Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company’s public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Performance of saliva compared with nasopharyngeal swab for diagnosis of COVID-19 by NAAT in cross-sectional studies: Systematic review and meta-analysis

Donald Brody Duncan a,b, Katharine Mackett c, Muhammad Usman Ali c, Deborah Yamamura a,b,d, Cynthia Balion a,b,c,*

a Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario L8S 4K1, Canada
b Microbiology Department, Hamilton Regional Laboratory Medicine Program, Hamilton Health Sciences, Hamilton, Ontario L8L 2X2, Canada
c Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario L8S 4K1, Canada
d Division of Infectious Diseases, Department of Medicine, McMaster University, Hamilton, Ontario L8V 1C3, Canada

ARTICLE INFO

Keywords:
COVID-19
Saliva
Nasopharyngeal swab
NAAT
Diagnostic accuracy
Systematic review

ABSTRACT

Nucleic acid amplification testing (NAAT) is the preferred method to diagnose coronavirus disease 2019 (COVID-19). Saliva has been suggested as an alternative to nasopharyngeal swabs (NPS), but previous systematic reviews were limited by the number and types of studies available. The objective of this systematic review and meta-analysis was to assess the diagnostic performance of saliva compared with NPS for COVID-19. We searched Ovid MEDLINE, Embase, Cochrane, and Scopus databases up to 24 April 2021 for studies that directly compared paired NPS and saliva specimens taken at the time of diagnosis. Meta-analysis was performed using an exact binomial rendition of the bivariate mixed-effects regression model. Risk of bias was assessed using the QUADAS-2 tool. Of 2683 records, we included 23 studies with 25 cohorts, comprising 11,582 paired specimens. A wide variety of NAAT assays and collection methods were used. Meta-analysis gave a pooled sensitivity of 87 % (95 % CI = 83–90 %) and specificity of 99 % (95 % CI = 98–99 %). Subgroup analyses showed the highest sensitivity when the suspected individual is tested in an outpatient setting and is symptomatic. Our results support the use of saliva NAAT as an alternative to NPS NAAT for the diagnosis of COVID-19.

1. Introduction

Testing for coronavirus disease 2019 (COVID-19) has become ubiquitous in inpatient, outpatient, and non-healthcare settings. There is no laboratory reference standard for COVID-19 diagnosis, but nucleic acid amplification testing (NAAT) for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) on respiratory specimens has been recommended as the preferred testing method since early in the pandemic [1]. Nasopharyngeal swabs (NPS) are one of the most common specimens for NAAT, but NPS collection can be problematic. NPS testing is uncomfortable, can rarely cause serious injury, and requires a steady supply of special disposable swabs and healthcare workers trained to use them [2–4]. Saliva has been suggested as an alternative to NPS because it does not require swabs, is non-invasive, and can be self-collected [5]. Guidelines from the Infectious Disease Society of America from early 2021 considered these benefits when making a conditional recommendation for the use of saliva (along with nasal swabs, mid-turbinate swabs, and NPS) for the diagnosis of COVID-19, although they commented that saliva is a complex sample matrix which can interfere with assay performance [6].

Multiple reviews have been published on the topic of saliva testing for SARS-CoV-2. However, a preponderance of studies published early in the pandemic have serious concerns for bias or applicability due to their study design. Many studies only enrolled patients who had already been diagnosed with COVID-19, without a control group for comparison [7–9]. These primary studies described the detection of viral RNA in saliva vs NPS at various time points in the natural history of COVID-19, but they did not directly answer questions about diagnostic accuracy of saliva NAAT. Another common study design was case-control, comparing saliva positivity in participants with prior positive and

* Corresponding author at: McMaster University Medical Centre, Department of Laboratory Medicine, 1200 Main St. W., HSC – 2N30, Hamilton, ON L8N 3Z5, Canada.
E-mail address: balion@hhsc.ca (C. Balion).

https://doi.org/10.1016/j.clinbiochem.2022.08.004
Received 22 April 2022; Received in revised form 14 July 2022; Accepted 5 August 2022
Available online 8 August 2022
0009-9120/© 2022 The Canadian Society of Clinical Chemists. Published by Elsevier Inc. All rights reserved.
negative NPS tests. These studies can give diagnostic performance characteristics that are different from cohort or cross-sectional studies [10]. They also often enrolled patients at different time points of COVID-19 illness and convalescence, so their findings are not directly applicable to patients presenting for diagnosis.

Given the evolving state of the literature, our goal was to create a methodologically rigorous systematic review and meta-analysis to describe the diagnostic accuracy of saliva testing in patients suspected of COVID-19. We focused on patients presenting for initial diagnosis or screening, because these are the contexts in which saliva testing is the most impactful. We also aimed to comprehensively describe testing characteristics that could affect test accuracy. Our primary study question was: “In patients being assessed for SARS-CoV-2 infection, is there a difference in diagnostic performance between saliva and NPS specimens tested by NAAT?”.  

---

**Fig. 1.** PRISMA 2020 flow diagram From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71.
2. Methods

2.1. Systematic review design

This systematic review was reported following the PRISMA guideline (Preferred Reporting Items for Systematic reviews and meta-Analyses) [11]. Our study was registered on PROSPERO (CRD42020209485).

2.2. Eligibility criteria

We included studies that assessed participants for diagnosis of SARS-CoV-2 infection using NAAT on paired saliva versus NPS samples. Combined nasopharyngeal swab and oropharyngeal swab (NPS + OPS) was also considered an acceptable comparator if both swabs were collected at the same time and tested as a single combined result. Our outcome of interest was the accuracy of saliva compared with NPS for NAAT, and we excluded studies that did not report sufficient data to calculate sensitivity and specificity.

We had no restrictions on setting, participant characteristics, or presence of symptoms. Because our research question was focused on the use of saliva specimens for diagnosis of infection rather than detection of SARS-CoV-2 virus at later time points, we excluded studies that enrolled participants who had already been diagnosed with SARS-CoV-2 infection on clinical or microbiological grounds. Studies that met our inclusion criteria but also assessed non-eligible specimen types, patient populations, or assays were included only if performance characteristics for saliva versus NPS on the subset of eligible patients could be determined separately.

Eligible study designs were cross-sectional, cohort, and randomized control trial, although for our meta-analysis we only extracted cross-sectional data at the time of initial assessment and ignored later time points. We excluded case-control studies, case reports, case series, and letters to the editor, as those studies generally described patients who had not been diagnosed with SARS-CoV-2 infection prior to salivary testing and did not reflect the purpose of our research question. We excluded pre-prints and other non-peer-reviewed literature. We excluded studies that prospectively collected saliva samples from all patients but used the results of NPS testing to pre-select a subset of saliva samples for testing.

2.3. Search strategy and selection process

We searched the Ovid MEDLINE, Embase, Cochrane, and Scopus databases up to 24 April 2021 using a predefined search strategy (Supplemental Fig. 1). Each record underwent an initial screen based on the title and abstract, and then an initial full-text review by two independent reviewers from the McMaster Evidence Review Synthesis Team. At each of these stages, if either reviewer chose to include the record it would proceed to the next stage. For the final inclusion stage, two content experts (DBD, KM) independently reviewed full-text articles for inclusion. Disagreements were resolved by discussion and consensus between four authors (DB, KM, DY, CB). In addition, we hand-searched the references of other systematic reviews on the topic for studies that met our inclusion criteria that were not found by our search strategy.

2.4. Data collection

Two authors (DBD, KM) independently extracted data from included studies into a standardized form. Disagreements were resolved by consensus. We emailed the corresponding authors for each study if outcomes of interest were not reported or unclear in the original manuscript.

To calculate our primary outcome, we collected data on the number of positive and negative results for paired saliva vs NPS specimens and the total number of specimens tested. For studies that included an “indeterminate” result category, we counted those results as negative. For studies with a cohort design and data gathered at multiple time points, only cross-sectional data on the first pair of specimens collected at the time of diagnosis were included.

Data on multiple other outcomes of interest were collected. Study characteristics were country, study design, inclusion and exclusion criteria, clinical setting, presence of symptoms, total number of patients enrolled, patient age, and patient sex or gender. Pre-analytical variables were technique of saliva collection, patient preparation, self-collection versus healthcare worker collection, time of collection, timespan between saliva and NPS collection, use of collection device and transport medium, volume of sample collected, and sample handling during transport and storage. Analytical variables collected were NAAT method, assay type and name, nucleic acid extraction performed, extra processing of saliva specimens, gene targets, use of adequacy control, cut-off value for positivity, reported limit of detection, number of targets needed to report a positive result, and definition of an indeterminate result.

2.5. Risk of bias assessment

Risk of bias of included studies was assessed independently by two authors (DBD, KM). Disagreements were resolved by consensus. We used the QUADAS-2 tool (Quality Assessment of Diagnostic Accuracy Studies), which evaluates the risk of bias of diagnostic accuracy studies across four domains (patient selection, index test, reference standard, and flow and timing) as well as concerns regarding applicability for the first three domains [12]. We made minor modifications to the original signaling questions to better reflect our study question (Supplemental Fig. 2).

2.6. Data analysis

For the meta-analysis of diagnostic test accuracy, we utilized the data from fourfold (two by two contingency) tables comparing test results for NAAT on saliva (as index) and NPS (as reference) tests as true positives (tp), false positives (fp), true negatives (tn) and false negatives (fn). Specifically, we used an exact binomial rendition of the bivariate mixed-effects regression models to generate the summary measures of effect as sensitivity, specificity, diagnostic odds ratio, and likelihood ratios along with their 95% confidence intervals (CI) [13-17].

A summary receiver operating curve (SROC) along with the area under the curve (AUC) and its 95% CIs was plotted based on parameters estimated by the bivariate model around summary sensitivity and specificity [18]. The AUC provides a global measure of test accuracy with values of 0.5 to 0.7, 0.7 to 0.9, and 0.9 to 1.0 considered as low, moderate, and high accuracy, respectively [19].

The I² statistic was employed to quantify the magnitude of statistical heterogeneity between studies where I² 30 % to 60 % represents moderate and I² 60 % to 90 % represents substantial heterogeneity across studies [20]. We carried out additional sub-group and meta-regression analysis based on pre-specified subgroups (methodological quality items, patient setting, symptoms, technique, preparation, transport medium, collection, reference specimen, extraction, adequacy control, extra processing, assay type, and number of positive gene targets) to quantify differential effect of screening and to facilitate exploratory analysis of observed heterogeneity across studies [14,21]. The results from meta-regression analysis are presented as likelihood ratio chi-squared (LRTChi²), I² heterogeneity statistic and corresponding p-values associated with the effect of categorized groups (covariates) on summary sensitivity and specificity. The degree of interdependence between performance measures (sensitivity and specificity) was tested using bivariate box plot. For subgroups with <4 studies, we used random effects logistic regression for meta-analysis of diagnostic accuracy data (an extension of generalized linear model for binomial family with a logit link) [22,23].

Publication bias was also assessed using both statistical tests and
visual inspection of funnel plot asymmetry [24]. All analyses were carried out using STATA ver. 16.0 software (MIDAS, METANDI and METADTA modules) [25–28].

3. Results

3.1. Study selection

Our search strategy yielded 2683 records (Fig. 1). Duplicates were removed and 2580 records were screened by title and abstract. 366 records underwent an initial full-text review, 148 were flagged for possible eligibility, and 22 studies were chosen for inclusion [29–50]. Of note, we excluded two studies that only did saliva testing on a pre-selected subset of NPS-negative specimens [51,52], and we emailed the authors of one excluded pre-print to confirm that it had not yet been submitted for peer-review [53].

We then hand-searched the reference lists of other systematic reviews on the topic [54–70] and did a full text review of 98 unique citations. We identified one additional study that met our inclusion criteria that was not identified in our initial search [71]. A total of 23 studies were therefore included in our systematic review [29–50,71]. Eight studies described in our review had not yet been included in any other systematic review [30,32,34,36,37,40,41,50].

3.2. Description of study characteristics

Overall, we included 11,582 paired specimens from 23 studies. Table 1 shows the number of studies and paired specimens that correspond to each pre-specified subgroup. The number of specimens for each study ranged from 71 to 2,107, with an average of 504 and a median of 354. Two studies reported outcome data on two different sub-populations, which we separated into unique patient cohorts for the meta-analysis [37,71].

Detailed study characteristics are presented in Supplemental Tables 1 and 2. These tables include additional information not in the publication but directly obtained from study authors [31,35,36,38,43,46,47]. Most studies took place in an outpatient setting, except three that were in a hospital setting and two with a mix of outpatients and inpatient. Nine studies enrolled symptomatic patients, three enrolled asymptomatic patients, and eleven enrolled both. Most studies enrolled a wide age range of young and older adults. Only three studies included pediatric patients younger than 16 years old.

Pre-analytic factors are presented in Supplemental Table 3. Saliva was collected by three main types of techniques: spitting/drooling of oral cavity saliva, deep-throat saliva of the oropharynx, and direct swab of saliva from the base of the lower jaw. Most studies collected saliva in a dry, sterile container, although four added a transport medium. A wide variety of swab types and transport media were used for NPS collection.

Analytic factors are described in Supplemental Table 4. All studies used RT-PCR to test for SARS-CoV-2 RNA, except one study that used both RT-PCR and reverse-transcriptase loop-mediated isothermal amplification (RT-LAMP) and one study that used transcription-mediated amplification (TMA). Many different NAAT protocols with different gene targets and cut-off values were used, including in-vitro diagnostic tests from fourteen different manufacturers and laboratory-developed tests based on four different primer sets. Within each study, the same assay was used to test paired saliva and NPS samples, except one study which tested some NPS samples using a different RT-PCR assay.

3.3. Risk of bias

The risk of bias for each study was assessed using the QUADAS-2 tool and is presented in Table 2. Overall risk of bias from patient selection was low, except 5 studies that had unclear patient enrollment procedures. Regarding the conduct of the index and reference tests, most...
Studies had unclear risk of bias because of a lack of information about blinding. Two studies had a high risk of bias in this domain because the results of NPS testing were already known before saliva testing was conducted. Sixteen studies had a low risk of bias from patient flow and timing, six had a high risk of bias, and one was of unclear risk. Almost all studies had no applicability concerns, except one study that used an unusual method of saliva collection with a swab of the lower jaw and two studies that did not describe their studies’ patient populations.

### Table 1

| Variable                                | Number of Studies, n = 23 | Number of Included Paired Specimens, n = 11,582 |
|-----------------------------------------|---------------------------|-----------------------------------------------|
| Overall                                 | 23                        | 11,582 (100 %)                                |
| Setting                                 |                           |                                               |
| Outpatient                              | 18                        | 10,272 (88.7 %)                               |
| Hospital                                | 3                         | 1165 (10.1 %)                                 |
| Mixed                                   | 2                         | 145 (1.2 %)                                   |
| Symptoms                                |                           |                                               |
| Asymptomatic                            | 3                         | 4593 (39.7 %)                                 |
| Symptomatic                             | 9                         | 2304 (19.9 %)                                 |
| Both                                    | 11                        | 4685 (40.4 %)                                 |
| Patient preparation                     |                           |                                               |
| Restrictions                            | 9                         | 4723 (40.8 %)                                 |
| None                                    | 4                         | 1137 (9.8 %)                                  |
| Not reported                            | 10                        | 5722 (49.4 %)                                 |
| Saliva collection technique             | 5                         | 3026 (26.1 %)                                 |
| Deep throat                             | 12                        | 2974 (25.7 %)                                 |
| Spitting                                | 1                         | 501 (4.3 %)                                   |
| Swab                                    | 5                         | 5081 (43.9 %)                                 |
| Not reported                            |                           |                                               |
| Saliva transport medium                 |                           |                                               |
| VTM/UTM                                 | 4                         | 2171 (18.7 %)                                 |
| Dry                                     | 17                        | 9241 (79.8 %)                                 |
| Not reported                            | 2                         | 170 (1.5 %)                                   |
| Collection by health care worker        | 19                        | 9214 (79.6 %)                                 |
| Self-collected                          | 4                         | 2368 (20.4 %)                                 |
| HCW collected                           |                           |                                               |
| Reference specimen                      |                           |                                               |
| NPS                                     | 20                        | 10,749 (92.8 %)                               |
| NPS + OPS                               | 3                         | 833 (7.2 %)                                   |
| NAAT method                             |                           |                                               |
| RT-PCR                                  | 21                        | 9304 (80.3 %)                                 |
| RT-PCR and RT-LAMP                      | 1                         | 354 (3.1 %)                                   |
| Extraction required                     |                           |                                               |
| Yes                                     | 21                        | 11,028 (95.2 %)                               |
| No                                      | 2                         | 554 (4.8 %)                                   |
| Extra processing of saliva              | 16                        | 7123 (61.5 %)                                 |
| Yes                                     | 1                         | 476 (4.1 %)                                   |
| None                                    | 6                         | 3983 (34.4 %)                                 |
| Not reported                            |                           |                                               |
| Adequacy control                        |                           |                                               |
| RNaseP                                  | 8                         | 3188 (27.5 %)                                 |
| None                                    | 15                        | 8394 (72.5 %)                                 |
| # positive gene targets                 |                           |                                               |
| One                                     | 10                        | 6219 (53.7 %)                                 |
| Multiple                                | 10                        | 4384 (37.9 %)                                 |
| Not reported                            | 3                         | 979 (8.4 %)                                   |
| IVD vs LDT                              |                           |                                               |
| IVD                                     | 15                        | 6805 (58.8 %)                                 |
| LDT                                     | 6                         | 746 (6.4 %)                                   |
| Unsure                                  | 1                         | 2107 (18.2 %)                                 |
| Both IVD, LDT                           | 1                         | 1924 (16.6 %)                                 |

IVD = in vitro diagnostic assay. LDT = lab-developed test. NAAT = nucleic acid amplification method. RT-PCR = reverse transcriptase polymerase chain reaction. RT-PCR = reverse transcriptase polymerase chain reaction. RT-LAMP = reverse transcriptase loop-mediated isothermal amplification. VTM = viral transport medium. UTM = universal transport medium. NPS = nasopharyngeal swab. OPS = oropharyngeal swab. HCW = healthcare worker.

### 3.4. Assessment of studies for inclusion in meta-Analysis

Our initial quantitative analyses of all included studies (n = 23 studies; n = 25 participant cohorts) revealed high heterogeneity, with three studies that exhibited very low sensitivity (<53 %) [34,35,45]. These three studies were investigated separately as outliers to understand the cause of this deviation from methodological homogeneity. Castelain et al. used an unusual collection technique of directly swabbing the lower jaw near the salivary glands [34], Dogan et al. was the only study utilizing RT-PCR methods without nucleic acid extraction [35], and Nacher et al. included a substantial proportion of patients more than 10 days from symptom onset [45]. Because of the methodological heterogeneity and low sensitivity of these three studies, they were excluded from pooling in the meta-analysis to ensure statistical stability, generalizability, and robustness of our results.

### 3.5. meta-Analysis and meta regression

For the main meta-analysis of interest (n = 20 studies; n = 22 participant cohorts), we obtained a pooled sensitivity of 87 % (95 % CI = 83–90 %) and specificity of 99 % (95 % CI = 98–99 %) with an overall I² of 94 %, p < 0.001 (Table 3). Individual study-estimated sensitivity, specificity, and I² are illustrated as a forest plot (Fig. 2). The pooled positive likelihood ratio (95 % CI) was 65.2 (37.9–112.2), negative likelihood ratio was 0.13 (0.10–0.18), and diagnostic odds ratio was 495 (254–966), implying that the test can accurately discriminate between patients with and without COVID-19.

The SROC (Supplemental Figure 3) showed best test performance for saliva testing at a summary sensitivity of 87 % (95 % CI = 83–90 %) and specificity of 99 % (95 % CI = 98–99 %). The area under the curve was 0.98 (95 % CI = 0.96–0.99), suggesting the test performance exhibits high accuracy.

To further explore observed statistical heterogeneity and ensure robustness of our findings, pre-specified subgroup and meta-regression analyses were also performed. The subgroup analyses revealed that the 95 % CIs for sensitivity and specificity overlapped for all categories except for study setting (Table 3). When considering each categorized subgroup within the extracted variables, outpatient setting, testing symptomatic participants, utilization of a transport medium for saliva samples, deep throat coughing collection technique, employing an adequacy control, and use of a laboratory developed assay showed higher values for sensitivity (range: 88 % to 93 %). Across these subgroups, negligible to low statistical heterogeneity across studies for summary sensitivity and specificity was observed for deep throat coughing collection technique and use of a laboratory developed assay. Further meta-regression based on pre-specified subgroups did not reveal any statistically significant effect of the levels of categorized groups on summary sensitivity and specificity (Table 4). However, given that the sample size directly corresponds to number of studies in the analysis with small number of studies across most pre-specified subgroup categories, this limited the power of meta-regression to detect statistically significant differences. Reference specimen, NAAT method, and extraction parameters were not included in the meta regression due to lack of adequate number of studies as a comparator group for the regression [20]. The bivariate box plot (Supplemental Figure 4) revealed high degree of interdependence between performance measures with most studies clustering with in the median distribution of the data points with only four outliers, suggesting a lower degree of heterogeneity.

### 3.6. Publication bias

The computed funnel plot showed no visual asymmetry along the slope of coefficient (Supplemental Figure 5) with a non-significant p-value (p = 0.79), suggesting a low likelihood of publication bias.
4. Discussion

Our systematic review and meta-analysis found saliva to be an excellent alternative specimen to NPS for diagnosis of COVID-19, with point estimates for sensitivity and specificity of 87 % and 99 %, respectively.

The strength of our study was selecting patients being assessed for initial diagnosis. This allows our findings to be generalizable to clinical and public health scenarios where saliva testing would be most useful, such as outpatient and mass-screening settings where the benefits of saliva testing (easier collection technique, better tolerability to the patient, and option of self-collection) may justify its use despite a lower sensitivity. This contrasts with reviews published earlier in the pandemic which made quantitative estimates of test accuracy based on the inclusion of patients at multiple time points of illness, including convalescence. For example, one widely cited study reported a significant difference in saliva vs NPS positivity for inpatients with confirmed COVID-19, but 33 % of those patients were more than two weeks into their course of illness and therefore had a high rate of negative NAAT on both saliva and NPS specimens [8]. Uncontrolled studies or those that enrolled patients at various time points after diagnosis cannot be used to directly answer questions about diagnostic test accuracy for COVID-19.

We excluded case-control studies. The majority of case-control studies examined saliva testing in patients who had already been diagnosed with COVID-19 by NPS NAAT at a prior time point, so NPS and saliva testing were not synchronous. Our study protocol also excluded studies that prospectively collected paired NPS and saliva specimens, and then retrospectively tested only a subset of saliva specimens based on a pre-determined ratio of NPS positivity/negativity [51,52]. The use of the same enrollment criteria for both cases and controls is less likely to introduce systemic bias [72].

Two other systematic reviews similarly excluded studies that had a case-control methodology or only enrolled known COVID-19 cases [56,70]. Their findings were similar to ours, although their confidence intervals were larger; Tsang et al. reported a sensitivity of 85 % (95 % CI = 75–93 %) and specificity of 99 % (95 % CI = 98–99 %) for saliva NAAT, while Butler-Laporte et al. reported a sensitivity of 83 % (95 % CI = 74–91 %) and specificity of 99 % (95 % CI = 98–99 %). However, they
Table 3
meta-Analysis with Subgroup Analysis.

| Variable                          | Number of Cohorts | I²   | p-value | Sensitivity (95% CI) | Specificity (95%CI) |
|-----------------------------------|-------------------|------|---------|----------------------|---------------------|
| Summary – all studies             | 22/18             | 0.00 | 0.87 [0.83, 0.90] | 0.99 [0.98, 0.99] |
| Setting                           |                   |      |         |                      |                     |
| Outpatient                        | 18/9              | 0.00 | 0.88 [0.83, 0.91] | 0.99 [0.98, 0.99] |
| Hospital                          | 2/0               | 0.91 | 0.98 [0.95, 0.99] |                     |
| Mixed                             | 2/0               | 0.85 [0.81, 0.88] | 0.97 [0.95, 0.98]  |                     |
| Symptom                           |                   |      |         |                      |                     |
| Asymptomatic                      | 4/0               | 0.00 | 0.82 [0.67, 0.89] | 0.99 [0.95, 0.99] |
| Symptomatic                       | 8/1               | 0.91 | 1.00    |                      |                     |
| Both                              | 10/5              | 0.84 [0.78, 0.89] | 0.99 [0.96, 0.99] |                     |
| Patient                           |                   |      |         |                      |                     |
| preparation                        | 8/3               | 0.00 | 0.88 [0.80, 0.93] | 0.99 [0.98, 0.99] |
| Restrictions                       | 0/0               | 0.83 [0.68, 0.91] | 0.99 [0.97, 1.00]  |                     |
| Not reported                       | 11/8              | 0.87 [0.80, 0.91] | 0.99 [0.96, 1.00]  |                     |
| Saliva collection technique       |                   |      |         |                      |                     |
| Deep throat                        | 4/0               | 0.84 [0.79, 0.90] | 0.99 [0.97, 0.98] |                     |
| Spitting                          | 15/8              | 0.91 [0.83, 0.95] | 1.00               |                     |
| Not reported                       | 3/0               | 0.91 [0.83, 0.95] | 1.00               |                     |
| Saliva transport medium           |                   |      |         |                      |                     |
| VTM/UTM                           | 2/18              | 0.92 [0.88, 0.95] | 0.99 [0.98, 1.00]  |                     |
| Dry                               | 2/18              | 0.86 [0.81, 0.90] | 0.99 [0.97, 0.99]  |                     |
| Not reported                       |                   |      |         |                      |                     |
| Collection by health care worker   |                   |      |         |                      |                     |
| Self-collected                    | 21/1              | 0.87 [0.83, 0.91] | 0.99 [0.98, 0.99]  |                     |
| HCW collected                     | 1/1               | 0.83 [0.79, 0.92] | 0.97 [0.97, 0.99]  |                     |
| Reference specimen                |                   |      |         |                      |                     |
| NPS                               | 19/3              | 0.87 [0.83, 0.91] | 0.99 [0.98, 0.99]  |                     |
| NPS + OPS                         |                   |      |         |                      |                     |
| NAAT method                       |                   |      |         |                      |                     |
| RT-PCR                            | 19/9              | 0.86 [0.81, 0.89] | 0.98 [0.98, 0.99]  |                     |
| TMA                               | 1/0               | 0.90 [0.90, 0.90] | 0.99 [0.98, 0.99]  |                     |
| RT-PCR and RT-LAMP                | 2/0               | 0.86 [0.86, 0.91] | 0.99 [0.95, 0.99]  |                     |
| Extraction required               |                   |      |         |                      |                     |
| Yes                               | 21/1              | 0.86 [0.82, 0.86] | 0.99 [0.95, 0.99]  |                     |
| None                              | 1/1               | 0.94 [0.86, 0.94] | 0.98 [0.95, 0.98]  |                     |
| Adequacy control                  |                   |      |         |                      |                     |
| RNaseP                            | 7/1               | 0.90 [0.76, 0.97] | 0.99 [0.97, 0.99]  |                     |
| None                              | 15/9              | 0.96 [0.96, 0.96] | 0.99 [0.98, 0.99]  |                     |

Table 3 (continued)

| Variable                          | Number of Cohorts | I²   | p-value | Sensitivity (95% CI) | Specificity (95%CI) |
|-----------------------------------|-------------------|------|---------|----------------------|---------------------|
| positive gene targets             | 11/9              | 0.00 | 0.86 [0.79, 0.90] | 0.99 [0.97, 0.99] |
| One                               | 9/0               | 0.91 | 1.00    |                      |                     |
| Multiple                          | 2/0               | 0.87 [0.80, 0.92] | 0.99 [0.97, 0.99]  |                     |
| Not reported                       |                   |      |         |                      |                     |
| IVD vs LDT                        |                   |      |         |                      |                     |
| IVD                               | 13/6              | 0.85 [0.80, 0.89] | 0.98 [0.97, 0.99]  |                     |
| LDT                               | 6/0               | 0.47 [0.33, 0.88] | 0.99 [0.97, 1.00]  |                     |
| Unsure                            | 1/0               | 0.93 [0.82, 0.94] | 0.98 [0.94, 0.99]  |                     |
| Both IVD, LDT                     | 2/0               | 0.78 [0.71, 0.84] | 0.99 [0.99, 1.00]  |                     |

CI = confidence interval. N/A = not available. IVD = in vitro diagnostic assay. LDT = lab-developed test. NAAT = nucleic acid amplification method. RT-PCR = reverse transcriptase polymerase chain reaction. TMA = transcription mediated amplification. RT-LAMP = reverse transcriptase loop-mediated isothermal amplification. VTM = viral transport medium. UTM = universal transport medium. NPS = nasopharyngeal swab. OPS = oropharyngeal swab. HCW = healthcare worker.

Heterogeneity statistics could not be computed for parameters with n < 2 studies.

Table 4
Meta Regression.

| Parameter                          | Category          | Number of cohorts | Chi² | I²   | p-value |
|-----------------------------------|-------------------|-------------------|------|------|---------|
| Setting                           |                   | 18/4              | 5.01 | 60   | 0.08    |
| Symptoms                          |                   | 15/4              | 3.10 | 36   | 0.21    |
| Patient                           |                   | 8/0               | 0.29 | 0.86 |         |
| Preparation                       |                   | 14/1              | 2.61 | 23   | 0.27    |
| Technique of saliva collection    |                   | 18/1              | 2.20 | 9    | 0.33    |
| Saliva transport medium           |                   | 4/0               | 0.85 | 0.66 |         |
| Collection by health care worker  |                   | 1/1               | 2.70 | 0.19 |         |
| Extra processing of saliva        |                   | 16/3              | 3.30 | 39   | 0.19    |
| Adequacy control                  |                   | 6/0               | 0.13 | 0.94 |         |
| Number of positive gene targets   |                   | 11/0              | 0.60 | 0.74 |         |
| IVD vs LDT                        |                   | 16/3              | 2.94 | 32   | 0.23    |

Note. HCW, healthcare worker; IVD, in vitro diagnostic assay; LDT, lab-developed test; UTM, universal transport medium; VTM, viral transport medium.

included some combined cohorts of undiagnosed and already diagnosed COVID-19 patients, which would systematically bias results towards the reference standard [73–80]. To our knowledge, there is only one meta-analysis that exclusively analyzed patients that had not yet been diagnosed with COVID-19, and it reported a sensitivity of 85 % (95 % CI = 77–91 %) and specificity of 99 % (95 % CI = 98–100 %) [62]. That study by Kivelä did not assess for risk of bias, had limited information of pre-
analytical and analytical testing characteristics, and examined literature up to 15 September 2020.

Our subgroup analyses showed the highest sensitivity when the suspected individual is tested within an outpatient setting and is symptomatic. However, the number of studies for many subgroups was small, so there is insufficient statistical power to conclusively identify differences in subgroups. There was also substantial heterogeneity between studies. This is likely due to the wide variety of pre-analytic and analytic factors, as described in Supplemental Tables 3 and 4. Although this reflects real-life practice because there is no standardization for saliva collection and testing, it is a limitation of the statistical power of our study.

An important variable that we could not quantitatively assess is the impact of extra processing of saliva. Saliva is a heterogeneous body fluid, and some studies used protocols for diluting or homogenizing samples that appeared viscous or contained mucus (Supplemental Table 4). Our meta-analysis was not able to determine if extra processing impacted diagnostic accuracy, but other researchers have shown that lack of homogenization can decrease NAAT sensitivity for lead to invalid test results [81]. The viscosity of saliva interferes with pipetting, which complicates the use of automated robotic processing equipment [82]. Saliva can also contain particulates, RNases, and other inhibitors that cause false-negative PCR results [83,84]. Many protocols require patients to refrain from eating, drinking, or smoking before saliva collection for this reason.

These factors play a role in the relative paucity of saliva-based assays compared with more common specimen types like NPS or nasal cavity swabs. The unique features of saliva must be considered before implementing saliva-based testing for SARS-CoV-2. Laboratories that are accustomed to swab-based testing may find unexpected difficulties when applying their usual operating procedures to saliva [82]. Efforts have been made to find simpler and less expensive methods of saliva testing, such as SalivaDirect [85]. Simplifying processing requirements will make saliva testing more accessible. Nonetheless, saliva testing is not appropriate for all patients, such as those that cannot generate a sufficient volume of liquid saliva for testing. The higher rate of invalid test results from inadequately collected or processed saliva may be unacceptable in some clinical contexts, such as sick inpatients [6]. These potential drawbacks to saliva testing must be carefully balanced against the benefits described earlier. Further head-to-head studies comparing these pre-analytic and processing factors will be helpful to guide laboratories in choosing the most accurate and reliable testing method.

In a broader context, our meta-analysis provides the strongest evidence to date that saliva is an appropriate specimen for respiratory virus testing. However, the data are not as robust for respiratory viral infections other than COVID-19, and there is some evidence that the accuracy of saliva testing may differ between viruses. A network meta-analysis comparing different sampling methods found that saliva was the best specimen for rhinovirus, parainfluenza virus, and adenovirus but was not as good for influenza virus or respiratory syncytial virus [92]. The possibility that saliva could be a superior specimen type for certain viruses is exciting, given the ease and low-cost of saliva collection. Nonetheless, for now the role that saliva testing will play in the pandemic remains unclear. We recommend manufacturers and researchers perform studies to validate saliva as an acceptable specimen type for viruses other than SARS-CoV-2. Bringing more saliva-based commercial assays to the market will be crucial for laboratories that do not have the capabilities to validate their own in-house test.

There is no single laboratory reference standard for COVID-19 diagnosis [86], so we chose NAAT on NPS or NPS + OPS specimens as the reference comparator. PCR on upper respiratory tract specimens is the first-line test in many settings, but it is an imperfect test. Accuracy of both microbiologic and non-microbiologic tests varies significantly over the course of this dynamic illness and in different populations [87]. To counterbalance this limitation, some other meta-analyses used a composite reference standard or latent class modelling to assess saliva testing for COVID-19, such as counting a positive result from any specimen type as a true positive and disregarding the possibility of false positives. These statistical approaches will give different estimates of accuracy and can have different sources of bias [88–90]. Another important limitation is that testing of upper respiratory specimens can miss infections that would be detected in other specimens, like bronchoalveolar lavage fluid [91]. Readers must be aware of the lack of a gold standard assay or specimen type when assessing the literature comparing different methods of diagnosing COVID-19.

The constantly changing dynamics of the global COVID-19 pandemic introduces some limitations of applicability of our results. All of the included studies in this systematic review were conducted before the emergence of the delta and omicron variants. The emergence of viral variants has affected some aspects of SARS-CoV-2 testing, including mutations in PCR target sites [93]. It is also possible that viral variants may have altered viral shedding from different anatomical sites which would affect the relative accuracy of different specimen types at different time points [94].

In conclusion, our systematic review and meta-analysis supports the use of saliva NAAT for diagnosis of COVID-19, allowing clinicians, laboratorians, and public health professionals to decide if saliva testing is appropriate for their specific clinical contexts.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Donna Fitzpatrick-Lewis, Dr. Diana Sherifali, and the McMaster Evidence Review Synthesis Team for their assistance with developing a search strategy, screening of records, and initial data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.clinbiochem.2022.08.004.

References

[1] interim guidance [Internet]. (2020). Available from: https://www.who.int/publi

[2] O. Vandenberg, D. Martiny, O. Rochas, A. van Belkum, Z. Kozlakidis, Considerations for diagnostic COVID-19 tests, Nat. Rev. Microbiol. 19 (3) (2021) 171–185.

[3] A. Koskinnen, M. Tolvi, M. Jauhiainen, E. Kekäläinen, A. Laulajainen-Hongisto, S. Lamminmäki, Complications of COVID-19 Nasopharyngeal Swab Test, JAMA Otolaryngol Head Neck Surg. 147 (7) (2021) 672.

[4] X. Moisnet, N. Gautier, T. Godet, S. Parabère, B. Pereira, E. Meunier, L. Gerbault, O. Lesens, C. Herquel, J. Beytout, P. Clavelou, Nasopharyngeal swab-induced pain for SARS-CoV-2 screening: A randomised controlled trial of conventional and self-swabbing, Eur. J. Pain 25 (4) (2021) 934–940.

[5] Azzi L. Saliva is the Key Element for Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Mass Screening. Clinical Infectious Diseases. 2021; 73 (3)656–658.

[6] K.E. Hanson, A.M. Caliendo, C.A. Arias, M.K. Hayden, J.A. Englund, M.J. Lee, M. Loeb, R. Patel, A. El Alyayi, O. Alsayar, A. Patel, Y. Falck-Ytter, V. Lavergne, R. L. Morgan, M.H. Murad, S. Sultan, A. Bhimraj, R.A. Mustafa, The infectious diseases society of america guidelines on the diagnosis of COVID-19: molecular diagnostic testing. Clin. Infect. Dis. (2021).

[7] To KK-W, Tsang OT-Y, Yip CC-Y, et al. Consistent Detection of 2019 Novel Coronavirus in Clinical Specimens. 2020; 129(20):566–571.

[8] Jamal AJ, Mozafarshahin M, Coomes E, et al. Sensitivity of Nasopharyngeal Swabs and Saliva for the Detection of Severe Acute Respiratory Syndrome Coronavirus 2. Clinical Infectious Diseases. 2021; 72(6):1064–1066.

[9] L. Azzi, G. Carcano, F. Gianfagna, P. Grotti, D.D. Gasperina, A. Genoni, M. Fasano, F. Sesko, L. Tettamanti, F. Carinci, V. Maurino, A. Rossi, A. Tagliabue, A. Baj, Saliva is a reliable tool to detect SARS-CoV-2, J. Infect. 81 (1) (2021) e45–e50.
Clinical Biochemistry 117 (2023) 84–93
93
Kojima N, Turner F, Slepnev V, et al. Self-Collected Oral Fluid and Nasal Swabs
Iwasaki S, Fujisawa S, Nakakubo S, et al. Comparison of SARS-CoV-2 detection in
A.W. Rutjes, J.B. Reitsma, J.P. Vandenbroucke, A.S. Glas, P.M. Bossuyt, Case-
Sagredo-Olivares K, Morales-G
O. Fakheran, M. Dehghannejad, A. Khademi, Saliva as a diagnostic specimen for
detection of SARS-CoV-2 in suspected patients: a scoping review, Infect Dis
K. Nasiri, A. Dimitrova, Comparing saliva and nasopharyngeal swab specimens in
C. Mishra, S. Meena, J.K. Meena, S. Tiwari, P. Mathur, Detection of three pandemic
J.M. Kivel
G. Butler-Laporte, A. Lawandi, I. Schiller, M. Yao, N. Dendukuri, E.G. McDonald, T.
D.B. Duncan et al.
K. Molberg, D. Cavuoti, A.J. McAdam, Saliva as an alternate specimen source for

93

[38x272]
[74]

[72]

[69]

[66]

[64]

[62]

[59]

[56]

[54x232]

K. Tao, P.L. Tzou, J. Nouhin, et al., The biological and clinical significance of


D. Hamilton, J. Barat, S. Das, V. De Giorgi, D.K. Henderson, S. Kopka, A.F. Lau, T. Miller,
M. Hegde, A.M. Rojiani, R. Kolhe, SalivaSTAT: Direct-PCR and pooling of saliva
samples collected in healthcare and community setting for SARS-CoV-2 mass
surveillance. Diagnostics. 11 (5) (2021) 904.

[82]

[81]

[80]

[79]

[78]

[77]

[76]

[75]

[74]

[73]

[72]

[71]

[70]

[69]

[68]

[67]

[66]

[65]

[64]

[63]

[62]

[61]

[60]

[59]

[58]

[57]

[56]

[55]

[54]

[53]

[52]

[51]

[50]

[49]

[48]

[47]

[46]

[45]

[44]

[43]

[42]

[41]

[40]

[39]

[38]

[37]

[36]

[35]

[34]

[33]

[32]

[31]

[30]

[29]

[28]

[27]

[26]

[25]

[24]

[23]

[22]

[21]

[20]

[19]

[18]

[17]

[16]

[15]

[14]

[13]

[12]

[11]

[10]

[9]

[8]

[7]

[6]

[5]

[4]

[3]

[2]

[1]

[0]