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.

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,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*

A B S T R A C T

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

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?".
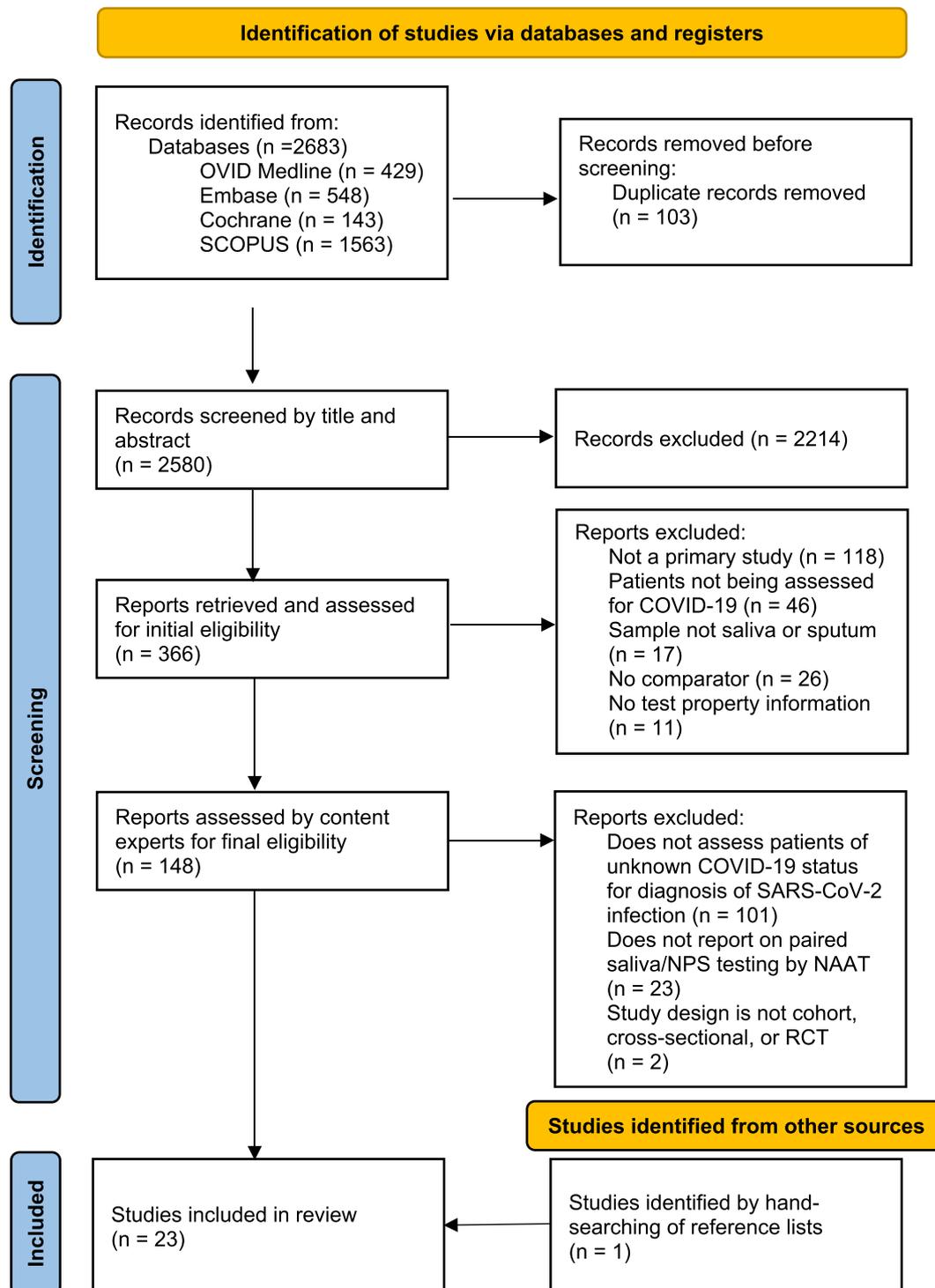

Fig. 1. PRISMA 2020 flow diagram *From:* Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71.

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 already 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 I^2 statistic was employed to quantify the magnitude of statistical heterogeneity between studies where I^2 30 % to 60 % represents moderate and I^2 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 (LRTChi^2), I^2 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].

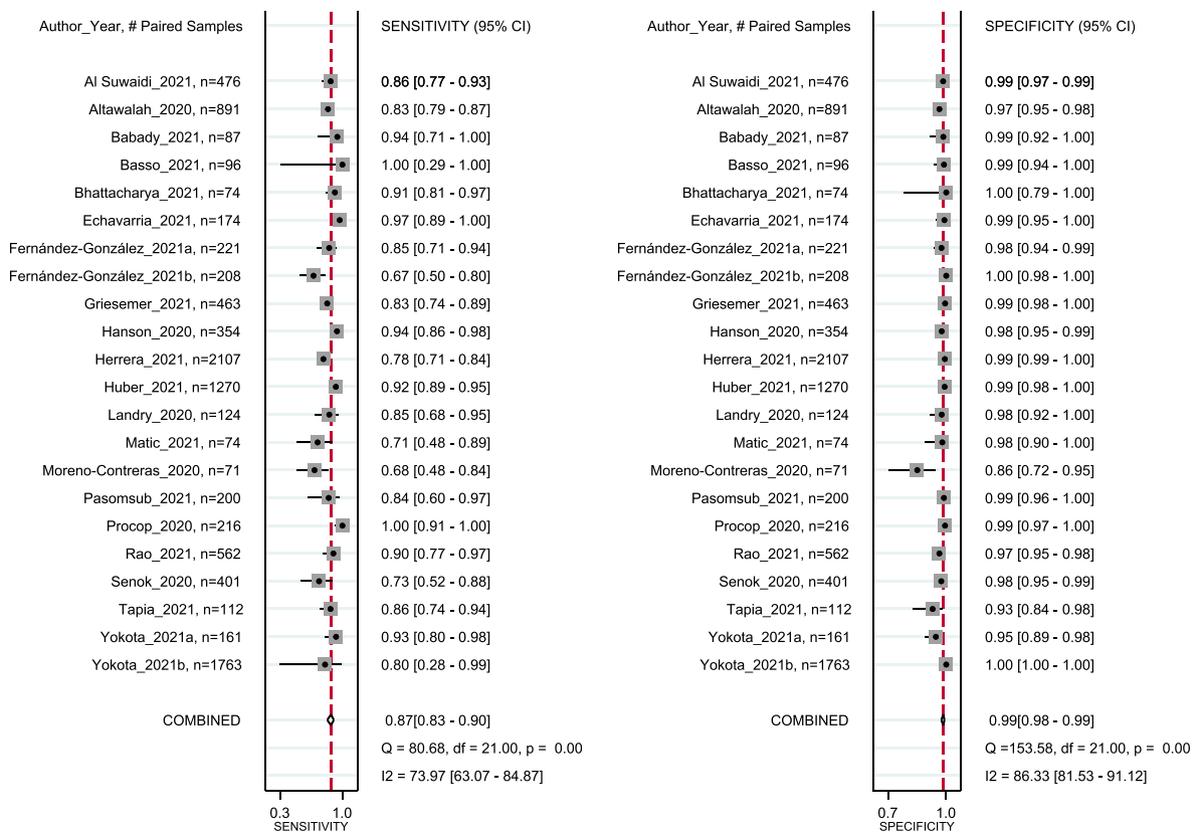Publication bias was also assessed using both statistical tests and

Author_Year, # Paired Samples		SENSITIVITY (95% CI)	Author_Year, # Paired Samples		SPECIFICITY (95% CI)
Al Suwaidi_2021, n=476		0.86 [0.77 - 0.93]	Al Suwaidi_2021, n=476		0.99 [0.97 - 0.99]
Altawalah_2020, n=891		0.83 [0.79 - 0.87]	Altawalah_2020, n=891		0.97 [0.95 - 0.98]
Babady_2021, n=87		0.94 [0.71 - 1.00]	Babady_2021, n=87		0.99 [0.92 - 1.00]
Basso_2021, n=96		1.00 [0.29 - 1.00]	Basso_2021, n=96		0.99 [0.94 - 1.00]
Bhattacharya_2021, n=74		0.91 [0.81 - 0.97]	Bhattacharya_2021, n=74		1.00 [0.79 - 1.00]
Echavarria_2021, n=174		0.97 [0.89 - 1.00]	Echavarria_2021, n=174		0.99 [0.95 - 1.00]
Fernández-González_2021a, n=221		0.85 [0.71 - 0.94]	Fernández-González_2021a, n=221		0.98 [0.94 - 0.99]
Fernández-González_2021b, n=208		0.67 [0.50 - 0.80]	Fernández-González_2021b, n=208		1.00 [0.98 - 1.00]
Griesemer_2021, n=463		0.83 [0.74 - 0.89]	Griesemer_2021, n=463		0.99 [0.98 - 1.00]
Hanson_2020, n=354		0.94 [0.86 - 0.98]	Hanson_2020, n=354		0.98 [0.95 - 0.99]
Herrera_2021, n=2107		0.78 [0.71 - 0.84]	Herrera_2021, n=2107		0.99 [0.99 - 1.00]
Huber_2021, n=1270		0.92 [0.89 - 0.95]	Huber_2021, n=1270		0.99 [0.98 - 1.00]
Landry_2020, n=124		0.85 [0.68 - 0.95]	Landry_2020, n=124		0.98 [0.92 - 1.00]
Matic_2021, n=74		0.71 [0.48 - 0.89]	Matic_2021, n=74		0.98 [0.90 - 1.00]
Moreno-Contreras_2020, n=71		0.68 [0.48 - 0.84]	Moreno-Contreras_2020, n=71		0.86 [0.72 - 0.95]
Pasomsub_2021, n=200		0.84 [0.60 - 0.97]	Pasomsub_2021, n=200		0.99 [0.96 - 1.00]
Procop_2020, n=216		1.00 [0.91 - 1.00]	Procop_2020, n=216		0.99 [0.97 - 1.00]
Rao_2021, n=562		0.90 [0.77 - 0.97]	Rao_2021, n=562		0.97 [0.95 - 0.98]
Senok_2020, n=401		0.73 [0.52 - 0.88]	Senok_2020, n=401		0.98 [0.95 - 0.99]
Tapia_2021, n=112		0.86 [0.74 - 0.94]	Tapia_2021, n=112		0.93 [0.84 - 0.98]
Yokota_2021a, n=161		0.93 [0.80 - 0.98]	Yokota_2021a, n=161		0.95 [0.89 - 0.98]
Yokota_2021b, n=1763		0.80 [0.28 - 0.99]	Yokota_2021b, n=1763		1.00 [1.00 - 1.00]
COMBINED		0.87[0.83 - 0.90]	COMBINED		0.99[0.98 - 0.99]
		Q = 80.68, df = 21.00, p = 0.00			Q =153.58, df = 21.00, p = 0.00
		I2 = 73.97 [63.07 - 84.87]			I2 = 86.33 [81.53 - 91.12]

0.3   1.0
SENSITIVITY

0.7   1.0
SPECIFICITY

Fig. 2. Meta-Analysis Forest Plot.

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

Table 1
Summary of Study Variables and Included Paired Specimens by Subgroup.

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 %)
TMA	1	354 (3.1 %)
RT-PCR and RT-LAMP	1	1924 (16.6 %)
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. 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.

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.

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 I^2 of 94 %, $p < 0.001$ (Table 3). Individual study-estimated sensitivity, specificity, and I^2 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.

Table 2
Risk of Bias.

Study	RISK OF BIAS				APPLICABILITY CONCERNS		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Al Suwaidi, 2021	Low	Low	Low	Low	Low	Low	Low
Altawalah, 2020	Low	Unclear	Unclear	Low	Low	Low	Low
Babady, 2021	Low	Unclear	Unclear	High	Low	Low	Low
Basso, 2021	Low	Low	Low	Low	Low	Low	Low
Bhattacharya, 2021	Unclear	Unclear	Unclear	Low	Low	Low	Low
Castelain, 2021	Unclear	Unclear	Unclear	Unclear	Low	High	Low
Dogan, 2021	Low	Unclear	Unclear	Low	Low	Low	Low
Echavarria, 2021	Low	Unclear	Low	Low	Low	Low	Low
Fernandez-Gonzalez, 2021	Low	High	Low	Low	Low	Low	Low
Griesmer, 2021	Low	Unclear	Unclear	Low	Low	Low	Low
Hanson, 2020	Low	Unclear	Unclear	High	Low	Low	Low
Herrera, 2021	Low	Unclear	Unclear	Low	Low	Low	Low
Huber, 2021	Low	Low	Low	Low	Low	Low	Low
Landry, 2020	Low	High	Low	High	Low	Low	Low
Matic, 2020	Unclear	Unclear	Unclear	High	Low	Low	Low
Moreno-Contreras, 2020	Unclear	Unclear	Unclear	Low	Unclear	Low	Low
Nacher, 2021	Low	Unclear	Unclear	Low	Low	Low	Low
Pasomsub, 2021	Low	Low	Low	Low	Low	Low	Low
Procop, 2020	Low	Low	Low	High	Low	Low	Low
Rao, 2021	Low	Unclear	Unclear	Low	Low	Low	Low
Senok, 2020	Low	Unclear	Unclear	Low	Low	Low	Low
Tapia, 2021	Unclear	Unclear	Unclear	Low	Unclear	Low	Low
Yokota, 2021	Low	Unclear	Unclear	High	Low	Low	Low

😊 Low Risk   ☹ High Risk   ? Unclear Risk.

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

D.B. Duncan et al.   *Clinical Biochemistry xxx (xxxx) xxx*

Table 3
meta-Analysis with Subgroup Analysis.

Variable	Number of Cohorts	I²	p-value	Sensitivity (95 % CI)	Specificity (95 %CI)
Summary – all studies	22	94	0.000	0.87 [0.83, 0.90]	0.99 [0.98, 0.99]
Setting					
Outpatient	18	94	0.000	0.88 [0.83, 0.91]	0.99 [0.98, 0.99]
Hospital	2	N/A	N/A	0.85 [0.81, 0.88]	0.97 [0.95, 0.98]
Mixed	2	N/A	N/A	0.69 [0.55, 0.81]	0.93 [0.85, 0.96]
Symptoms					
Asymptomatic	4	94	0.000	0.82 [0.67, 0.91]	0.99 [0.95, 1.00]
Symptomatic	8	0	0.491	0.91 [0.83, 0.95]	0.98 [0.96, 0.99]
Both	10	5	0.175	0.84 [0.78, 0.89]	0.99 [0.98, 0.99]
Patient preparation					
Restrictions	8	84	0.001	0.88 [0.80, 0.93]	0.99 [0.98, 0.99]
None	3	0.0	0.998	0.83 [0.68, 0.91]	0.99 [0.97, 1.00]
Not reported	11	88	0.000	0.87 [0.80, 0.91]	0.99 [0.96, 1.00]
Saliva collection technique					
Deep throat	4	0	0.357	0.91 [0.83, 0.95]	0.99 [0.96, 0.99]
Spitting	15	88	0.000	0.84 [0.79, 0.88]	0.98 [0.97, 0.99]
Not reported	3	0	0.000	0.91 [0.83, 0.95]	1.00 [0.36, 1.00]
Saliva transport medium					
VTM/UTM	2	N/A	N/A	0.92 [0.88, 0.95]	0.99 [0.98, 1.00]
Dry	18	95	0.000	0.86 [0.81, 0.90]	0.99 [0.97, 0.99]
Not reported	2	N/A	N/A	0.92 [0.82, 0.97]	0.99 [0.94, 1.00]
Collection by health care worker					
Self-collected	21	94	0.000	0.87 [0.83, 0.91]	0.99 [0.98, 0.99]
HCW collected	1	N/A	N/A	0.83 [0.79, 0.87]	0.97 [0.97, 0.99]
Reference specimen					
NPS	19	93	0.000	0.87 [0.83, 0.91]	0.99 [0.98, 0.99]
NPS + OPS	3	0.01	0.071	0.83 [0.68, 0.92]	0.96 [0.88, 0.99]
NAAT method					
RT-PCR	19	91	0.000	0.86 [0.81, 0.90]	0.98 [0.98, 0.99]
TMA	1	N/A	N/A	0.94 [0.86, 0.98]	0.98 [0.95, 0.99]
RT-PCR and RT-LAMP	2	N/A	N/A	0.91 [0.79, 0.97]	1.00 [0.99, 1.00]
Extraction required					
Yes	21	94	0.000	0.86 [0.82, 0.90]	0.86 [0.82, 0.90]
None	1	N/A	N/A	0.94 [0.86, 0.98]	0.98 [0.95, 0.99]
Extra processing of saliva					
Yes	16	77	0.007	0.89 [0.84, 0.92]	0.98 [0.97, 0.99]
None	1	N/A	N/A	0.86 [0.77, 0.93]	0.99 [0.97, 0.99]
Not reported	5	51	0.064	0.82 [0.72, 0.89]	0.99 [0.98, 1.00]
Adequacy control					
RNaseP	7	77	0.007	0.90 [0.76, 0.96]	0.99 [0.97, 1.00]
None	15	91	0.000	0.86 [0.82, 0.90]	0.99 [0.97, 0.99]
# positive gene targets					
One	11	90	0.000	0.86 [0.79, 0.91]	0.99 [0.97, 1.00]
Multiple	9	86	0.000	0.87 [0.80, 0.92]	0.99 [0.97, 0.99]
Not reported	2	N/A	N/A	0.91 [0.84, 0.95]	0.98 [0.96, 0.99]
IVD vs LDT					
IVD	13	67	0.024	0.85 [0.80, 0.89]	0.98 [0.97, 0.99]
LDT	6	0	0.475	0.93 [0.82, 0.97]	0.98 [0.94, 0.99]
Unsure	1	N/A	N/A	0.78 [0.71, 0.84]	0.99 [0.99, 1.00]
Both IVD, LDT	2	N/A	N/A	0.91 [0.79, 0.97]	1.00 [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	Outpatient	18	5.01	60	0.08
	Hospital, mixed	4			
Symptoms	Symptomatic	8	3.10	36	0.21
	Asymptomatic, both	14			
Patient preparation	Restrictions	8	0.29	0	0.86
	No restrictions, not reported	14			
Technique of saliva collection	Deep throat	4	2.61	23	0.27
	Spitting, not reported	18			
Saliva transport medium	Dry	18	2.20	9	0.33
	VTM/UTM, not reported	4			
Collection by health care worker	Self-collected	21	0.85	0	0.66
	HCW collected	1			
Extra processing of saliva	Yes	16	3.30	39	0.19
	None, not reported	6			
Adequacy control	RNaseP	7	0.13	0	0.94
	None	15			
Number of positive gene targets	One target	11	0.60	0	0.74
	Multiple targets, not reported	11			
IVD vs LDT	LDT	6	2.94	32	0.23
	IVD, both	16			

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 quantitively assess is the impact of extra processing of saliva. Saliva is a heterogenous 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 and 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 routine diagnostics of viral respiratory tract infections after 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 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

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/publications/i/item/10665-331501.

[2] O. Vandenberg, D. Martiny, O. Rochas, A. van Belkum, Z. Kozlakidis, Considerations for diagnostic COVID-19 tests, Nat. Rev. Microbiol. 19 (3) (2021) 171–183.

[3] A. Koskinen, 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. Moisset, N. Gautier, T. Godet, S. Parabère, B. Pereira, E. Meunier, L. Gerbaud, O. Lesens, C. Henquell, 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) 924–929.

[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):e566–e568.

[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 Alayli, O. Altayar, P. 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 Saliva. Clinical Infectious Diseases. **2020**; 71(15):841–843.

[8] Jamal AJ, Mozafarihashjin 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. Grossi, D.D. Gasperina, A. Genoni, M. Fasano, F. Sessa, 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) (2020) e45–e50.

[10] J.G. Lijmer, Empirical evidence of design-related bias in studies of diagnostic tests, JAMA 282 (11) (1999) 1061.

[11] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021; n71.

[12] P.F. Whiting, QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies, Ann. Intern. Med. 155 (8) (2011) 529.

[13] H. Chu, S.R. Cole, Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach, J. Clin. Epidemiol. 59 (12) (2006) 1331–1332.

[14] H.C. van Houwelingen, L.R. Arends, T. Stijnen, Advanced methods in meta-analysis: multivariate approach and meta-regression, Statist Med. 21 (4) (2002) 589–624.

[15] H.C. Van Houwelingen, K.H. Zwinderman, T. Stijnen, A bivariate approach to meta-analysis, Statist Med. 12 (24) (1993) 2273–2284.

[16] J.B. Reitsma, A.S. Glas, A.W.S. Rutjes, R.J.P.M. Scholten, P.M. Bossuyt, A. H. Zwinderman, Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews, J. Clin. Epidemiol. 58 (10) (2005) 982–990.

[17] Riley R, Abrams K, Sutton A, Lambert P, Thompson J. The benefits and limitations of multivariate meta-analysis, with application to diagnostic and prognostic studies. Technical Report 05-04. 2005.

[18] L.R. Arends, T.H. Hamza, J.C. van Houwelingen, M.H. Heijenbrok-Kal, M.G. M. Hunink, T. Stijnen, Bivariate random effects meta-analysis of ROC curves, Med. Decis. Making 28 (5) (2008) 621–638.

[19] J.A. Swets, Measuring the accuracy of diagnostic systems, Science 240 (4857) (1988) 1285–1293.

[20] Deeks J, Higgins J, Altman D, editors. Chapter 10: Analysing data and undertaking meta-analyses. Cochrane Handbook for Systematic Reviews of Interventions version 62 (updated February 2021) [Internet]. Cochrane; 2021. Available from: www.training.cochrane.org/handbook.

[21] Dinnes J, Deeks J, Kirby J, Roderick P. A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. Health Technol Assess [Internet]. 2005 [cited 2021 Dec 9]; 9(12). Available from: https://www.journalslibrary.nihr.ac.uk/hta/hta9120/.

[22] T.H. Hamza, H.C. van Houwelingen, T. Stijnen, The binomial distribution of meta-analysis was preferred to model within-study variability, J. Clin. Epidemiol. 61 (1) (2008) 41–51.

[23] Brown LD, Cai TT, DasGupta A. Interval Estimation for a Binomial Proportion. Statist Sci [Internet]. 2001 [cited 2021 Dec 9]; 16(2). Available from: https://projecteuclid.org/journals/statistical-science/volume-16/issue-2/Interval-Estimation-for-a-Binomial-Proportion/10.1214/ss/1009213286.full.

[24] J.J. Deeks, P. Macaskill, L. Irwig, The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed, J. Clin. Epidemiol. 58 (9) (2005) 882–893.

[25] StataCorp., Stata statistical software, StataCorp LLC, College Station, TX, 2019.

[26] B. Dwamena, midas,, A program for meta-analytical integration of diagnostic accuracy studies in Stata. Ann Arbor, Michigan: division of nuclear medicine. Department of Radiology, University of Michigan Medical School, 2007.

[27] R. Harbord, metandi: Stata module for meta-analysis of diagnostic accuracy, Boston College Department of Economics, Statistical Software Components, 2008.

[28] V.N. Nyaga, A.M. Metadta, A stata command for meta-analysis and meta-regression of diagnostic test accuracy data – A Tutorial. [Internet], Jul. Available from: In Review (2021) https://www.researchsquare.com/article/rs-721114/v1.

[29] Altawalah H, AlHuraish F, Alkandari WA, Ezzikouri S. Saliva specimens for detection of severe acute respiratory syndrome coronavirus 2 in Kuwait: A cross-sectional study. Journal of Clinical Virology. 2020; 132:104652.

[30] H. Al Suwaidi, A. Senok, R. Varghese, Z. Deesi, I. Khansaheb, S. Pokasirakath, B. Chacko, I. Abufara, T. Loney, A. Alsheikh-Ali, Saliva for molecular detection of SARS-CoV-2 in school-age children, Clin. Microbiol. Infect. 27 (9) (2021) 1330–1335.

[31] N.E. Babady, T. McMillen, K. Jani, A. Viale, E.V. Robilotti, A. Aslam, M. Diver, D. Sokoli, G. Mason, M.K. Shah, D. Korenstein, M. Kamboj, Performance of severe acute respiratory syndrome coronavirus 2 real-time RT-PCR tests on oral rinses and saliva samples, J. Mol. Diagn. 23 (1) (2021) 3–9.

[32] D. Basso, A. Aita, A. Padoan, C. Cosma, F. Navaglia, S. Moz, N. Contran, C.-F. Zambon, A. Maria Cattelan, M. Plebani, Salivary SARS-CoV-2 antigen rapid detection: A prospective cohort study, Clin. Chim. Acta 517 (2021) 54–59.

[33] D. Bhattacharya, D. Parai, U.K. Rout, P. Dash, R.R. Nanda, G.C. Dash, S. Kanungo, S.K. Palo, S. Giri, H.R. Choudhary, J.S. Kshatri, J. Turuk, B.K. Mishra, R.K. Lenka, S. Dash, S. Pati, Saliva for diagnosis of SARS-CoV-2: First report from India, J. Med. Virol. 93 (4) (2021) 2529–2533.

[34] Castelain S, François C, Demey B, et al. Comparison of nasopharyngeal and saliva swabs for the detection of RNA SARS-CoV-2 during mass screening (SALICOV study). 3.

[35] Dogan OA, Kose B, Agaoglu NB, et al. Does sampling saliva increase detection of SARS-CoV-2 by RT-PCR? Comparing saliva with oro-nasopharyngeal swabs. Journal of Virological Methods. 2021; 290:114049.

[36] M. Echavarria, N.S. Reyes, P.E. Rodriguez, M. Ypas, C. Ricarte, M.P. Rodriguez, M. G. Perez, A. Seoane, A. Martinez, C. Videla, M.E. Stryjewski, G. Carballal, Self-collected saliva for SARS-CoV-2 detection: A prospective study in the emergency room, J. Med. Virol. 93 (5) (2021) 3268–3272.

[37] M. Fernández-González, V. Agulló, A. de la Rica, A. Infante, M. Carvajal, J. A. García, N. Gonzalo-Jiménez, C. Cuartero, M. Ruiz-García, C. de Gregorio, M. Sánchez, M. Masiá, F. Gutiérrez, M.B. Miller, Performance of saliva specimens for the molecular detection of SARS-CoV-2 in the community setting: does sample collection method matter? J. Clin. Microbiol. 59 (4) (2021).

[38] S.B. Griesemer, G. Van Slyke, D. Ehrbar, K. Strle, T. Yildirim, D.A. Centurioni, A. C. Walsh, A.K. Chang, M.J. Waxman, K. St. George, A.J. McAdam, Evaluation of specimen types and saliva stabilization solutions for SARS-CoV-2 testing, J. Clin. Microbiol. 59 (5) (2021).

[39] K.E. Hanson, A.P. Barker, D.R. Hillyard, N. Gilmore, J.W. Barrett, R.R. Orlandi, S. M. Shakir, A.J. McAdam, Self-collected anterior nasal and saliva specimens versus health care worker-collected nasopharyngeal swabs for the molecular detection of SARS-CoV-2, J. Clin. Microbiol. 58 (11) (2020).

[40] L.A. Herrera, A. Hidalgo-Miranda, N. Reynoso-Noverón, A.A. Meneses-García, A. Mendoza-Vargas, J.P. Reyes-Grajeda, F. Vadillo-Ortega, A. Cedro-Tanda, F. Peñaloza, E. Frías-Jimenez, C. Arriaga-Canon, R. Ruiz, O. Angulo, I. López-Villaseñor, C. Amador-Bedolla, D. Vilar-Compte, P. Cornejo, M. Cisneros-Villanueva, E. Hurtado-Cordova, M. Cendejas-Orozco, J.S. Hernández-Morales, B. Moreno, I.A. Hernández-Cruz, C.A. Herrera, M.A. González-Woge, P. Munguía-Garza, F. Luna-Maldonado, A. Sánchez-Vizcarra, V.G. Osnaya, N. Medina-Molotla, Y. Alfaro-Mora, R.E. Cáceres-Gutiérrez, L. Tolentino-García, P. Rosas-Escobar, S.A. Román-González, M.A. Escobar-Arrazola, J.C. Canseco-Méndez, D.R. Ortiz-Soriano, J. Domínguez-Ortiz, A.D. González-Barrera, D. I. Aparicio-Bautista, A. Cruz-Rangel, A.P. Alarcón-Zendejas, L. Contreras-Espinosa, R. González, L. Guerra-Calderas, M.A. Meraz-Rodríguez, M. Montalvo-Casimiro, R. Montiel-Manríquez, K. Torres-Arciga, D. Venegas, V. Juárez-González, X. Guajardo-Barreto, V. Monroy-Martínez, D. Guillén, J. Fernández, J. Herrera, R. León-Rodriguez, I. Canela-Pérez, B.H. Ruíz-Ordaz, R. Valdez-Vazquez, J. Bertin-Montoya, M. Niembro-Ortega, L. Villegas-Acosta, D. López-Castillo, A. Soriano-Ríos, M. Gastelum-Ramos, T. Zamora-Barandas, J. Morales-Baez, M. García-Rodríguez, M. García-Martínez, E. Nieto-Patlán, M. Quirasco-Baruch, I. López-Martínez, E. Ramírez-Gonzalez, H. Olivera-Díaz, N. Escobar-Escamilla, Saliva is a reliable and accessible source for the detection of SARS-CoV-2, International Journal of Infectious Diseases. 105 (2021) 83–90.

[41] M. Huber, P.W. Schreiber, T. Scheier, A. Audigé, R. Buonomano, A. Rudiger, D. L. Braun, G. Eich, D.I. Keller, B. Hasse, J. Böni, C. Berger, H.F. Günthard, A. Manrique, A. Trkola, High efficacy of saliva in detecting SARS-CoV-2 by RT-PCR in adults and children, Microorganisms. 9 (3) (2021) 642.

[42] Landry ML, Criscuolo J, Peaper DR. Challenges in use of saliva for detection of SARS CoV-2 RNA in symptomatic outpatients. Journal of Clinical Virology. 2020; 130:104567.

[43] N. Matic, A. Stefanovic, V. Leung, T. Lawson, G. Ritchie, L. Li, S. Champagne, M. G. Romney, C.F. Lowe, Practical challenges to the clinical implementation of saliva for SARS-CoV-2 detection, Eur. J. Clin. Microbiol. Infect. Dis. 40 (2) (2021) 447–450.

[44] J. Moreno-Contreras, M.A. Espinoza, C. Sandoval-Jaime, M.A. Cantú-Cuevas, H. Barón-Olivares, O.D. Ortiz-Orozco, A.V. Muñoz-Rangel, M. Hernández-de la Cruz, C.M. Eroza-Osorio, C.F. Arias, S. López, A.M. Caliendo, Saliva sampling and its direct lysis, an excellent option to increase the number of SARS-CoV-2 diagnostic tests in settings with supply shortages, J. Clin. Microbiol. 58 (10) (2020).

[45] Nacher M, Mergeay-Fabre M, Blanchet D, et al. Prospective Comparison of Saliva and Nasopharyngeal Swab Sampling for Mass Screening for COVID-19. Front Med. 2021; 8:621160.

[46] E. Pasomsub, S.P. Watcharananan, K. Boonyawat, P. Janchompoo, G. Wongtabtim, W. Suksuwan, S. Sungkanuparph, A. Phuphuakrat, Saliva sample as a non-invasive specimen for the diagnosis of coronavirus disease 2019: a cross-sectional study, Clin. Microbiol. Infect. 27 (2) (2021) 285.e1–285.e4.

[47] G.W. Procop, N.K. Shrestha, S. Vogel, K. Van Sickle, S. Harrington, D.D. Rhoads, B. P. Rubin, P. Terpeluk, R. Hayden, A direct comparison of enhanced saliva to nasopharyngeal swab for the detection of SARS-CoV-2 in symptomatic patients, J. Clin. Microbiol. 58 (11) (2020).

[48] M. Rao, F.A. Rashid, F.S.A.H. Sabri, N.N. Jamil, V. Seradja, N.A. Abdullah, H. Ahmad, S.L. Aren, S.A.S. Ali, M. Ghazali, A.A. Manaf, H. Talib, R. Hashim, R. Zain, R. Thayan, F. Amran, T. Aris, N. Ahmad, COVID-19 screening test by using random oropharyngeal saliva, J. Med. Virol. 93 (4) (2021) 2461–2466.

[49] A. Senok, H. Alsuwaidi, Y. Atrah, et al., Saliva as an alternative specimen for molecular COVID-19 testing in community settings and population-based screening, IDR. 13 (2020) 3393–3399.

[50] C.V. Tapia, C. Marcia, M. Ivone, P. Nadia, G. Camila, A. Valentina, I. Paula, M. Fabien, Performance of saliva samples for COVID-19 diagnosis by using the AllplexTM 2019-nCoV assay kit, Front Med. 8 (2021), 617399.

[51] C. Kandel, J. Zheng, J. McCready, M. Serbanescu, H. Racher, M. Desaulnier, J. Powis, K. Vojdani, L. Finlay, E. Sheldrake, C. Vermeiren, K. Katz, A. McGeer, R. Kozak, L. Goneau, Detection of SARS-CoV-2 from saliva as compared to nasopharyngeal swabs in outpatients, Viruses. 12 (11) (2020) 1314.

[52] C. Trobajo-Sanmartín, M. Adelantado, A. Navascués, M.J. Guembe, I. Rodrigo-Rincón, J. Castilla, C. Ezpeleta, Self-collection of saliva specimens as a suitable alternative to nasopharyngeal swabs for the diagnosis of SARS-CoV-2 by RT-qPCR, JCM. 10 (2) (2021) 299.

[53] Becker D, Sandoval E, Amin A, et al. Saliva is less sensitive than nasopharyngeal swabs for COVID-19 detection in the community setting [Internet]. Infectious Diseases (except HIV/AIDS); 2020 May. Available from: http://medrxiv.org/lookup/doi/10.1101/2020.05.11.20092338.

[54] M.A. Atieh, M. Guirguis, N.H.M. Alsabeeha, R.D. Cannon, The diagnostic accuracy of saliva testing for SARS-CoV-2: A systematic review and meta-analysis, Oral Dis. (2021) :odi.13934.

[55] M.L. Bastos, S. Perlman-Arrow, D. Menzies, J.R. Campbell, The sensitivity and costs of testing for SARS-CoV-2 infection with saliva versus nasopharyngeal swabs: A systematic review and meta-analysis, Ann. Intern. Med. 174 (4) (2021) 501–510.

[56] G. Butler-Laporte, A. Lawandi, I. Schiller, M. Yao, N. Dendukuri, E.G. McDonald, T. C. Lee, Comparison of saliva and nasopharyngeal swab nucleic acid amplification testing for detection of SARS-CoV-2: A systematic review and meta-analysis, JAMA Intern Med. 181 (3) (2021) 353.

[57] M.G. Cañete, I.M. Valenzuela, P.C. Garcés, I.C. Massó, M.J. González, S. G. Providell, Saliva sample for the massive screening of SARS-CoV-2 infection: a systematic review, Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology. 131 (5) (2021) 540–548.

[58] L.M. Czumbel, S. Kiss, N. Farkas, I. Mandel, A. Hegyi, Á. Nagy, Z. Lohinai, Z. Szakács, P. Hegyi, M.C. Steward, G. Varga, Saliva as a candidate for COVID-19 diagnostic testing: A meta-analysis, Front Med. 7 (2020).

[59] 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 Poverty. 9 (1) (2020) 100.

[60] Ibrahimi N, Delaunay-Moisan A, Hill C, et al. Screening for SARS-CoV-2 by RT-PCR: Saliva or nasopharyngeal swab? Rapid review and meta-analysis. Darlix J-LE, editor. PLoS ONE. **2021**; 16(6):e0253007.

[61] K. Khiabani, M.H. Amirzade-Iranaq, Are saliva and deep throat sputum as reliable as common respiratory specimens for SARS-CoV-2 detection? A systematic review and meta-analysis, Am. J. Infect. Control 49 (9) (2021) 1165–1176.

[62] J.M. Kivelä, H. Jarva, M. Lappalainen, S. Kurkela, Saliva-based testing for diagnosis of SARS-CoV-2 infection: A meta-analysis, J. Med. Virol. 93 (3) (2021) 1256–1258.

[63] R.A. Lee, J.C. Herigon, A. Benedetti, N.R. Pollock, C.M. Denkinger, R. M. Humphries, Performance of saliva, oropharyngeal swabs, and nasal swabs for SARS-CoV-2 molecular detection: a systematic review and meta-analysis, J. Clin. Microbiol. 59 (5) (2021).

[64] C. Mishra, S. Meena, J.K. Meena, S. Tiwari, P. Mathur, Detection of three pandemic causing coronaviruses from non-respiratory samples: systematic review and meta-analysis, Sci. Rep. 11 (1) (2021) 16131.

[65] V.M. Moreira, P. Mascarenhas, V. Machado, J. Botelho, J.J. Mendes, N. Taveira, M. G. Almeida, Diagnosis of SARS-Cov-2 infection by RT-PCR using specimens other than naso- and oropharyngeal swabs: A systematic review and meta-analysis, Diagnostics. 11 (2) (2021) 363.

[66] K. Nasiri, A. Dimitrova, Comparing saliva and nasopharyngeal swab specimens in the detection of COVID-19: A systematic review and meta-analysis, Journal of Dental Sciences. 16 (3) (2021) 799–805.

[67] T.J. O'Leary, Relative sensitivity of saliva and upper airway swabs for initial detection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in ambulatory patients, J. Mol. Diagn. 23 (3) (2021) 265–273.

[68] M. Riccò, S. Ranzieri, S. Peruzzi, et al., RT-qPCR assays based on saliva rather than on nasopharyngeal swabs are possible but should be interpreted with caution: results from a systematic review and meta-analysis.: Diagnosis of SARS-CoV-2 through salivary fluids, Acta Bio Medica Atenei Parmensis. 91 (3) (2020) e2020025.

[69] Sagredo-Olivares K, Morales-Gómez C, Aitken-Saavedra J. Evaluation of saliva as a complementary technique to the diagnosis of COVID-19: a systematic review. Med Oral. **2021**; e526–e532.

[70] N.N.Y. Tsang, H.C. So, K.Y. Ng, B.J. Cowling, G.M. Leung, D.K.M. Ip, Diagnostic performance of different sampling approaches for SARS-CoV-2 RT-PCR testing: a systematic review and meta-analysis, Lancet Infect. Dis. 21 (9) (2021) 1233–1245.

[71] Yokota I, Shane PY, Okada K, et al. Mass Screening of Asymptomatic Persons for Severe Acute Respiratory Syndrome Coronavirus 2 Using Saliva. Clinical Infectious Diseases. **2021**; 73(3):e559–e565.

[72] A.W. Rutjes, J.B. Reitsma, J.P. Vandenbroucke, A.S. Glas, P.M. Bossuyt, Case-control and two-gate designs in diagnostic accuracy studies, Clin. Chem. 51 (8) (2005) 1335–1341.

[73] Iwasaki S, Fujisawa S, Nakakubo S, et al. Comparison of SARS-CoV-2 detection in nasopharyngeal swab and saliva. Journal of Infection. 2020; 81(2):e145–e147.

[74] Kojima N, Turner F, Slepnev V, et al. Self-Collected Oral Fluid and Nasal Swabs Demonstrate Comparable Sensitivity to Clinician Collected Nasopharyngeal Swabs for Coronavirus Disease 2019 Detection. Clinical Infectious Diseases. **2020**; ciaa1589.

[75] C. McCormick-Baw, K. Morgan, D. Gaffney, Y. Cazares, K. Jaworski, A. Byrd, K. Molberg, D. Cavuoti, A.J. McAdam, 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. 58 (8) (2020).

[76] A.K.J. Teo, Y. Choudhury, I.B. Tan, C.Y. Cher, S.H. Chew, Z.Y. Wan, L.T.E. Cheng, L.L.E. Oon, M.H. Tan, K.S. Chan, L.Y. Hsu, Saliva is more sensitive than nasopharyngeal or nasal swabs for diagnosis of asymptomatic and mild COVID-19 infection, Sci. Rep. 11 (1) (2021).

[77] Wong SCY, Tse H, Siu HK, et al. Posterior Oropharyngeal Saliva for the Detection of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). Clinical Infectious Diseases. **2020**; 71(11):2939–2946.

[78] Miller M, Jansen M, Bisignano A, et al. Validation of a Self-administrable, Saliva-based RT-qPCR Test Detecting SARS-CoV-2 [Internet]. Infectious Diseases (except HIV/AIDS); 2020 Jun. Available from: http://medrxiv.org/lookup/doi/10.1101/2020.06.05.20122721.

[79] B. Barat, S. Das, V. De Giorgi, D.K. Henderson, S. Kopka, A.F. Lau, T. Miller, T. Moriarty, T.N. Palmore, S. Sawney, C. Spalding, P. Tanjutco, G. Wortmann, A. M. Zelazny, K.M. Frank, A.J. McAdam, Pooled saliva specimens for SARS-CoV-2 testing, J. Clin. Microbiol. 59 (3) (2021).

[80] Migueres M, Mengelle C, Dimeglio C, et al. Saliva sampling for diagnosing SARS-CoV-2 infections in symptomatic patients and asymptomatic carriers. Journal of Clinical Virology. **2020**; 130:104580.

[81] N.S. Sahajpal, A.K. Mondal, S. Ananth, A. Njau, P. Ahluwalia, G. Newnam, A. Lozoya-Colinas, N.V. Hud, V. Kota, T.M. Ross, M.D. Reid, S. Fulzele, A. Chaubey, 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] Hamilton JR, Stahl EC, Tsuchida CA, et al. Robotic RNA extraction for SARS-CoV-2 surveillance using saliva samples. Kalendar R, editor. PLoS ONE. **2021**; 16(8): e0255690.

[83] D. White, J. Gu, C.-J. Steinberg, D. Yamamura, B.J. Salena, C. Balion, C.D.M. Filipe, A. Capretta, Y. Li, J.D. Brennan, Investigation of discordant SARS-CoV-2 RT-PCR results using minimally processed saliva, Sci. Rep. 12 (1) (2022).

[84] Nishibata Y, Koshimoto S, Ogaki K, et al. RNase in the saliva can affect the detection of severe acute respiratory syndrome coronavirus 2 by real-time one-step polymerase chain reaction using saliva samples. Pathology - Research and Practice. **2021**; 220:153381.

[85] Vogels CBF, Watkins AE, Harden CA, et al. SalivaDirect: A simplified and flexible platform to enhance SARS-CoV-2 testing capacity. Med. **2021**; 2(3):263-280.e6.

[86] D.A. Korevaar, J. Toubiana, M. Chalumeau, M.D.F. McInnes, J.F. Cohen, Evaluating tests for diagnosing COVID-19 in the absence of a reliable reference standard: pitfalls and potential solutions, J. Clin. Epidemiol. 138 (2021) 182–188.

[87] M.T. Hernández-Huerta, L. Pérez-Campos Mayoral, L.M. Sánchez Navarro, G. Mayoral-Andrade, E. Pérez-Campos Mayoral, E. Zenteno, E. Pérez-Campos, Should RT-PCR be considered a gold standard in the diagnosis of COVID-19? J. Med. Virol. 93 (1) (2021) 137–138.

[88] I. Schiller, M. van Smeden, A. Hadgu, M. Libman, J.B. Reitsma, N. Dendukuri, Bias due to composite reference standards in diagnostic accuracy studies: Bias due to Composite Reference Standards, Statist Med. 35 (9) (2016) 1454–1470.

[89] M.R. Schofield, M.J. Maze, J.A. Crump, M.P. Rubach, R. Galloway, K.J. Sharples, On the robustness of latent class models for diagnostic testing with no gold standard, Stat. Med. 40 (22) (2021) 4751–4763.

[90] N. Dendukuri, A. Hadgu, L. Wang, Modeling conditional dependence between diagnostic tests: A multiple latent variable model, Statist Med. 28 (3) (2009) 441–461.

[91] G.M. Bwire, M.V. Majigo, B.J. Njiro, A. Mawazo, Detection profile of SARS-CoV-2 using RT-PCR in different types of clinical specimens: A systematic review and meta-analysis, J. Med. Virol. 93 (2) (2021) 719–725.

[92] Hou N, Wang K, Zhang H, et al. Comparison of detection rate of 16 sampling methods for respiratory viruses: a Bayesian network meta-analysis of clinical data and systematic review. BMJ Glob Health. **2020**; 5(11):e003053.

[93] K. Tao, P.L. Tzou, J. Nouhin, et al., The biological and clinical significance of emerging SARS-CoV-2 variants [cited 2021 Nov 4];. Available from: Nat Rev Genet [Internet]. (2021) https://www.nature.com/articles/s41576-021-00408-x.

[94] M.J. Siedner, J. Boucau, R.F. Gilbert, R. Uddin, J. Luu, S. Haneuse, T. Vyas, Z. Reynolds, S. Iyer, G.C. Chamberlin, R.H. Goldstein, C.M. North, C.A. Sacks, J. Regan, J.P. Flynn, M.C. Choudhary, J.M. Vyas, A.K. Barczak, J.E. Lemieux, J. Z. Li, Duration of viral shedding and culture positivity with post-vaccination SARS-CoV-2 delta variant infections, JCI Insight [Internet]. 7 (2) (2022).