vol) with 2 \times Tris-borate-EDTA buffer (0.089 M Tris, 0.089 M borate, 0.002 M EDTA in a final 1 \times buffer solution), followed by vigorous vortexing to ensure thorough mixing. The diluted saliva was then subjected to RT-qPCR (1 reaction per sample) using the TaqPath COVID-19 combo kit (Thermo Fisher Scientific), which includes three primer/probe sets specific to SARS-CoV-2 genes (ORF1ab, S gene, N gene) and one MS2 bacteriophage control target. Briefly, 5 μL diluted saliva was added to a freshly prepared reaction mix containing 2.5 μL TaqPath 1-step multiplex master mix (no ROX) (4 \times), 0.5 μL COVID-19 real-time PCR assay multiplex, 0.5 μL MS2 phage control, and 1.5 μL nuclease-free water. Reactions were set up in a 96-well plate format (0.1 mL MicroAmp fast optical 96-well reaction plate [Applied Biosystems]), with each plate containing a positive control diluted to 4 copies/ μL and a no-template control (nuclease-free water), as well as the MS2 control internal to each sample. All RT-qPCRs were carried out using QuantStudio RT-qPCR instruments (Applied Biosystems). The reaction parameters were set as follows: *hold stage*, 25°C for 2 min, 53°C for 10 min, 95°C for 2 min; *PCR stage* (40 \times), 95°C for 3 s, 60°C for 30 s; 1.6°C/s ramp for all stages; run mode “fast.” Targets, reporter dyes, and quencher information for the RT-qPCR instrument were set up according to the TaqPath COVID-19 combo kit manufacturer’s instructions.

A presence/absence analysis of the viral targets was performed using Applied Biosystems Design and Analysis v2.4 software with the baseline set at 5 and the quantification cycle (C_q) cutoff for all targets set at 37. The results of the RT-qPCR test were interpreted as positive, negative, or invalid. A positive test had at least 2 of the 3 gene targets present within the threshold settings. All positive and invalid tests were subjected to repeat testing for confirmation.

(ii) SARS-CoV-2 commercial and rapid antigen detection. Self-administered nasal swab samples were outsourced to LabCorp, Inc., for viral detection using a RT-PCR protocol (EUA200011). Rapid antigen assays were performed on self-administered nasal swab samples using the Sofia2 fluorescent

immunoassay (FIA) analyzer and the SARS antigen fluorescent immunoassay for qualitative detection of the nucleocapsid protein from SARS-CoV-2 (Quidel).

Statistical analysis. (i) Model. For our analysis, we estimated eight parameters (Table 1), which together determine the probability of each type of testing outcome. The likelihood of a given set of values of these parameters is equal to the probability of the observed testing outcomes, given those parameter values. The data were defined according to the number of individuals with a given combination of testing outcomes as $n_{ij,k}$, where i , j , and k refer to positive, negative, or missing results, respectively, for each type of test (commercial, saliva, and antigen) (Table 2).

By definition, the test sensitivity and specificity are specified in reference to the true infection status of an individual, which we define as having been infected with SARS-CoV-2 recently enough to still contain sufficient RNA to be detectable. We note that being infected under this definition does not necessarily imply that a person is infectious (42). Because we did not know the true infection status of any individual with certainty, we defined the probability of a given set of testing outcomes (i.e., i , j , k) conditional on the true status, which we refer to as s (this could be either + or -). This probability, $\Pr(i,j,k|s)$, is defined as the product of the probabilities of each testing outcome given status s . It is important to note that our model does not assume or infer the true infection status of any study participants; rather, it treats the true infection status probabilistically as an unknown state.

To account for the fact that the true status of any given infection is unknown, we used the law of total probability to calculate the overall probability of the observed testing outcomes, $\Pr(i,j,k)$, as the weighted average of the conditional probabilities of the observed testing outcomes:

$$\Pr(i,j,k) = \Pr(i,j,k|+)Prev + \Pr(i,j,k|-)(1 - Prev)$$

Denoting the set of all parameters as θ , we defined the likelihood of the parameters, given the data, \mathbf{n} , as

$$L(\theta|\mathbf{n}) = \prod_{\{i,j,k\} \in \mathbf{n}} \Pr(i,j,k)^{n_{i,j,k}}$$

In these calculations, different values of $Prev$ were used depending on whether the individuals were tested as part of surveillance efforts or for other reasons.

(ii) Estimation procedure. Taking a Bayesian approach to parameter estimation, the posterior probability of the parameters in θ was defined as

$$\Pr(\theta|\mathbf{n}) = \frac{L(\theta|\mathbf{n}) \Pr(\theta)}{\Pr(\mathbf{n})}$$

where $L(\theta|\mathbf{n})$ is the likelihood defined above, $\Pr(\theta)$ is the prior probability of the parameters, and $\Pr(\mathbf{n})$ is the probability of the data. We assumed noninformative priors for the sensitivity and specificity parameters, and we assumed informative priors for the two types of prevalence that were in loose alignment with the estimated prevalence at the time and location of sample collection. We avoided calculation of $\Pr(\mathbf{n})$ by using Markov chain Monte Carlo (MCMC) sampling. Details about the prior assumptions and MCMC algorithm are provided in the supplemental text.

(iii) Validation. To validate our model, we applied it to 100 simulated data sets and compared inferred parameter values to the true parameter values used to simulate the data. To ensure that the model's inferences were valid for data resembling those used in this study, we simulated the same number of individuals tested with the same combination of tests as in our empirical data. Simulated parameter values were drawn uniformly and independently from the 95% credible interval of each parameter. We examined coverage probabilities and correlations between median and true parameter values across the 100 simulated data sets.

(iv) Predictive value. Using the posterior parameter estimates, we calculated predictive values of the three tests under two different contexts. These values represent the probability that the test's indication, whether positive or negative, reflects the true status of the individual being tested. The positive predictive value is defined as

$$PPV = Se \text{ Prev} / (Se \text{ Prev} + (1 - Sp)(1 - Prev))$$

and the negative predictive value is defined as

$$NPV = Sp(1 - Prev) / ((1 - Se)Prev + Sp(1 - Prev))$$

First, we calculated PPV and NPV during the 1-week period of our study, accounting for the uncertainty in $Prev$ in doing so. Second, we calculated PPV and NPV on a daily basis over the course of the entire fall 2020 semester, accounting for the daily changes in prevalence over time. Our estimates for time-varying prevalence were based on an extrapolation of the daily incidence of symptomatic cases (40) that accounted for the proportion symptomatic (43), the incubation period (44), and the probability that a test administered on a given day of infection would be positive (11). This method is described in further detail in the supplemental text.

Data availability. All code and data from this study are available at https://github.com/TAlexPerkins/SARSCoV2_BLCM to facilitate applications or extensions of this work by others.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.7 MB.

ACKNOWLEDGMENTS

We thank Paul Hergenrother for generously sharing his expertise and experience designing and implementing a saliva-based COVID-19 surveillance program at Illinois University, Champaign-Urbana. Special thanks to Joanna McNulty and Carol Mullaney for coordinating the sample collection efforts and the technicians in the Notre Dame COVID Surveillance Lab who assisted with sample processing and data collection. This study was made possible by the enthusiastic participation of Notre Dame undergraduate students, staff, and faculty.

Funding was provided by the University of Notre Dame and Notre Dame Research. This study was conducted under University of Notre Dame IRB number 21-06-6687.

REFERENCES

- Aleta A, Martín-Corral D, Pastore y Piontti A, Ajelli M, Litvinova M, Chinazzi M, Dean NE, Halloran ME, Longini IM, Jr, Merler S, Pentland A, Vespignani A, Moro E, Moreno Y. 2020. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat Hum Behav* 4:964–971. <https://doi.org/10.1038/s41562-020-0931-9>.
- Ng T-C, Cheng H-Y, Chang H-H, Liu C-C, Yang C-C, Jian S-W, Liu D-P, Cohen T, Lin H-H. 2021. Comparison of estimated effectiveness of case-based and population-based interventions on COVID-19 containment in Taiwan. *JAMA Intern Med* 181:913. <https://doi.org/10.1001/jamainternmed.2021.1644>.
- Chin ET, Huynh BQ, Chapman LAC, Murrill M, Basu S, Lo NC. 2021. Frequency of routine testing for coronavirus disease 2019 (COVID-19) in high-risk healthcare environments to reduce outbreaks. *Clin Infect Dis* 73: e3127–e3129. <https://doi.org/10.1093/cid/ciaa1383>.
- Evans S, Agnew E, Vynnycky E, Stimson J, Bhattacharya A, Rooney C, Warne B, Robotham J. 2021. The impact of testing and infection prevention and control strategies on within-hospital transmission dynamics of COVID-19 in English hospitals. *Philos Trans R Soc Lond B Biol Sci* 376: 20200268. <https://doi.org/10.1098/rstb.2020.0268>.
- VanderWaal K, Black L, Hodge J, Bedada A, Dee S. 2021. Modeling transmission dynamics and effectiveness of worker screening programs for SARS-CoV-2 in pork processing plants. *PLoS One* 16:e0249143. <https://doi.org/10.1371/journal.pone.0249143>.
- Wells CR, Townsend JP, Pandey A, Moghadas SM, Krieger G, Singer B, McDonald RH, Fitzpatrick MC, Galvani AP. 2021. Optimal COVID-19 quarantine and testing strategies. *Nat Commun* 12:356. <https://doi.org/10.1038/s41467-020-20742-8>.
- Panovska-Griffiths J, Kerr CC, Stuart RM, Mistry D, Klein DJ, Viner RM, Bonell C. 2020. Determining the optimal strategy for reopening schools, the impact of test and trace interventions, and the risk of occurrence of a second COVID-19 epidemic wave in the UK: a modelling study. *Lancet Child Adolesc Health* 4:817–827. [https://doi.org/10.1016/S2352-4642\(20\)30250-9](https://doi.org/10.1016/S2352-4642(20)30250-9).
- Paltiel AD, Zheng A, Walensky RP. 2020. Assessment of SARS-CoV-2 screening strategies to permit the safe reopening of college campuses in the United States. *JAMA Netw Open* 3:e2016818. <https://doi.org/10.1001/jamanetworkopen.2020.16818>.
- Gostic K, Gomez AC, Mummah RO, Kucharski AJ, Lloyd-Smith JO. 2020. Estimated effectiveness of symptom and risk screening to prevent the spread of COVID-19. *Elife* 9:e55570. <https://doi.org/10.7554/eLife.55570>.
- Kiang MV, Chin ET, Huynh BQ, Chapman LAC, Rodriguez-Barraquer I, Greenhouse B, Rutherford GW, Bibbins-Domingo K, Havlir D, Basu S, Lo NC. 2021. Routine asymptomatic testing strategies for airline travel during the COVID-19 pandemic: a simulation study. *Lancet Infect Dis* 21: 929–938. [https://doi.org/10.1016/S1473-3099\(21\)00134-1](https://doi.org/10.1016/S1473-3099(21)00134-1).
- Grassly NC, Pons-Salort M, Parker EPK, White PJ, Ferguson NM, Imperial College COVID-19 Response Team. 2020. Comparison of molecular testing strategies for COVID-19 control: a mathematical modelling study. *Lancet Infect Dis* 20:1381–1389. [https://doi.org/10.1016/S1473-3099\(20\)30630-7](https://doi.org/10.1016/S1473-3099(20)30630-7).
- Larremore DB, Wilder B, Lester E, Shehata S, Burke JM, Hay JA, Tambe M, Mina MJ, Parker R. 2021. Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening. *Sci Adv* 7:eabd5393. <https://doi.org/10.1126/sciadv.abd5393>.
- Watson J, Whiting PF, Brush JE. 2020. Interpreting a covid-19 test result. *BMJ* 369:m1808.
- Arevalo-Rodriguez I, Buitrago-Garcia D, Simancas-Racines D, Zambrano-Achig P, Del Campo R, Ciapponi A, Sued O, Martinez-García L, Rutjes AW, Low N, Bossuyt PM, Perez-Molina JA, Zamora J. 2020. False-negative results of initial RT-PCR assays for COVID-19: a systematic review. *PLoS One* 15:e0242958. <https://doi.org/10.1371/journal.pone.0242958>.
- Wang W, Xu Y, Gao R, Lu R, Han K, Wu G, Tan W. 2020. Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA* 323:1843–1844.
- Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, Niemeyer D, Jones TC, Vollmar P, Rothe C, Hoelscher M, Bleicker T, Brünink S, Schneider J, Ehmann R, Zwirgmaier K, Drosten C, Wendtner C. 2020. Virological assessment of hospitalized patients with COVID-2019. *Nature* 581:465–469. <https://doi.org/10.1038/s41586-020-2196-x>.
- Kucirka LM, Lauer SA, Laeyendecker O, Boon D, Lessler J. 2020. Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based SARS-CoV-2 tests by time since exposure. *Ann Intern Med* 173: 262–267. <https://doi.org/10.7326/M20-1495>.
- Wikramaratna PS, Paton RS, Ghafari M, Lourenço J. 2020. Estimating the false-negative test probability of SARS-CoV-2 by RT-PCR. *Euro Surveill* 25: 2000568. <https://doi.org/10.2807/1560-7917.ES.2020.25.50.2000568>.
- Chung E, Chow EJ, Wilcox NC, Burstein R, Brandstetter E, Han PD, Fay K, Pfau B, Adler A, Lacombe K, Lockwood CM, Uyeki TM, Shendure J, Duchin JS, Rieder MJ, Nickerson DA, Boeckh M, Famulare M, Hughes JP, Starita LM, Bedford T, Englund JA, Chu HY. 2021. Comparison of symptoms and RNA levels in children and adults with SARS-CoV-2 infection in the community setting. *JAMA Pediatr* 175:e212025. <https://doi.org/10.1001/jamapediatrics.2021.2025>.
- Dramé M, Tabue Teguo M, Proye E, Hequet F, Hentzien M, Kanagaratnam L, Godaert L. 2020. Should RT-PCR be considered a gold standard in the diagnosis of COVID-19? *J Med Virol* 92:2312–2313. <https://doi.org/10.1002/jmv.25996>.
- Rindskopf D, Rindskopf W. 1986. The value of latent class analysis in medical diagnosis. *Stat Med* 5:21–27. <https://doi.org/10.1002/sim.4780050105>.
- Kostoulas P, Eusebi P, Hartnack S. 2020. Diagnostic accuracy estimates for COVID-19 RT-PCR and lateral flow immunoassay tests with Bayesian latent class models. *Res Square*.
- Butler-Laporte G, Lawandi A, Schiller I, Yao M, Dendukuri N, McDonald EG, Lee TC. 2021. Comparison of saliva and nasopharyngeal swab nucleic acid amplification testing for detection of SARS-CoV-2: a systematic review and meta-analysis. *JAMA Intern Med* 181:353–360. <https://doi.org/10.1001/jamainternmed.2020.8876>.

24. Hartnack S, Eusebi P, Kostoulas P. 2021. Bayesian latent class models to estimate diagnostic test accuracies of COVID-19 tests. *J Med Virol* 93: 639–640. <https://doi.org/10.1002/jmv.26405>.
25. Jones TC, Biele G, Mühlemann B, Veith T, Schneider J, Beheim-Schwarzbach J, Bleicker T, Tesch J, Schmidt ML, Sander LE, Kurth F, Menzel P, Schwarzer R, Zuchowski M, Hofmann J, Krumbholz A, Stein A, Edelmann A, Corman VM, Drosten C. 2021. Estimating infectiousness throughout SARS-CoV-2 infection course. *Science* 373:eabi5273. <https://doi.org/10.1126/science.abi5273>.
26. Kuwelker K, Zhou F, Blomberg B, Lartey S, Brokstad KA, Trieu MC, Bansal A, Madsen A, Krammer F, Mohn KG, Tøndel C, Linchusen DW, Cox RJ, Langeland N, Bergen COVID-19 Research Group. 2021. Attack rates amongst household members of outpatients with confirmed COVID-19 in Bergen, Norway: a case-ascertained study. *Lancet Reg Health Eur* 3: 100014. <https://doi.org/10.1016/j.lanpe.2020.100014>.
27. Wyllie AL, Fournier J, Casanovas-Massana A, Campbell M, Tokuyama M, Vijayakumar P, Warren JL, Geng B, Muenker MC, Moore AJ, Vogels CBF, Petrone ME, Ott IM, Lu P, Venkataraman A, Lu-Culligan A, Klein J, Earnest R, Simonov M, Datta R, Handoko R, Naushad N, Sewanan LR, Valdez J, White EB, Lapidus S, Kalinich CC, Jiang X, Kim DJ, Kudo E, Linehan M, Mao T, Moriyama M, Oh JE, Park A, Silva J, Song E, Takahashi T, Taura M, Weizman O-E, Wong P, Yang Y, Bermejo S, Odio CD, Omer SB, Dela Cruz CS, Farhadian S, Martinello RA, Iwasaki A, Grubaugh ND, Ko AI. 2020. Saliva or nasopharyngeal swab specimens for detection of SARS-CoV-2. *N Engl J Med* 383:1283–1286. <https://doi.org/10.1056/NEJMc2016359>.
28. Hanson KE, Barker AP, Hillyard DR, Gilmore N, Barrett JW, Orlandi RR, Shakir SM. 2020. Self-collected anterior nasal and saliva specimens versus health care worker-collected nasopharyngeal swabs for the molecular detection of SARS-CoV-2. *J Clin Microbiol* 58:e01824-20. <https://doi.org/10.1128/JCM.01824-20>.
29. Teo AKJ, Choudhury Y, Tan IB, Cher CY, Chew SH, Wan ZY, Cheng LTE, Oon LLE, Tan MH, Chan KS, Hsu LY. 2021. Saliva is more sensitive than nasopharyngeal or nasal swabs for diagnosis of asymptomatic and mild COVID-19 infection. *Sci Rep* 11:3134. <https://doi.org/10.1038/s41598-021-82787-z>.
30. Lee RA, Herigon JC, Benedetti A, Pollock NR, Denkinger CM. 2021. Performance of saliva, oropharyngeal swabs, and nasal swabs for SARS-CoV-2 molecular detection: a systematic review and meta-analysis. *J Clin Microbiol* 59:e02881-20. <https://doi.org/10.1128/JCM.02881-20>.
31. Pray IW. 2021. Performance of an antigen-based test for asymptomatic and symptomatic SARS-CoV-2 testing at two university campuses—Wisconsin, September–October 2020. *MMWR Morb Mortal Wkly Rep* 69: 1642–1647. <https://doi.org/10.15585/mmwr.mm695152a3>.
32. Smith RL, Gibson LL, Martinez PP, Ke R, Mirza A, Conte M, Gallagher N, Conte A, Wang L, Fredrickson R, Edmonson DC, Baughman ME, Chiu KK, Choi H, Jensen TW, Scardina KR, Bradley S, Gloss SL, Reinhart C, Yedetore J, Owens AN, Broach J, Barton B, Lazar P, Henness D, Young T, Dunnett A, Robinson ML, Mostafa HH, Pekosz A, Manabe YC, Heetderks WJ, McManus DD, Brooke CB. 2021. Longitudinal assessment of diagnostic test performance over the course of acute SARS-CoV-2 Infection. *J Infect Dis* 224: 976–982. <https://doi.org/10.1093/infdis/jiab337>.
33. Corman VM, Haage VC, Bleicker T, Schmidt ML, Mühlemann B, Zuchowski M, Jo WK, Tscheak P, Möncke-Buchner E, Müller MA, Krumbholz A, Drexler JF, Drosten C. 2021. Comparison of seven commercial SARS-CoV-2 rapid point-of-care antigen tests: a single-centre laboratory evaluation study. *Lancet Microbe* 2:e311–e319. [https://doi.org/10.1016/S2666-5247\(21\)00056-2](https://doi.org/10.1016/S2666-5247(21)00056-2).
34. Mina MJ, Parker R, Larremore DB. 2020. Rethinking Covid-19 test sensitivity—a strategy for containment. *N Engl J Med* 383:e120. <https://doi.org/10.1056/NEJMp2025631>.
35. Elliott J, Whitaker M, Bodinier B, Eales O, Riley S, Ward H, Cooke G, Darzi A, Chadeau-Hyam M, Elliott P. 2021. Predictive symptoms for COVID-19 in the community: REACT-1 study of over 1 million people. *PLoS Med* 18: e1003777. <https://doi.org/10.1371/journal.pmed.1003777>.
36. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, Lau YC, Wong JY, Guan Y, Tan X, Mo X, Chen Y, Liao B, Chen W, Hu F, Zhang Q, Zhong M, Wu Y, Zhao L, Zhang F, Cowling BJ, Li F, Leung GM. 2020. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med* 26:672–675. <https://doi.org/10.1038/s41591-020-0869-5>.
37. Christensen H, Turner K, Trickey A, Booton RD, Hemani G, Nixon E, Relton C, Danon L, Hickman M, Brooks-Pollock E, Part of the University of Bristol UNCOVER Group. 2020. COVID-19 transmission in a university setting: a rapid review of modelling studies. medRxiv. <https://doi.org/10.1101/2020.09.07.20189688>.
38. Hamer DH, White LF, Jenkins HE, Gill CJ, Landsberg HE, Klapperich C, Bulekova K, Platt J, Decarie L, Gilmore W, Pilkington M, MacDowell TL, Faria MA, Densmore D, Landaverde L, Li W, Rose T, Burgay SP, Miller C, Doucette-Stamm L, Lockard K, Elmore K, Schroeder T, Zaia AM, Kolaczyk ED, Waters G, Brown RA. 2021. Assessment of a COVID-19 control plan on an urban university campus during a second wave of the pandemic. *JAMA Netw Open* 4:e2116425. <https://doi.org/10.1001/jamanetworkopen.2021.16425>.
39. Motta FC, McGoff KA, Deckard A, Wolfe CR, Bonsignori M, Moody MA, Cavanaugh K, Denny TN, Harer J, Haase SB. 2021. Assessment of simulated surveillance testing and quarantine in a SARS-CoV-2–vaccinated population of students on a university campus. *JAMA Health Forum* 2: e213035. <https://doi.org/10.1001/jamahealthforum.2021.3035>.
40. Cavany S, Bivins A, Wu Z, North D, Bibby K, Perkins TA. 2021. Inferring SARS-CoV-2 RNA shedding into wastewater relative to time of infection. medRxiv. <https://doi.org/10.1101/2021.06.03.21258238>.
41. Ranoa DRE, Holland RL, Alnaji FG, Green KJ, Wang L, Brooke CB, Burke MD, Fan TM, Hergenrother PJ. 2020. Saliva-based molecular testing for SARS-CoV-2 that bypasses RNA extraction. bioRxiv. <https://doi.org/10.1101/2020.06.18.159434>.
42. Singanayagam A, Patel M, Charlett A, Lopez Bernal J, Saliba V, Ellis J, Ladhani S, Zambon M, Gopal R. 2020. Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. *Euro Surveill* 25:2001483. <https://doi.org/10.2807/1560-7917.ES.2020.25.32.2001483>.
43. Kasper MR, Geibe JR, Sears CL, Riegodedios AJ, Luse T, Von Thun AM, McGinnis MB, Olson N, Houskamp D, Fenequito R, Burgess TH, Armstrong AW, DeLong G, Hawkins RJ, Gillingham BL. 2020. An outbreak of Covid-19 on an aircraft carrier. *N Engl J Med* 383:2417–2426. <https://doi.org/10.1056/NEJMoa2019375>.
44. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich NG, Lessler J. 2020. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med* 172:577–582. <https://doi.org/10.7326/M20-0504>.