JAMA Internal Medicine | **Original Investigation**

Comparison of Saliva and Nasopharyngeal Swab Nucleic Acid Amplification Testing for Detection of SARS-CoV-2
A Systematic Review and Meta-analysis

Guillaume Butler-Laporte, MD; Alexander Lawandi, MD, MSc; Ian Schiller, MSc; Mandy C. Yao, MSc; Nandini Dendukuri, PhD; Emily G. McDonald, MD, MSc; Todd C. Lee, MD, MPH

+ **Supplemental content**

IMPORTANCE Nasopharyngeal swab nucleic acid amplification testing (NAAT) is the noninvasive criterion standard for diagnosis of coronavirus disease 2019 (COVID-19). However, it requires trained personnel, limiting its availability. Saliva NAAT represents an attractive alternative, but its diagnostic performance is unclear.

OBJECTIVE To assess the diagnostic accuracy of saliva NAAT for COVID-19.

DATA SOURCES In this systematic review, a search of the MEDLINE and medRxiv databases was conducted on August 29, 2020, to find studies of diagnostic test accuracy. The final meta-analysis was performed on November 17, 2020.

STUDY SELECTION Studies needed to provide enough data to measure salivary NAAT sensitivity and specificity compared with imperfect nasopharyngeal swab NAAT as a reference test. An imperfect reference test does not perfectly reflect the truth (ie, it can give false results). Studies were excluded if the sample contained fewer than 20 participants or was neither random nor consecutive. The Quality Assessment of Diagnostic Accuracy Studies 2 tool was used to assess the risk of bias.

DATA EXTRACTION AND SYNTHESIS Preferred Reporting Items for Systematic Reviews and Meta-analyses reporting guideline was followed for the systematic review, with multiple authors involved at each stage of the review. To account for the imperfect reference test sensitivity, we used a bayesian latent class bivariate model for the meta-analysis.

MAIN OUTCOMES AND MEASURES The primary outcome was pooled sensitivity and specificity. Two secondary analyses were performed: one restricted to peer-reviewed studies, and a post hoc analysis limited to ambulatory settings.

RESULTS The search strategy yielded 385 references, and 16 unique studies were identified for quantitative synthesis. Eight peer-reviewed studies and 8 preprints were included in the meta-analyses (5922 unique patients). There was significant variability in patient selection, study design, and stage of illness at which patients were enrolled. Fifteen studies included ambulatory patients, and 9 exclusively enrolled from an outpatient population with mild or no symptoms. In the primary analysis, the saliva NAAT pooled sensitivity was 83.2% (95% credible interval [CrI], 74.7%-91.4%) and the pooled specificity was 99.2% (95% CrI, 98.2%-99.8%). The nasopharyngeal swab NAAT had a sensitivity of 84.8% (95% CrI, 76.8%-92.4%) and a specificity of 98.9% (95% CrI, 97.4%-99.8%). Results were similar in secondary analyses.

CONCLUSIONS AND RELEVANCE These results suggest that saliva NAAT diagnostic accuracy is similar to that of nasopharyngeal swab NAAT, especially in the ambulatory setting. These findings support larger-scale research on the use of saliva NAAT as an alternative to nasopharyngeal swabs.

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Guillaume Butler-Laporte, MD, McGill University Health Centre, Royal Victoria Hospital, 1001 Decarie Blvd, Room E5.1820, Montréal, QC H4A 3J1, Canada (guillaume.butler-laporte@mail.mcgill.ca).

T esting is the cornerstone of a successful public health response to coronavirus disease 2019 (COVID-19). Nasopharyngeal sampling for nucleic acid amplification testing (NAAT) is the current noninvasive criterion standard diagnostic test. However, nasopharyngeal testing requires trained personnel and handling of a specially designed swab, and the technique cannot be easily or reliably performed in all populations (eg, children and quarantined individuals).[1] A test that can be self-administered would greatly increase the options for case identification in the community. Consequently, alternative sampling sites are being explored to allow broader deployment of testing, increased access in remote regions, and testing among underserved populations who may have limited access to highly qualified testing personnel. Of the currently available routes for testing, salivary testing likely represents the most practical option because it avoids the inconvenience, discomfort, and required technical expertise of nasopharyngeal sampling and could potentially be broadly deployed in the community. To date, the results of studies on the diagnostic performance of saliva NAAT are conflicting.[2,3] Furthermore, other studies have relied on the assumption that the sensitivity and specificity of the nasopharyngeal swab is perfect, which is not the case.[4] To summarize the evidence on saliva NAAT performance, we conducted a systematic review of the operating characteristics of saliva NAAT for detection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the diagnosis of COVID-19 using nasopharyngeal NAAT as the comparator. We performed a bayesian latent class meta-analysis that adjusts for the imperfect accuracy of the comparator.

Methods

Search Strategy and Review

This study followed the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) reporting guideline[5] for meta-analysis of diagnostic test accuracy. The protocol was registered on PROSPERO.[6]

On August 29, 2020, we used the PubMed tool to search the MEDLINE database using the following terms: (*novel coronavirus* OR *SARS-CoV-2* OR *COVID19* OR *COVID-19*) and (*PCR* OR *polymerase chain reaction* OR *NAAT* OR *nucleic acid amplification testing* OR *detection*) and (*saliva* OR *oral swabs* OR *salivary*). On the same day, we used the following search terms on MedRxiv for applicable preprints: (*SARS-Cov-2* OR *COVID-19*) and (*PCR* OR *polymerase chain reaction* OR *NAAT* OR *nucleic acid amplification testing*) and *saliva*. The MedRxiv search was simplified owing to the limited search functionality compared with PubMed. If preprints were also published in a peer-reviewed journal, we used the peer-reviewed version. We included studies that directly compared saliva and nasopharyngeal NAAT for the detection of SARS-CoV-2 and provided enough data to construct the 2 × 2 diagnostic table. Given that some studies used a combination or nasopharyngeal swab and oropharyngeal or oral swabs for diagnosis in routine clinical activities, we also allowed this composite test as a criterion standard. Allowing for the use of this composite refer-

Key Points

Question Is saliva nucleic acid amplification testing (NAAT) comparable to nasopharyngeal NAAT, the current noninvasive criterion standard test for diagnosis of coronavirus disease 2019?

Findings In this systematic review and latent class meta-analysis adjusting for the imperfect reference standard, saliva NAAT had a similar sensitivity and specificity to that of nasopharyngeal NAAT.

Meaning Given the ease of use and good diagnostic performances, these findings suggest that saliva NAAT represents an attractive alternative to nasopharyngeal swab NAAT and may significantly bolster massive testing efforts.

ence test could make saliva NAAT appear less sensitive, hence making our comparison more conservative. Similarly, we allowed for studies using different NAAT platforms so long as the reference was used for routine clinical microbiology diagnostic purposes, ensuring its validity as a comparator. We excluded studies without random or consecutive samples. Owing to well-recognized concerns over publication bias and inflated effect estimates inherent to smaller studies,[7,8] we further excluded studies with fewer than 20 patients. This threshold was prespecified before our search. Case-control studies were likewise excluded. Finally, we screened references from all included studies for other potential studies to include in the meta-analysis.

Studies were first reviewed by title, then by abstract, and finally by full-text review by 2 authors with experience in systematic reviews (G.B.-L. and T.C.L.). For studies selected after full-text review, we extracted raw data on true-positive, false-positive, true-negative, and false-negative test results for saliva vs nasopharyngeal NAAT. Using a prewritten data extraction chart, we also retrieved information on study design, country, age and sex of patients, proportion of symptomatic patients, and NAAT platforms used. Study selection, review, and data extraction were performed by 2 authors (G.B.-L. and T.C.L.), with disagreements resolved by consensus.

Quality Assessment

Two authors (G.B.-L. and A.L.) used the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool[9] to assess for the risk of bias in each selected study, with disagreements resolved by consensus. The QUADAS-2 tool evaluates 4 bias categories: patient selection, index test, reference test, and flow and timing of testing. It also grades the risk that this bias decreases the applicability of the results in the target population (eg, patients being tested for SARS-CoV-2). Studies that were considered at high risk of bias were not included in the analysis.

Analysis

In the primary meta-analysis, we pooled assay sensitivities and specificities from the studies selected. As a secondary prespecified meta-analysis, we limited the analysis to only peer-reviewed studies. Finally, after finishing the data extraction and risk of bias assessment, we decided to perform a post hoc

meta-analysis restricted to studies performed in ambulatory patients and that did not allow repeated testing on multiple days. This analysis was performed to better assess diagnostic accuracy in a setting where saliva sampling is likely to have the greatest effect on public health policy.

To address the fact that the nasopharyngeal NAAT represents an imperfect reference standard, we used a latent class bayesian bivariate random-effects model.[10] This model accounts for the imperfect reference test (nasopharyngeal swab NAAT) by estimating the true disease status of each participant using results from both the index (saliva NAAT) and the reference test. Hence, as opposed to traditional bivariate models, this model does not require that the reference test have a perfect (or near perfect) accuracy. For example, in the unlikely scenario where no tests presented a false-positive result, the latent class model would use a composite test to determine the true SARS-CoV-2 test result. Conditional dependence (ie, dependence that may arise between the tests if they are prone to making the same false-positive or false-negative errors) is also accounted for. The latent class model was also parameterized using a random-effects approach, allowing for correction owing to heterogeneity in sensitivity and specificity between the included studies.

Through the model, sensitivity and specificity of saliva NAAT are then obtained using the true disease estimate as the reference. For each analysis, we present sensitivity and specificity estimates using the latent true disease status as the criterion standard, as well as pooled sensitivity and specificity. We also computed 95% credible intervals (CrIs) for each result, and an overall 95% prediction interval[11,12] for pooled sensitivity and specificity. Briefly, in a bayesian framework, the 95% CrI for a parameter is the interval within which it is estimated to fall, with a probability of 95%. The 95% prediction interval is the interval within which the parameter obtained in a future (still unmeasured) study, performed in similar populations, would fall with 95% probability. It therefore provides a measure of between-study heterogeneity as well as a measure of uncertainty about the applicability of our results in the studied population. The CrI is therefore always contained within the prediction interval. When between-study heterogeneity is low, the CrI and the prediction interval are expected to be similar. However, when between-study heterogeneity is large, the prediction interval widens. This highlights the fact that unmeasured results from a new study can differ significantly from the pooled estimate, which can suggest additional unmeasured factors explaining differences in diagnostic accuracy that may not be accounted for.

A sample from the posterior distributions of the parameters of interest was obtained using Monte Carlo Markov Chain methods implemented using the JAGS software[13] packages (version 4.10) accessed through the R statistical environment (version 4.0.2 [R Foundation for Statistical Computing]).[14] The programs were accessed through an app on R Shiny, version 1.5.0, written by 3 authors (M.C.Y., I.S., and N.D.) (https://bayesdta.shinyapps.io/meta-analysis/). We used low-information prior distribution on each parameter. For true COVID-19 prevalence in each study, we used a uniform (0, 1) prior distribution. For the hyper prior distribution for the index test sensitivity and specificity (on the logit scale), we used a normal distribution, with mean of 0 and variance of 100. For the between-study variance parameters, we used a gamma prior distribution with a shape parameter of 2 and a rate parameter of 0.5. Finally, we used a uniform (−1, 1) prior distribution for the correlation term between test sensitivity and specificity to account for a threshold effect. Three Gibbs sampling chains were used, with 40 000 iterations each. Pooled estimates were obtained by removing the first 10 000 iterations from each chain and thinning the remaining 30 000 by a factor of 10. Convergence was assessed by visual inspection of each Markov chain, with Brooks-Gelman-Rubin statistics (R-hat <1.1), and by comparing results of Markov chains using different randomly generated starting values.

As a comparator, we also performed a usual bayesian bivariate random-effects model analysis. This approach accounts for both within- and between-study variation in sensitivity and specificity and for the correlation between sensitivity and specificity across studies. However, it does not account for the imperfect reference test.

Results

Review

Our MEDLINE and MedRxiv search yielded 129 and 256 studies, respectively. Only 7 studies were published before 2020, all from the MEDLINE search, suggesting that our search was specific to COVID-19. We initially retained 62 studies for full-text review, including 7 duplicate preprints that were subsequently published in peer-reviewed journals. We identified no further studies from reference parsing. After full-text review, we retained 16 studies for data extraction.[3,15-29] The most common reason for exclusion was insufficient information to build the 2 × 2 diagnostic accuracy table (10 of 39 studies), followed by the presence of review articles (7 of 39). Eight of the studies[15-22] were from the MedRxiv preprint server, whereas the rest[3,23-29] were from peer-reviewed journals. However, during our review process, 2 of these studies were published in peer-reviewed literature[18,21] and were therefore included in the analysis restricted to peer-reviewed literature. The PRISMA flow diagram is shown in **Figure 1**.

We ultimately retained 16 studies[3,15-29] for the quantitative analysis. From these, the mean (SD) number of sample pairs was 370 (616), and the median number of sample pairs was 117 (interquartile range, 88-256). The difference is explained by 2 studies[21,23] of more than 1900 samples that yielded a high proportion of concordant negative test results (96%) for saliva and nasopharyngeal swab NAAT. Both studies were performed in the context of mass population screening, explaining the low disease prevalence found.

The studies varied by setting and patient population. Although the US was the most highly represented country with 6 studies,[16,18,22,24,25,28] we also found studies from Europe,[15,17] Asia,[3,19,21,26,27] and Australia.[29] Patients were selected from drive-through testing clinics,[24] at hospital admission,[15,20,25] through a laboratory information system,[26] through screening campaigns in returning travelers,[21] or participants were mi-

Figure 1. PRISMA Flow Diagram
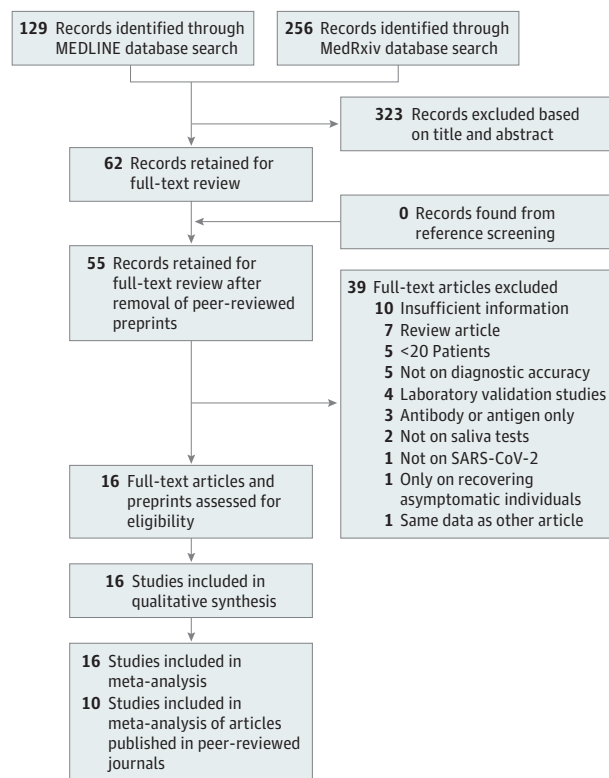
SARS-CoV-2 indicates severe acute respiratory syndrome coronavirus 2.

grant workers.[19] Five studies[3,15,17,27,28] required symptoms compatible with COVID-19 for enrollment. Given that local indications for testing evolved rapidly during the pandemic, some studies[3,18,23] only enrolled patients who fit certain criteria for symptoms or with a known exposure to a contact case. One study[20] enrolled both patients and health care workers, although the exact inclusion and exclusion criteria were unclear. Finally, 1 study[15] enrolled hospitalized patients, whereas 10 studies[3,16,18,19,21-24,28,29] enrolled patients exclusively tested in an ambulatory setting.

Testing methods also differed between studies. All laboratories used different NAAT assays, but the assays used for the reference standard of nasopharyngeal swab NAAT had all been previously validated by the manufacturer, or in-house by the clinical laboratory in cases of home-brew tests developed by local clinical laboratories. The reference standard used was always that used by the clinical laboratory for diagnosis of COVID-19 in usual clinical practice. Moreover, apart from 2 studies,[20,23] all cohorts used the same assay to compare the saliva and the nasopharyngeal swab NAATs (other than the different sample types). Furthermore, 1 study[21] used 2 different assays, but each pair of saliva-nasopharyngeal swab samples from each individual patient was analyzed using the same assay. Of note, 2 studies[16,26] compared results from different NAAT assays on the same patients. For this meta-analysis, we selected the most sensitive of the compared assays (ie, the assay with the highest number of positive results using saliva

NAAT). In 1 study,[16] the authors also included an indeterminate category, which we merged with the positive results for the purposes of this meta-analysis. Two studies[19,26] allowed for repeated tests on the same patients to be performed on separate days, and it was impossible to remove these duplicates from our meta-analysis. These studies also had the 2 highest rates of discordant results between tests with positive saliva findings and negative nasopharyngeal swab findings. Finally, 4 studies[3,15,17,23] used a composite test that included nasopharyngeal swab in combination with either the oral or the oropharyngeal swab as the reference test. The details of the included studies are found in eTable 1 in the Supplement.

Quality Assessment
The QUADAS-2 tool highlights the results mentioned above. Whereas the index and reference tests were unlikely to contribute bias to our analysis, patient selection methods varied between studies, which may have provided biased estimates of sensitivity and specificity. As such, all studies scored a high or unclear risk of bias owing to patient selection, but most of the studies scored as low risk in other categories. The QUADAS-2 graphical summary is shown in **Figure 2**, and the tabular summary is available in eTable 2 in the Supplement.

Meta-analysis
In total, we included 5922 samples in the primary meta-analysis, of which 941 had a positive result (by saliva NAAT, nasopharyngeal swab NAAT, or both), and 4981 had a concordant negative result (by saliva and nasopharyngeal swab NAAT). In the primary analysis, we obtained a pooled saliva sensitivity of 83.2% (95% CrI, 77.4%-91.4%) and a pooled saliva specificity of 99.2% (95% CrI, 98.2%-99.8%). Individual study-estimated sensitivity and specificity (accounting for each study's imperfect reference test) can be found in **Figure 3**, and **Figure 4** shows the resulting summary receiver-operator curve. Nasopharyngeal swab NAAT performed similarly with a pooled sensitivity of 84.8% (95% CrI, 76.8%-92.4%) and a pooled specificity of 98.9% (95% CrI, 97.4%-99.8%).

By comparison, the traditional bivariate model (which assumes a perfect reference test) yielded similar sensitivity for the saliva NAAT (86%; 95% CrI, 80%-92%), but with a slight decrease in specificity (97%; 95% CrI, 93%-99%) for saliva NAAT. The prediction interval was also considerably larger and reached values of specificity as low as 40% (eFigure 1 in the Supplement). This is explained by the fact that cases with positive findings for the saliva NAAT and negative findings for the nasopharyngeal swab NAAT would worsen the estimate of saliva NAAT specificity using traditional methods.

The secondary analysis, restricted to peer-reviewed literature (eFigure 2 in the Supplement), found a pooled sensitivity for saliva NAAT of 85.6% (95% CrI, 77.0%-92.7%) and a pooled saliva specificity of 99.1% (95% CrI, 98.0%-99.8%). The nasopharyngeal swab NAAT had a sensitivity of 85.7% (95% CrI, 76.5%-93.4%) and a specificity of 98.9% (95% CrI, 97.4%-99.7%).

Finally, we performed a post hoc meta-analysis limited to the ambulatory setting, after removal of studies that allowed for repeated testing on multiple days. We retained 9 studies[3,16,18,21-24,28,29] for this meta-analysis, for a total of 4851

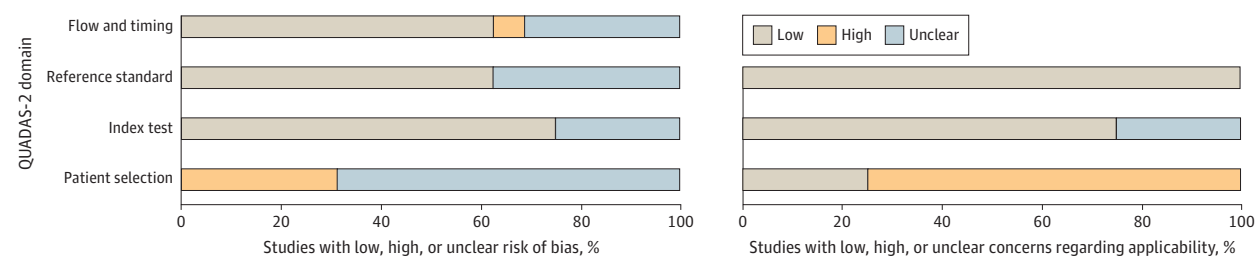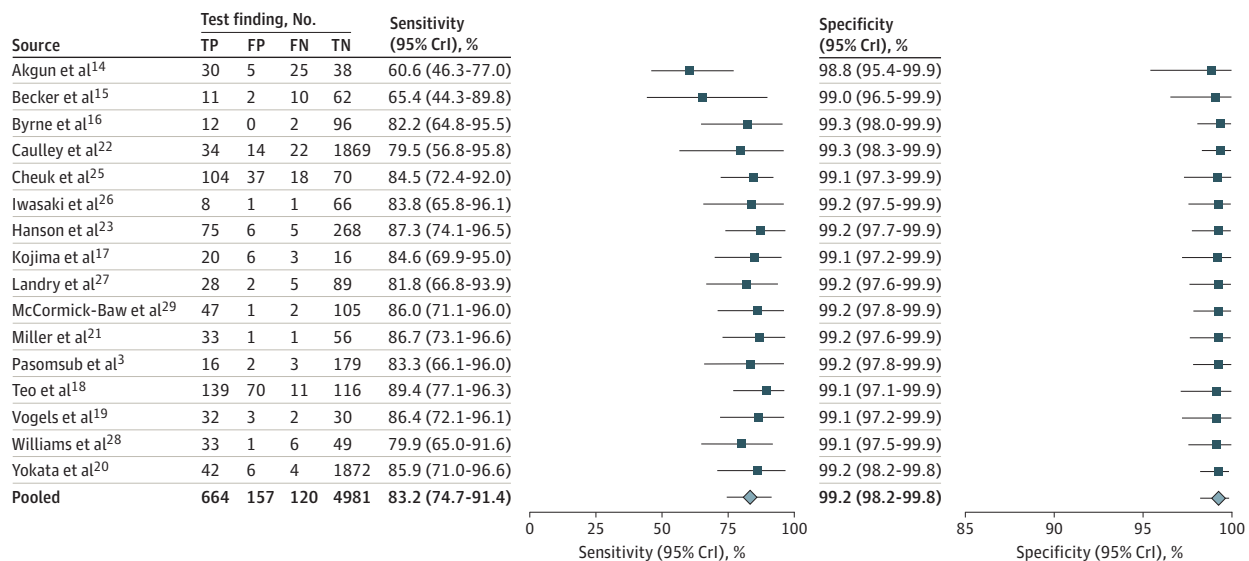Figure 2. Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) Study Quality Summary

Figure 3. Primary Meta-analysis Results for the Detection of Severe Acute Respiratory Syndrome Coronavirus 2 in Saliva Samples

Source	TP	FP	FN	TN	Sensitivity (95% CrI), %		Specificity (95% CrI), %	
Akgun et al[14]	30	5	25	38	60.6 (46.3-77.0)		98.8 (95.4-99.9)	
Becker et al[15]	11	2	10	62	65.4 (44.3-89.8)		99.0 (96.5-99.9)	
Byrne et al[16]	12	0	2	96	82.2 (64.8-95.5)		99.3 (98.0-99.9)	
Caulley et al[22]	34	14	22	1869	79.5 (56.8-95.8)		99.3 (98.3-99.9)	
Cheuk et al[25]	104	37	18	70	84.5 (72.4-92.0)		99.1 (97.3-99.9)	
Iwasaki et al[26]	8	1	1	66	83.8 (65.8-96.1)		99.2 (97.5-99.9)	
Hanson et al[23]	75	6	5	268	87.3 (74.1-96.5)		99.2 (97.7-99.9)	
Kojima et al[17]	20	6	3	16	84.6 (69.9-95.0)		99.1 (97.2-99.9)	
Landry et al[27]	28	2	5	89	81.8 (66.8-93.9)		99.2 (97.6-99.9)	
McCormick-Baw et al[29]	47	1	2	105	86.0 (71.1-96.0)		99.2 (97.8-99.9)	
Miller et al[21]	33	1	1	56	86.7 (73.1-96.6)		99.2 (97.6-99.9)	
Pasomsub et al[3]	16	2	3	179	83.3 (66.1-96.0)		99.2 (97.8-99.9)	
Teo et al[18]	139	70	11	116	89.4 (77.1-96.3)		99.1 (97.1-99.9)	
Vogels et al[19]	32	3	2	30	86.4 (72.1-96.1)		99.1 (97.2-99.9)	
Williams et al[28]	33	1	6	49	79.9 (65.0-91.6)		99.1 (97.5-99.9)	
Yokata et al[20]	42	6	4	1872	85.9 (71.0-96.6)		99.2 (98.2-99.8)	
Pooled	664	157	120	4981	83.2 (74.7-91.4)		99.2 (98.2-99.8)	

Sensitivity and specificity estimates (with their 95% credible intervals [CrI]) are obtained from the latent class bivariate bayesian model. The true-positive (TP), false-positive (FP), false-negative (FN), and true-negative (TN) counts in the sensitivity graph are with respect to the nasopharyngeal swab nucleic acid amplification testing. Note the y-axis scale difference between the 2 forest plots. Diamond indicates pooled data.

patients, including 391 (8.1%) with a positive test result (saliva, nasopharyngeal swab, or both). Pooled saliva NAAT estimates remained similar (**Figure 5**) with a sensitivity of 84.5% (95% CrI, 73.0%-95.3%) and a specificity of 99.0% (95% CrI, 97.7%-99.7%). For nasopharyngeal swab NAAT, the sensitivity was 88.0% (95% CrI, 77.5%-95.8%) and the specificity was 98.7% (95% CrI, 96.2%-99.8%). The summary receiver-operator curve was similar to that of the primary analysis (eFigure 3 in the Supplement).

Detailed results and the summary receiver-operator curves for the meta-analysis on peer-reviewed studies can be found in eFigure 2 in the Supplement. Forest plots for the diagnostic accuracy of the nasopharyngeal swabs can be found in eFigure 4 in the Supplement. Gibbs sample chains and posterior densities for all analyses can be found in eFigures 5 to 7 in the Supplement.

Discussion

Our meta-analysis found that the diagnostic sensitivity for saliva NAAT is approximately 83.2% (95% CrI, 74.7%-91.4%), which

is comparable to that reported for nasopharyngeal swab NAAT[4] and to the result obtained using our latent class model analysis (84.8%; 95% CrI, 76.8%-92.4%). Given the ease of sample procurement and increased patient comfort, testing centers should strongly consider adopting saliva as their first sample choice, especially in community mass screening programs.

Strengths and Limitations

Our study has several strengths. Our search strategy was sensitive, because no additional studies were identified from reference screening or by other means than our primary search strategy. We also addressed a highly clinically relevant question that could potentially rapidly affect global public health policy for testing strategies. It is incredibly important to validate this sampling technique for possible deployment in countries or communities with continually high case rates and especially for those with developing health care systems and less access to specialized care. An increasingly large number of published studies on the subject suggest that a reliable alternative to nasopharyngeal swab NAAT is clearly being sought.[30] Finally, we used a meta-analysis that accounts for the imper-

Figure 4. Summary Receiver-Operator Curve for the Primary Meta-analysis of Detection of Severe Acute Respiratory Syndrome Coronavirus 2 in Saliva Samples
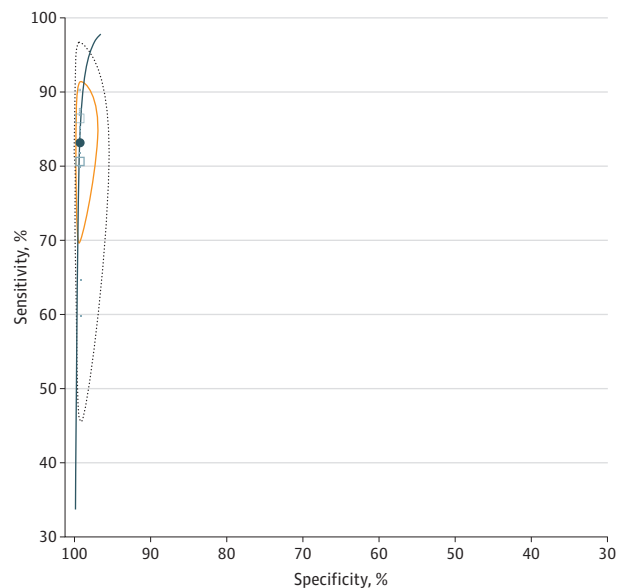
Each study is represented by a box of size proportional to the number of samples tested. Some boxes are small and overlap. The orange line is the saliva nucleic acid amplification testing (NAAT) 95% credible interval. The dashed line is the 95% saliva NAAT prediction interval, and the dark solid line is the summary receiver-operator curve. The dark circle is the pooled sensitivity and specificity estimate. The dark solid line is the summary receiver-operating characteristic curve.

fect reference standard. The standard meta-analysis methods would have provided negatively biased estimates of the true specificity of the saliva NAAT, with a larger amount of uncertainty (as can be seen from the prediction intervals). Accounting for this, we were able to obtain a more precise estimate of saliva NAAT diagnostic accuracy.

Our results are limited by the heterogeneity of the included studies, which varied in terms of the study population and the timing of testing. For example, hospitalized and critically ill patients were significantly underrepresented compared with studies performed in the community, and so our results may not apply in this setting. Similarly, many studies compared saliva and nasopharyngeal swabs later in the disease course, and only a few were able to obtain paired samples on the first day of presentation. Two studies[19,26] also allowed for repeated samples on multiple days, and correlation between samples may have been imperfectly accounted for, which might have biased our results. However, after removal of these studies in the post hoc analysis, we obtained similar pooled sensitivity and specificity estimates. Furthermore, although each study used assays with similar technology that were all used in clinical practice, small differences between the assays and their targets could have contributed to heterogeneity.

Of note, most studies did not provide details on patient symptoms, and it remains unclear whether certain clinical syndromes (eg, predominantly lower respiratory tract disease or critical illness) may warrant a specific specimen type for optimal

NAAT diagnostic accuracy. Such differences may explain the wide prediction intervals obtained in the traditional meta-analysis (without adjustment for the imperfect reference test), and some caution is warranted in applying the findings of our study to populations that were underrepresented in our meta-analysis. Nevertheless, all latent class meta-analyses showed much less uncertainty in the pooled diagnostic accuracy, including the post hoc analysis in the ambulatory patient population. Hence, despite these limitations, we believe that saliva sampling is a reasonable alternative to nasopharyngeal swabs, especially in community testing centers.

The timing of testing also deserves careful consideration; it is plausible that the timing of testing may have affected the diagnostic accuracy of saliva and nasopharyngeal swab testing differentially. In this review, studies with the highest discordant rates of positive findings for saliva and negative findings for nasopharyngeal swab paired samples included patients with repeated tests performed on multiple days. This finding suggests that saliva NAAT results may remain positive for longer than nasopharyngeal swab NAAT results. In fact, the 1 study[31] we excluded from our meta-analysis, performed exclusively in patients who had clinically recovered from proven COVID-19, found a high rate of positive findings for saliva and negative findings for nasopharyngeal swab paired samples. This hypothesis has been supported by other studies that compared viral load in saliva and nasopharyngeal specimens[22,32,33] and found that more viral RNA could be amplified from saliva than nasopharyngeal swabs and for longer periods. Given that NAAT does not necessarily imply presence of live virus,[34] more research is needed to determine the role of saliva NAAT in following viral shedding time, and whether this approach overestimates the duration of infectiousness.

Last, other limitations of saliva NAAT may emerge with time and more widespread use. We therefore caution that any large-scale deployment of saliva NAAT should be accompanied by an ongoing rigorous quality control program. Despite these limitations, community testing of presymptomatic or mildly symptomatic persons was well represented in our analyses, and a subgroup analysis limited to this population also demonstrated excellent diagnostic accuracy, implying that saliva NAAT could be deployed in a community setting with a fair level of confidence. Saliva testing could be particularly useful for larger screening efforts or in patients who are reliable enough to self-report to a testing center (eg, health care workers). Our findings open the door for patient self-collected tests, which would make self-isolation easier to implement and increase community research capacity. In theory, because saliva samples are self-collected, this could also reduce occupational exposures for personnel collecting samples. This sampling strategy could also conceivably facilitate community prophylaxis and early treatment studies of COVID-19.[35] Importantly, saliva NAAT allows for large-scale testing in high-risk populations that are difficult to reach using hospital-based approaches, an approach that has been deemed to be cost-effective when using nasopharyngeal NAAT. The use of techniques to address inefficiencies, such as saliva testing or pooling of test results, could reduce costs by as much as 40% and personnel by 20%.[36] These populations include college stu-

Figure 5. Post hoc Meta-analysis Results for the Detection of Severe Acute Respiratory Syndrome Coronavirus 2 in Saliva Samples From Ambulatory Patients
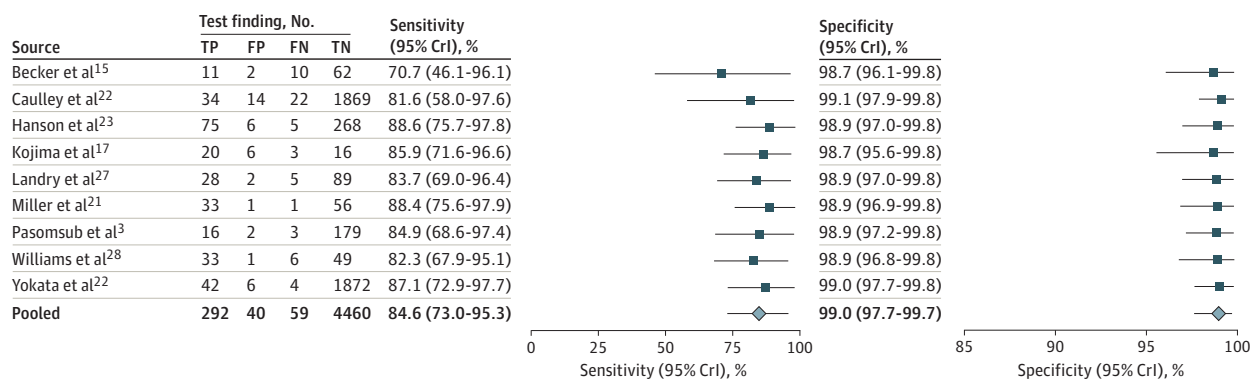
Source	Test finding, No.				Sensitivity (95% CrI), %	Specificity (95% CrI), %
	TP	FP	FN	TN		
Becker et al[15]	11	2	10	62	70.7 (46.1-96.1)	98.7 (96.1-99.8)
Caulley et al[22]	34	14	22	1869	81.6 (58.0-97.6)	99.1 (97.9-99.8)
Hanson et al[23]	75	6	5	268	88.6 (75.7-97.8)	98.9 (97.0-99.8)
Kojima et al[17]	20	6	3	16	85.9 (71.6-96.6)	98.7 (95.6-99.8)
Landry et al[27]	28	2	5	89	83.7 (69.0-96.4)	98.9 (97.0-99.8)
Miller et al[21]	33	1	1	56	88.4 (75.6-97.9)	98.9 (96.9-99.8)
Pasomsub et al[3]	16	2	3	179	84.9 (68.6-97.4)	98.9 (97.2-99.8)
Williams et al[28]	33	1	6	49	82.3 (67.9-95.1)	98.9 (96.8-99.8)
Yokata et al[22]	42	6	4	1872	87.1 (72.9-97.7)	99.0 (97.7-99.8)
Pooled	292	40	59	4460	84.6 (73.0-95.3)	99.0 (97.7-99.7)

Sensitivity and specificity estimates (with their 95% credible intervals [CrI]) are obtained from the latent class bivariate bayesian model. The true-positive (TP), false-positive (FP), false-negative (FN), and true-negative (TN) counts in the sensitivity graph are with respect to the nasopharyngeal swab nucleic acid amplification testing. Note the y-axis scale difference between the 2 forest plots. Diamond indicates pooled data.

dents and children, 2 populations for whom lockdown causes significant harm,[37] but that are also likely to catalyze the spread of the virus.[38] However, these populations and such a testing strategy warrant dedicated studies to better assess the role of saliva NAAT as a public health intervention.

Conclusions

During a pandemic, when the pretest probability of infection is elevated and lack of exposure is unreliable in ruling out COVID-19, large community-based testing programs are required for an effective public health response. Although some questions remain about saliva NAAT diagnostic accuracy in certain populations and settings, especially hospitalized and critically ill patients, saliva NAAT yields a sensitivity and specificity comparable to nasopharyngeal swab NAAT in ambulatory patients presenting with minimal or mild symptoms. Saliva NAAT should therefore be prioritized for larger-scale deployment with prospective studies conducted by clinical microbiology laboratories and public health authorities.

REFERENCES

1. Marty FM, Chen K, Verrill KA. How to obtain a nasopharyngeal swab specimen. *N Engl J Med*. 2020;382(22):e76. doi:10.1056/NEJMvcm2010260

2. Jamal AJ, Mozafarihashjin M, Coomes E, et al; Toronto Invasive Bacterial Diseases Network COVID-19 Investigators. Sensitivity of nasopharyngeal swabs and saliva for the detection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Clin Infect Dis*. 2020;ciaa848. doi:10.1093/cid/ciaa848

3. Pasomsub E, Watcharananan SP, Boonyawat K, et al. Saliva sample as a non-invasive specimen for the diagnosis of coronavirus disease 2019: a cross-sectional study. *Clin Microbiol Infect*. 2020; S1198-743X(20)30278-0. Published online May 15, 2020. doi:10.1016/j.cmi.2020.05.001

4. Ridgway JP, Pisano J, Landon E, Beavis KG, Robicsek A. Clinical sensitivity of severe acute respiratory syndrome coronavirus 2 nucleic acid amplification tests for diagnosing coronavirus disease 2019. *Open Forum Infect Dis*. 2020;7(8):a315. doi:10.1093/ofid/ofaa315

5. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and

elaboration. *BMJ*. 2009;339:b2700. doi:10.1136/bmj.b2700

6. Butler-Laporte G, Lawandi A, McDonald E, Lee T. Saliva PCR for the diagnosis of SARS-CoV-2: a systematic review and meta-analysis. PROSPERO 2020 CRD42020206551. Accessed September 15, 2020. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020206551

7. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med*. 2001;135(11):982-989. doi:10.7326/0003-4819-135-11-200112040-00010

8. Dechartres A, Altman DG, Trinquart L, Boutron I, Ravaud P. Association between analytic strategy and estimates of treatment outcomes in meta-analyses. *JAMA*. 2014;312(6):623-630. doi:10.1001/jama.2014.8166

9. Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536. doi:10.7326/0003-4819-155-8-201110180-00009

10. Ling DI, Pai M, Schiller I, Dendukuri N. A Bayesian framework for estimating the incremental value of a diagnostic test in the absence of a gold standard. *BMC Med Res Methodol*. 2014;14:67. doi:10.1186/1471-2288-14-67

11. Graham PL, Moran JL. Robust meta-analytic conclusions mandate the provision of prediction intervals in meta-analysis summaries. *J Clin Epidemiol*. 2012;65(5):503-510. doi:10.1016/j.jclinepi.2011.09.012

12. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ*. 2011;342:d549. doi:10.1136/bmj.d549

13. Plummer M. rJAGS: Bayesian graphical models using MCMC. Published November 6, 2019. Accessed September 15, 2020. https://cran.r-project.org/web/packages/rjags/index.html

14. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2019.

15. Akgun Dogan O, Kose B, Agaoglu NB, et al. Does sampling saliva increase detection of SARS-CoV-2 by RT-PCR? comparing saliva with oro-nasopharyngeal swabs. Preprint. Posted online July 28, 2020. *medRxiv*. 2020.2007.2026.20158618.

16. Becker D, Sandoval E, Amin A, et al. Saliva is less sensitive than nasopharyngeal swabs for COVID-19 detection in the community setting. Preprint.

Posted online May 17, 2020. *medRxiv*. doi:10.1101/2020.05.11.20092338

17. Byrne RL, Kay GA, Kontogianni K, et al. Saliva offers a sensitive, specific and non-invasive alternative to upper respiratory swabs for SARS-CoV-2 diagnosis. Preprint. Posted online July 11, 2020. *medRxiv*. doi:10.1101/2020.07.09.20149534

18. Kojima N, Turner F, Slepnev V, et al. Self-collected oral fluid and nasal swab specimens demonstrate comparable sensitivity to clinician-collected nasopharyngeal swab specimens for the detection of SARS-CoV-2. Published online October 19, 2020. *Clin Infect Dis*. doi:10.1093/cid/ciaa1589

19. Teo AKJ, Choudhury Y, Tan IB, et al. Validation of saliva and self-administered nasal swabs for COVID-19 testing. Preprint. Posted online August 14, 2020. *medRxiv*. doi:10.1101/2020.08.13.20173807

20. Vogels CBF, Brackney D, Wang J, et al. SalivaDirect: simple and sensitive molecular diagnostic test for SARS-CoV-2 surveillance. Preprint. Posted online September 28, 2020. *medRxiv*. doi:10.1101/2020.08.03.20167791

21. Yokota I, Shane PY, Okada K, et al. Mass screening of asymptomatic persons for SARS-CoV-2 using saliva. *Clin Infect Dis*. Published online September 25, 2020. doi:10.1093/cid/ciaa1388

22. Miller M, Jansen M, Bisignano A, et al. Validation of a self-administrable, saliva-based RT-qPCR test detecting SARS-CoV-2. Preprint. Posted online June 9, 2020. *medRxiv*. doi:10.1101/2020.06.05.20122721

23. Caulley L, Corsten M, Eapen L, et al. Salivary detection of COVID-19. *Ann Intern Med*. Published online August 28, 2020. doi:10.7326/M20-4738

24. Hanson KE, Barker AP, Hillyard DR, et al. 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*. 2020;58(11):e01824-20. doi:10.1128/JCM.01824-20

25. McCormick-Baw C, Morgan K, Gaffney D, et al. Saliva as an alternate specimen source for detection of SARS-CoV-2 in symptomatic patients using Cepheid Xpert Xpress SARS-CoV-2. *J Clin Microbiol*. 2020;58(8):e01109-20. doi:10.1128/JCM.01109-20

26. Cheuk S, Wong Y, Tse H, et al. Posterior oropharyngeal saliva for the detection of SARS-CoV-2. *Clin Infect Dis*.Published online June 21, 2020.

27. Iwasaki S, Fujisawa S, Nakakubo S, et al. Comparison of SARS-CoV-2 detection in

nasopharyngeal swab and saliva. *J Infect*. 2020;81(2):e145-e147. doi:10.1016/j.jinf.2020.05.071

28. Landry ML, Criscuolo J, Peaper DR. Challenges in use of saliva for detection of SARS CoV-2 RNA in symptomatic outpatients. *J Clin Virol*. 2020;130:104567. doi:10.1016/j.jcv.2020.104567

29. Williams E, Bond K, Zhang B, Putland M, Williamson DA. Saliva as a noninvasive specimen for detection of SARS-CoV-2. *J Clin Microbiol*. 2020;58(8):e00776-20. doi:10.1128/JCM.00776-20

30. Goldfarb DM, Tilley P, Al-Rawahi GN, et al. Self-collected saline gargle samples as an alternative to healthcare worker collected nasopharyngeal swabs for COVID-19 diagnosis in outpatients. Preprint. Posted online September 14, 2020. *medRxiv*. doi:10.1101/2020.09.13.20188334

31. Chau NVV, Thanh Lam V, Thanh Dung N, et al; OUCRU COVID-19 research group. The natural history and transmission potential of asymptomatic SARS-CoV-2 infection. *Clin Infect Dis*. 2020;71(10):2679-2687. doi:10.1093/cid/ciaa711

32. Wyllie AL, Fournier J, Casanovas-Massana A, et al. Saliva or nasopharyngeal swab specimens for detection of SARS-CoV-2. *N Engl J Med*. 2020;383(13):1283-1286. doi:10.1056/NEJMc2016359

33. Rao M, Rashid FA, Sabri FSAH, et al. Comparing nasopharyngeal swab and early morning saliva for the identification of SARS-CoV-2. *Clin Infect Dis*. Published online August 6, 2020. doi:10.1093/cid/ciaa1156

34. To KK, Tsang OT, Yip CC, et al. Consistent detection of 2019 novel coronavirus in saliva. *Clin Infect Dis*. 2020;71(15):841-843. doi:10.1093/cid/ciaa149

35. Boulware DR, Pullen MF, Bangdiwala AS, et al. A randomized trial of hydroxychloroquine as postexposure prophylaxis for COVID-19. *N Engl J Med*. 2020;383(6):517-525. doi:10.1056/NEJMoa2016638

36. Campbell JR, Uppal A, Oxlade O, et al. Active testing of groups at increased risk of acquiring SARS-CoV-2 in Canada: costs and human resource needs. *CMAJ*. 2020;192(40):E1146-E1155. doi:10.1503/cmaj.201128

37. Van Lancker W, Parolin Z. COVID-19, school closures, and child poverty: a social crisis in the making. *Lancet Public Health*. 2020;5(5):e243-e244. doi:10.1016/S2468-2667(20)30084-0

38. Stage HB, Shingleton J, Ghosh S, Scarabel F, Pellis L, Finnie T. Shut and re-open: the role of schools in the spread of COVID-19 in Europe. Preprint. Posted online June 26, 2020. *medRxiv*. doi:10.1101/2020.06.24.20139634