Diagnostic performance of different sampling approaches for SARS-CoV-2 RT-PCR testing: a systematic review and meta-analysis

Nicole Ngai Yung Tsang, Hau Chi So, Ka Yan Ng, Benjamin J Cowling, Gabriel M Leung, Dennis Kai Ming Ip

Summary

Background The comparative performance of different clinical sampling methods for diagnosis of SARS-CoV-2 infection by RT-PCR among populations with suspected infection remains unclear. This meta-analysis aims to systematically compare the diagnostic performance of different clinical specimen collection methods.

Methods In this systematic review and meta-analysis, we systematically searched PubMed, Embase, MEDLINE, Web of Science, *medRxiv*, *bioRxiv*, SSRN, and Research Square from Jan 1, 2000, to Nov 16, 2020. We included original clinical studies that examined the performance of nasopharyngeal swabs and any additional respiratory specimens for the diagnosis of SARS-CoV-2 infection among individuals presenting in ambulatory care. Studies without data on paired samples, or those that only examined different samples from confirmed SARS-CoV-2 cases were not useful for examining diagnostic performance of a test and were excluded. Diagnostic performance, including sensitivity, specificity, positive predictive value, and negative predictive value, was examined using random effects models and double arcsine transformation.

Findings Of the 5577 studies identified in our search, 23 studies including 7973 participants with 16762 respiratory samples were included. Respiratory specimens examined in these studies included 7973 nasopharyngeal swabs, 1622 nasal swabs, 6110 saliva samples, 338 throat swabs, and 719 pooled nasal and throat swabs. Using nasopharyngeal swabs as the gold standard, pooled nasal and throat swabs gave the highest sensitivity of 97% (95% CI 93–100), whereas lower sensitivities were achieved by saliva (85%, 75–93) and nasal swabs (86%, 77–93) and a much lower sensitivity by throat swabs (68%, 35–94). A comparably high positive predictive value was obtained by pooled nasal and throat (97%, 90–100) and nasal swabs (96%, 87–100) and a slightly lower positive predictive value by saliva (93%, 88–97). Throat swabs have the lowest positive predictive value of 75% (95% CI 45–96). Comparably high specificities (range 97–99%) and negative predictive value (range 95–99%) were observed among different clinical specimens. Comparison between health-care-worker collection and self-collection for pooled nasal and throat swabs and nasal swabs showed comparable diagnostic performance. No significant heterogeneity was observed in the analysis of pooled nasal and throat swabs and throat swabs, whereas moderate to substantial heterogeneity ($I^2 \geq 30\%$) was observed in studies on saliva and nasal swabs.

Interpretation Our review suggests that, compared with the gold standard of nasopharyngeal swabs, pooled nasal and throat swabs offered the best diagnostic performance of the alternative sampling approaches for diagnosis of SARS-CoV-2 infection in ambulatory care. Saliva and nasal swabs gave comparable and very good diagnostic performance and are clinically acceptable alternative specimen collection methods. Throat swabs gave a much lower sensitivity and positive predictive value and should not be recommended. Self-collection for pooled nasal and throat swabs and nasal swabs was not associated with any significant impairment of diagnostic accuracy. Our results also provide a useful reference framework for the proper interpretation of SARS-CoV-2 testing results using different clinical specimens.

Introduction

SARS-CoV-2 infection emerged in late 2019 and has spread globally, with the number of newly confirmed cases growing to more than 122 million.[1] COVID-19 has a broad clinical spectrum, ranging from asymptomatic, mild clinical illness to one with severe complications and death.[2] Common symptoms include fever, cough, and fatigue, which overlap with those of other acute respiratory infections. Accurate and efficient diagnosis of SARS-CoV-2 infection is therefore necessary, especially in ambulatory care settings, to enable downstream clinical and infection control procedures, including case management and isolation, field investigation, contact tracing and quarantine, and community disease surveillance, to prevent further disease transmission in the community.

Research in context

Evidence before this study

Accurate and efficient diagnosis of SARS-CoV-2 infection is required for downstream clinical and public health procedures. Although several studies have examined the use of different respiratory specimens for detection of SARS-CoV-2 RNA by RT-PCR, the comparative performance of different clinical sampling methods among population with suspected infection remains unclear. A search of PubMed on Feb 20, 2021, using search terms describing SARS-CoV-2 infection and different respiratory specimens and with no language restrictions identified ten reviews on the issue. Five of these reviews exclusively included studies of patients with confirmed COVID-19 and only reported the agreement of positivity rate of different respiratory specimens in infected people. Because of the absence of non-infected participants in these studies, they are not useful in assessing diagnostic performance in identifying or excluding the disease among individuals with suspected infection in terms of false positivity or negative predictive value. The other five reviews only examined the performance of saliva in comparison with nasopharyngeal swabs, rather than giving a comprehensive comparison of all commonly used sampling methods. No review has examined the performance of pooled nasal and throat swabs, nasal swabs, or throat swabs or reported positive and negative predictive values of different sampling approaches, two indicators necessary for understanding the implication of testing results from the perspective of both patients and health-care workers. Additionally, none of the reviews examined the comparative performance of samples collected by health-care workers or by self-collection.

Added value of this study

To fill this research gap, we did a systematic review and meta-analysis of studies comparing different clinical sampling methods of respiratory specimens for the detection of SARS-CoV-2 RNA. We systematically searched PubMed, Embase, MEDLINE, Web of Science, *medRxiv*, *bioRxiv*, SSRN, and Research Square from Jan 1, 2000, to Nov 16, 2020. We included original clinical studies that examined the performance of nasopharyngeal swabs and any additional respiratory specimens for the diagnosis of SARS-CoV-2 infection among individuals presenting in ambulatory care. We examined diagnostic performance, including sensitivity, specificity, positive predictive value, and negative predictive value using random effects models and double arcsine transformation.

To our knowledge, this is the first systematic review and meta-analysis examining the comparative diagnostic performance of different clinical sampling methods for SARS-CoV-2 testing in an ambulatory care setting and assesses sensitivity, specificity, positive predictive value, and negative predictive value. Our review suggested that, compared with the gold standard of nasopharyngeal swabs, pooled nasal and throat swabs offer the best diagnostic performance and represent the optimal alternative sampling approach. Saliva and nasal swabs also gave very good and comparable performance and are clinically acceptable alternative specimen collection methods for the accurate diagnosis of SARS-CoV-2 infection. For all the three methods, self-collection of these clinical specimens did not associate with any significant impairment of diagnostic accuracy.

Implications of all the available evidence

Our results provide a useful reference framework for the proper interpretation of SARS-CoV-2 testing results using different clinical specimens. They also support the use of pooled nasal and throat swabs, saliva, and nasal swabs as alternative sampling methods for SARS-CoV-2 RNA detection and the use of self-collection to facilitate efficient scaling up of testing in appropriate community settings.

RT-PCR is regarded as the gold-standard laboratory technique for the identification of SARS-CoV-2 in a clinical setting. In people with confirmed SARS-CoV-2 infection, lower respiratory tract specimens were generally reported to have higher positivity rates than other biosamples,[3] a finding consistent with the current understanding of the pathogenetic mechanism of SARS-CoV-2 infection over the disease course.[4] For clinical diagnosis of the infection, a variety of respiratory specimens are used for laboratory testing, with nasopharyngeal swabs so far regarded as the gold-standard sampling method for the diagnosis of SARS-CoV-2 infection.[5,6] However, several drawbacks have hindered the widespread use of nasopharyngeal swabs in ambulatory care settings, including the technical difficulty of specimen collection,[7] discomfort associated with the procedure, manpower implications, and the requirement for trained and technically experienced health-care workers, high-level personal protective equipment,[8] and

standardised negative-pressure settings, which are often not readily available, resulting in increased occupational risk exposure to health-care workers.

With the aim of improving the scalability of SARS-CoV-2 testing in ambulatory care settings under heightened demand,[9] several alternative sampling approaches have been explored, including pooled nasal and throat swabs, saliva,[10,11] nasal swabs,[12–17] and oropharyngeal swabs (throat swabs).[18–20] These alternative approaches to SARS-CoV-2 testing have the theoretical advantages of reduced invasiveness and simpler procedures than nasopharyngeal swabs, potentially making testing more acceptable and accessible.[10,11,16] The less stringent manpower and expertise requirement also allows for self-collection to be explored,[20,21] which might also help to reduce the risk to health-care workers.[10,11] However, a comprehensive understanding of the comparative diagnostic performance of these alternative sampling approaches is needed.

Although several published studies have investigated the performance of various alternative sampling approaches for SARS-CoV-2 testing, they generally had methodological limitations, including the inclusion of only confirmed positive cases,[22–25] inadequate sample size,[26] absence of differentiation between populations with suspected and confirmed infection,[26–28] absence of comparison with a suitable gold standard,[23] reporting of viral loads alone,[29] or comparison of aggregated result rather than use of a head-to-head comparison.[29] Ten reviews have tried to summarise existing evidence, but similarly all had multiple design limitations, and neither addressed clinically important measures such as positive or negative predictive values nor compared samples collected by health-care workers or by self-collection.[3,23,27,30–36]

A systematic review of the diagnostic accuracy of different sampling approaches for SARS-CoV-2 testing in individuals with suspected infection is therefore needed to properly evaluate their diagnostic performance. We aimed to systematically examine the comparative diagnostic performance of different clinical specimen collection methods for SARS-CoV-2 in populations with suspected infection presenting to ambulatory care settings, with a view to informing clinical and public health workers on the best tool for the diagnosis of SARS-CoV-2 infection in an evidence-based manner.

Methods

Search strategy and selection criteria
For this systematic review and meta-analysis, a standardised search was done in PubMed, OVID MEDLINE, Embase, and Web of Science, using the search term "(((((novel coronavirus) OR (ncov)) OR (SARS CoV 2)) OR (COVID19)) OR (covid)) AND (((((((((saliva) OR (nasopharyngeal swab)) OR (nasal swab)) OR (throat swab))) OR (oropharyngeal swab)) OR (posterior oropharyngeal saliva)) OR (nasopharyngeal aspirate)))". Given the role of preprints in timely dissemination of research studies during the COVID-19 pandemic, a search of the *medRxiv* and *bioRxiv* servers was also done using the search term "((SARS CoV 2) OR COVID19 OR covid) AND (saliva OR nasopharyngeal OR nasal OR throat OR oropharyngeal OR swab)". We also searched SSRN and Research Square for preprint literature containing the word "nasopharyngeal" in title, abstract, and keywords. The search on preprint servers was simplified because of their reduced search functionality. The search was done on Nov 16, 2020, with no language restrictions. Additional relevant articles from the reference sections were also reviewed.

Original clinical studies comparing the performance of nasopharyngeal swabs and any additional respiratory specimens for the diagnosis of suspected SARS-CoV-2 infection in individuals presenting in ambulatory care settings were included. Studies without data on paired samples or those that only examined different types of

sample from confirmed SARS-CoV-2 cases were excluded. Studies examining samples from confirmed cases can only report the positivity rate, result concordance, and viral load of different samples, which are not useful for examining the diagnostic performance of a test as a public health tool. Because clinical diagnosis involves the correct identification of individuals with infection from individuals with suspected clinical features or exposure history, only studies involving populations with suspected infection, including both positive and negative cases, allow the examination of diagnostic performance, through the calculation of sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV; appendix p 1). Studies testing only one part of a collection of specimens were excluded from this review to avoid potential bias due to selective testing.

Two authors (NNYT and HCS) screened articles, with disagreement resolved by consensus together with a third author (DKMI). Studies identified from different databases were de-duplicated after screening. Three authors (NNYT, HCS, and KYN) independently extracted data from the included studies, with disagreement resolved by consensus with a fourth author (DKMI). Two authors (NNYT and DKMI) assessed studies for methodological quality, including risk of bias and applicability, by use of the Scottish Intercollegiate Guidelines Network methodology checklist, adapted from Quality Assessment of Diagnostic Accuracy Studies tool for diagnostic studies.[37,38]

Data analysis
For included studies, either individual data or summary estimates of sample size, number of true positive, true negative, false positive, and false negative results in each study were extracted. By use of a standardised data extraction chart, we also retrieved information on study period, country, setting, disease prevalence, symptomatic status of population, sampling approaches, peer-reviewed status, and target genes assessed. These findings were checked for agreement.

The sensitivity, specificity, PPV, and NPV of RT-PCR tests and the 95% CIs were calculated and compared for different sampling methods, including pooled nasal and throat swabs, saliva, nasal swabs, throat swabs, with random effects meta-analyses using the inverse variance method[39] and restricted maximum likelihood estimator for heterogeneity,[40,41] with nasopharyngeal swabs as the reference because they are preferred by established guidelines. Freeman-Tukey double arcsine transformation was incorporated for normalising and stabilising the variance of sampling distribution of proportions, and pooled estimates were back-transformed using harmonic mean.[42] Sensitivity, specificity, PPV, and NPV of alternative clinical specimens were compared by constructing a fixed-effect model using the standard errors obtained from the random effects meta-analysis. Q statistic and its p value were calculated to test whether effect sizes depart
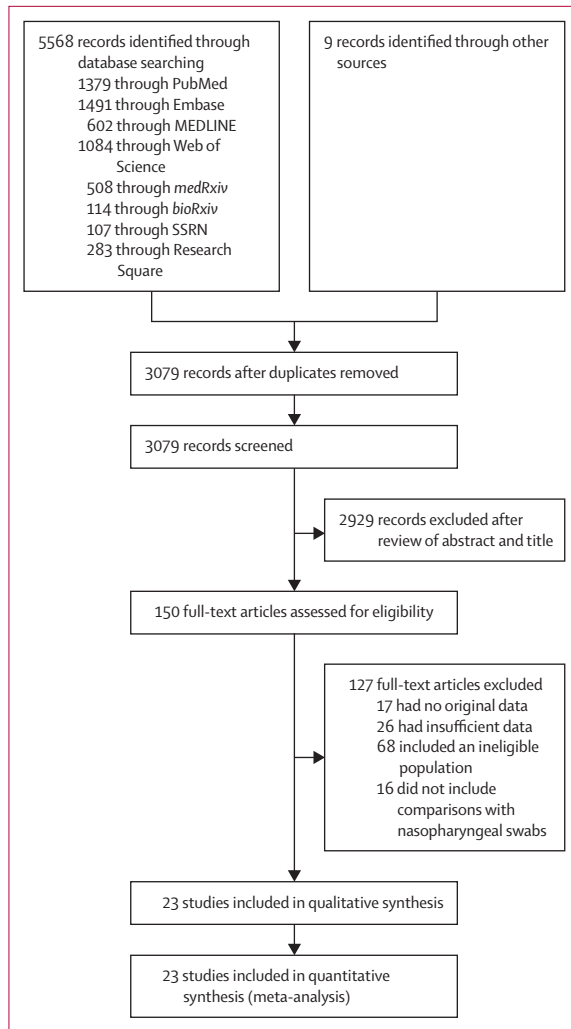
See **Online** for appendix

Figure 1: Study profile

5568 records identified through database searching
1379 through PubMed
1491 through Embase
602 through MEDLINE
1084 through Web of Science
508 through *medRxiv*
114 through *bioRxiv*
107 through SSRN
283 through Research Square

9 records identified through other sources

3079 records after duplicates removed

3079 records screened

2929 records excluded after review of abstract and title

150 full-text articles assessed for eligibility

127 full-text articles excluded
17 had no original data
26 had insufficient data
68 included an ineligible population
16 did not include comparisons with nasopharyngeal swabs

23 studies included in qualitative synthesis

23 studies included in quantitative synthesis (meta-analysis)

from homogeneity, and I^2 statistic and its 95% CI were calculated to examine the proportion of dispersion due to heterogeneity.[43,44] To assess possible factors contributing to heterogeneity, subgroup analyses were done for saliva, nasal swabs, and pooled nasal and throat swabs. Subgroup analysis for throat swab was precluded by the availability of only two studies. Study-level characteristics stratified included disease prevalence (<10% or ≥10%), geographical regions (USA or non-USA), symptomatic status of population (symptomatic only or symptomatic and asymptomatic), peer-reviewed status (yes or no), and number of target genes assessed in RT-PCR (one gene or two or more genes). For nasal swabs and pooled nasal and throat swabs, additional factors stratified included collection personnel (by health-care worker or self-collected by patient), swab materials (flocked or unflocked), and number of nostrils sampled (one or both). Comparison between samples collected by health-care workers and self-collection was presented using

forest plots for pooled nasal and throat swabs and nasal swabs, as allowable by the availability of studies examined the two different approaches. Scatterplots were used to present associations between the disease prevalence and the four performance indicators in included studies. Data were analysed with the metafor (version 2.4.0) and robvis (version 0.3.0) packages in R (version 3.6.0).

Role of the funding source
The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results
From the 5577 studies identified in our search, 2498 duplicates were excluded. After screening the titles and abstracts of the remaining articles, 150 full texts were screened (figure 1). On the basis of our selection criteria, 127 of those studies were excluded and 23 studies[21,47–69] met our inclusion criteria (table). Of these, 14 studies were from the USA,[21,47–59] four were from European countries,[60–63] two were from Eastern Mediterranean countries,[64,65] and the rest were from Canada,[66] India,[67] and China.[68] 7973 individuals with suspected SARS-CoV-2 infection, who were mostly symptomatic outpatients presenting to dedicated testing sites or emergency departments, were included from the 23 eligible studies. All studies used nasopharyngeal swabs collected by health-care workers as the reference gold standard, three studies examined pooled nasal and throat swabs (two on samples collected by health-care workers and one on self-collected samples), 13 examined self-collected saliva, seven examined nasal swabs (two on samples collected by health-care workers and five on self-collected samples), and two examined throat swabs collected by health-care workers. 16762 respiratory samples were included in our analysis, including 7973 nasopharyngeal swabs from each of the participants in the 23 studies and 1622 nasal swabs from seven studies, 6110 saliva samples from 13 studies, 338 throat swabs from two studies, and 719 pooled nasal and throat swabs from three studies. Among the 7973 individuals with suspected SARS-CoV-2 infection included in these studies, 1353 patients tested positive by nasopharyngeal swabs, giving an overall prevalence of 17·0% (95% CI 16·2–17·8). For individual studies, the prevalence of SARS-CoV-2 infection ranged from 4·3% to 84·1% (table).

Sensitivity measures the ability of a diagnostic test to correctly identify patients who have the disease with a positive test result. Using nasopharyngeal swabs as the reference, pooled nasal and throat swabs gave a sensitivity of 97% (95% CI 93–100), whereas saliva achieved a sensitivity of 85% (75–93), nasal swabs 86% (77–93), and throat swabs 68% (35–94; figure 2). The sensitivity of SARS-CoV-2 testing by pooled nasal and throat swabs was significantly higher than that of throat swabs (p=0·017). Specificity measures the ability of a test to

Author (year)	Study period	Location	Setting	Prevalence of SARS-CoV-2 (95% CI)	Patient characteristics	Sampling approaches	Target gene	Peer review	Evidence rating
Kojima et al (2020)[58]	Not reported	USA	Dedicated COVID-19 drive-through testing sites or specimen collection through home visit	48·8% (33·3–64·5)	Symptomatic and asymptomatic outpatients	Nasopharyngeal swab, self-collected nasal swab	N	Yes	High
Landry et al (2020)[47]	April 16, 2020–April 28, 2020	USA	Dedicated COVID-19 drive-through testing sites	26·6% (19·1–35·3)	Symptomatic outpatients	Nasopharyngeal swab, saliva	N	Yes	High
McCormick-Baw et al (2020)[48]	Not reported	USA	Accident and emergency department	31·6% (24·4–39·6)	Symptomatic outpatients	Nasopharyngeal swab, saliva	N	Yes	High
Migueres et al (2020)[60]	Not reported	France	Hospital	33·3% (25·1–42·4)	Symptomatic and asymptomatic outpatients	Nasopharyngeal swab, saliva	RdRp	Yes	High
Miller et al (2020)[49]	Not reported	USA	Two primary care medicine facilities	37·4% (27·4–48·1)	Symptomatic and asymptomatic outpatients	Nasopharyngeal swab, saliva	N	No	High
Callahan et al (2020)[50]	Not reported	USA	Dedicated COVID-19 drive-through or walk-up testing sites	23·3% (13·4–36·0)	Symptomatic and asymptomatic outpatients	Nasopharyngeal swab, self-collected nasal swab	N	No	High
Péré et al (2020)[61]	March, 2020	France	Hospital	84·1% (69·9–93·4)	Symptomatic outpatients	Nasopharyngeal swab, nasal swab	N, S	Yes	High
Tu et al (2020)[51]	March 16, 2020–March 21, 2020	USA	Ambulatory clinics	10·3% (7·8–13·3)	Symptomatic outpatients	Nasopharyngeal swab, self-collected nasal and mid-turbinate swabs	N	Yes	High
Patel et al (2020)[52]	Jan 27, 2020–Feb 29, 2020	USA	Sample submitted through Centers for Disease Control and Prevention	15·1% (9·7–21·9)	Symptomatic outpatients ≤7 days since illness onset	Nasopharyngeal swab, oropharyngeal swab	N	Yes	High
Wang et al (2020)[68]	Feb 16, 2020–March 2, 2020	China	Hospital	7·3% (4·0–11·9)	Outpatients with fever and x-ray abnormality	Nasopharyngeal swab, oropharyngeal swab	N, ORF	Yes	High
LeBlanc et al (2020)[66]	Not reported	Canada	Dedicated COVID-19 testing sites	17·9% (12·7–24·1)	Symptomatic and asymptomatic outpatients	Nasopharyngeal swab, pooled nasal and throat swab	E, ORF	Yes	High
Vlek et al (2020)[62]	April 21, 2020–April 29, 2020	Netherlands	Hospital	23·4% (15·7–32·5)	Symptomatic health-care workers	Nasopharyngeal swab, pooled nasal and throat swab	E	Yes	High
Griesemer et al (2020)[53]	March 20, 2020–March 26, 2020	USA	Two dedicated COVID-19 drive-through testing sites	22·2% (18·5–26·3)	Symptomatic and asymptomatic outpatients	Nasopharyngeal swab, nasal swab, saliva	N	No	High
Hanson et al (2020)[54]	May 29, 2020–June 25, 2020	USA	Dedicated COVID-19 drive-through testing sites	22·6% (18·3–27·3)	Symptomatic outpatients	Nasopharyngeal swab, saliva, self-collected nasal swab	ORF	Yes	High
Altawalah et al (2020)[64]	July 19, 2020–July 21, 2020	Kuwait	Hospital	38·6% (35·4–41·9)	Suspected COVID-19 admitted case	Nasopharyngeal swab, saliva	N, S, ORF	Yes	High
Barat et al (2020)[55]	July 13, 2020–Sept 18, 2020	USA	Drive-through testing sites and emergency department	6·4% (4·3–9·1)	Symptomatic outpatients	Nasopharyngeal swab, saliva	N	Yes	High
Procop et al (2020)[56]	Not reported	USA	Outpatient testing centre (ie, drive-through)	17·6% (12·8–23·3)	Symptomatic outpatients	Nasopharyngeal swab, saliva	N	Yes	High
Senok et al (2020)[65]	June 29, 2020–July 14, 2020	United Arab Emirates	Community-based COVID-19 screening facility	6·5% (4·3–9·4)	Symptomatic and asymptomatic outpatients	Nasopharyngeal swab, saliva	N	Yes	High
McCulloch et al (2020)[21]	March 31, 2020–April 13, 2020	USA	Drive-through testing clinics	7·1% (3·6–12·4)	Symptomatic outpatients and health-care workers	Nasopharyngeal swab, self-collected nasal swab	N	Yes	High
Shakir et al (2020)[57]	Not reported	USA	Dedicated COVID-19 drive-through testing sites	27·7% (23·5–32·3)	Symptomatic outpatients	Nasopharyngeal swab, self-collected pooled nasal and throat swab	E, ORF	Yes	High
Bhattacharya et al (2020)[67]	Not reported	India	Hospital	78·4% (67·3–87·1)	Symptomatic suspected patients	Nasopharyngeal swab, saliva	E, ORF	No	Acceptable
Yee et al (2020)[59]	June 8, 2020–Aug 28, 2020	USA	Hospital	22·7% (17·9–28·1)	Symptomatic and asymptomatic suspected patients	Nasopharyngeal swab, saliva	N, S, ORF	No	Acceptable
Mestdagh et al (2020)[63]	June, 2020–July, 2020	Belgium	Triage centres	4·3% (3·5–5·2)	Symptomatic and asymptomatic outpatients	Nasopharyngeal swab, saliva	E	No	Acceptable

E=envelope protein. N=nucleocapsid protein. ORF=open reading frame. RdRp=RNA-dependent RNA-polymerase. S=spike protein.

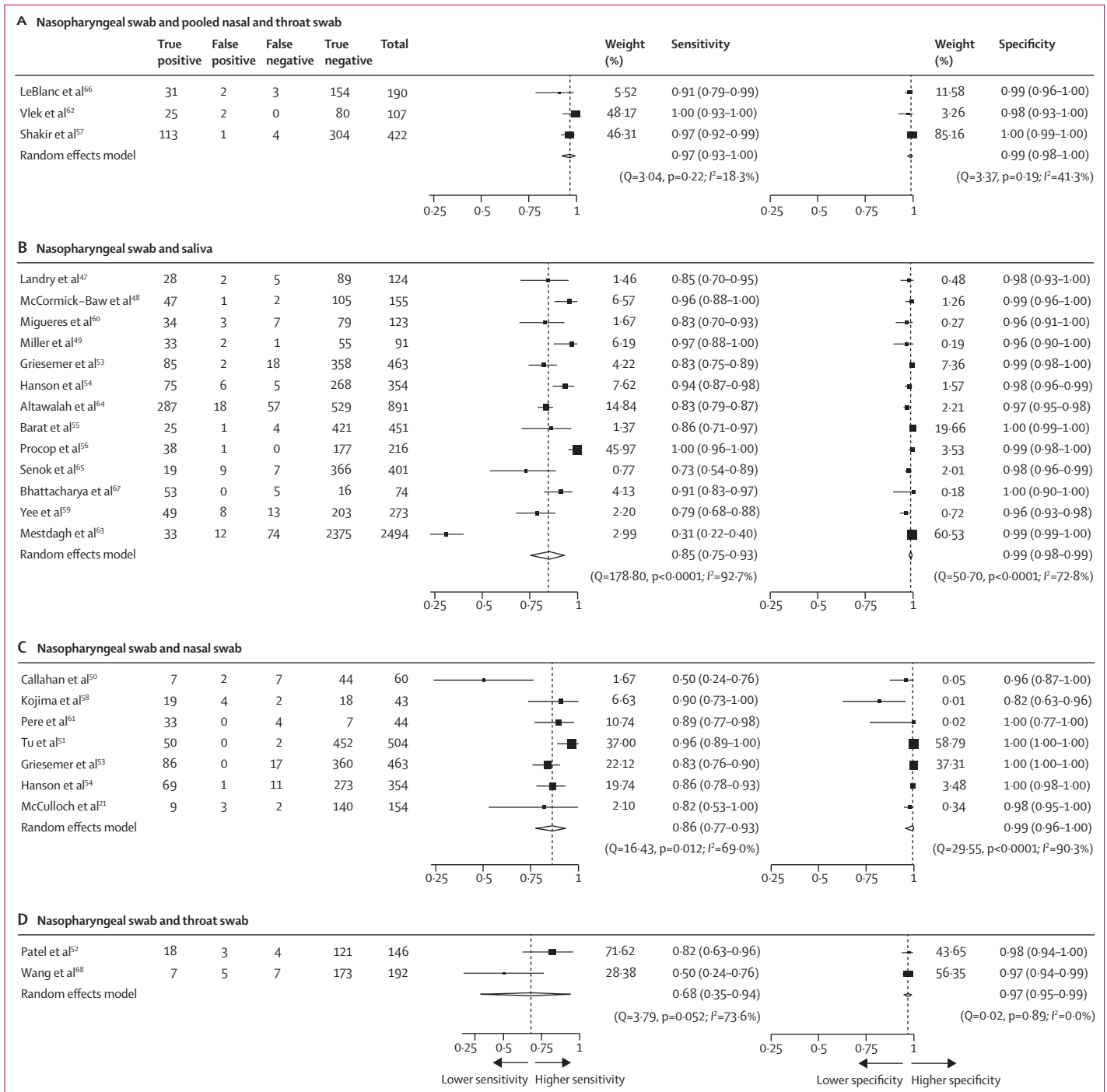*Table:* Characteristics of included studies

A Nasopharyngeal swab and pooled nasal and throat swab

	True positive	False positive	False negative	True negative	Total		Weight (%)	Sensitivity		Weight (%)	Specificity
LeBlanc et al[66]	31	2	3	154	190		5·52	0·91 (0·79–0·99)		11·58	0·99 (0·96–1·00)
Vlek et al[62]	25	2	0	80	107		48·17	1·00 (0·93–1·00)		3·26	0·98 (0·93–1·00)
Shakir et al[57]	113	1	4	304	422		46·31	0·97 (0·92–0·99)		85·16	1·00 (0·99–1·00)
Random effects model								0·97 (0·93–1·00)			0·99 (0·98–1·00)
								(Q=3·04, p=0·22; I²=18·3%)			(Q=3·37, p=0·19; I²=41·3%)

B Nasopharyngeal swab and saliva

	True positive	False positive	False negative	True negative	Total		Weight (%)	Sensitivity		Weight (%)	Specificity
Landry et al[47]	28	2	5	89	124		1·46	0·85 (0·70–0·95)		0·48	0·98 (0·93–1·00)
McCormick–Baw et al[48]	47	1	2	105	155		6·57	0·96 (0·88–1·00)		1·26	0·99 (0·96–1·00)
Migueres et al[60]	34	3	7	79	123		1·67	0·83 (0·70–0·93)		0·27	0·96 (0·91–1·00)
Miller et al[49]	33	2	1	55	91		6·19	0·97 (0·88–1·00)		0·19	0·96 (0·90–1·00)
Griesemer et al[53]	85	2	18	358	463		4·22	0·83 (0·75–0·89)		7·36	0·99 (0·98–1·00)
Hanson et al[54]	75	6	5	268	354		7·62	0·94 (0·87–0·98)		1·57	0·98 (0·96–0·99)
Altawalah et al[64]	287	18	57	529	891		14·84	0·83 (0·79–0·87)		2·21	0·97 (0·95–0·98)
Barat et al[55]	25	1	4	421	451		1·37	0·86 (0·71–0·97)		19·66	1·00 (0·99–1·00)
Procop et al[56]	38	1	0	177	216		45·97	1·00 (0·96–1·00)		3·53	0·99 (0·98–1·00)
Senok et al[65]	19	9	7	366	401		0·77	0·73 (0·54–0·89)		2·01	0·98 (0·96–0·99)
Bhattacharya et al[67]	53	0	5	16	74		4·13	0·91 (0·83–0·97)		0·18	1·00 (0·90–1·00)
Yee et al[59]	49	8	13	203	273		2·20	0·79 (0·68–0·88)		0·72	0·96 (0·93–0·98)
Mestdagh et al[63]	33	12	74	2375	2494		2·99	0·31 (0·22–0·40)		60·53	0·99 (0·99–1·00)
Random effects model								0·85 (0·75–0·93)			0·99 (0·98–0·99)
								(Q=178·80, p<0·0001; I²=92·7%)			(Q=50·70, p<0·0001; I²=72·8%)

C Nasopharyngeal swab and nasal swab

	True positive	False positive	False negative	True negative	Total		Weight (%)	Sensitivity		Weight (%)	Specificity
Callahan et al[50]	7	2	7	44	60		1·67	0·50 (0·24–0·76)		0·05	0·96 (0·87–1·00)
Kojima et al[58]	19	4	2	18	43		6·63	0·90 (0·73–1·00)		0·01	0·82 (0·63–0·96)
Pere et al[61]	33	0	4	7	44		10·74	0·89 (0·77–0·98)		0·02	1·00 (0·77–1·00)
Tu et al[51]	50	0	2	452	504		37·00	0·96 (0·89–1·00)		58·79	1·00 (1·00–1·00)
Griesemer et al[53]	86	0	17	360	463		22·12	0·83 (0·76–0·90)		37·31	1·00 (1·00–1·00)
Hanson et al[54]	69	1	11	273	354		19·74	0·86 (0·78–0·93)		3·48	1·00 (0·98–1·00)
McCulloch et al[21]	9	3	2	140	154		2·10	0·82 (0·53–1·00)		0·34	0·98 (0·95–1·00)
Random effects model								0·86 (0·77–0·93)			0·99 (0·96–1·00)
								(Q=16·43, p=0·012; I²=69·0%)			(Q=29·55, p<0·0001; I²=90·3%)

D Nasopharyngeal swab and throat swab

	True positive	False positive	False negative	True negative	Total		Weight (%)	Sensitivity		Weight (%)	Specificity
Patel et al[52]	18	3	4	121	146		71·62	0·82 (0·63–0·96)		43·65	0·98 (0·94–1·00)
Wang et al[68]	7	5	7	173	192		28·38	0·50 (0·24–0·76)		56·35	0·97 (0·94–0·99)
Random effects model								0·68 (0·35–0·94)			0·97 (0·95–0·99)
								(Q=3·79, p=0·052; I²=73·6%)			(Q=0·02, p=0·89; I²=0·0%)

Lower sensitivity Higher sensitivity   Lower specificity Higher specificity

Figure 2: **Meta-analysis of the sensitivity and specificity, using nasopharyngeal swab as a reference standard**
Forest plots of sensitivity and specificity. Squares (proportional to the sample size, disease prevalence, and heterogeneity) represent point estimates.

correctly identify people who do not have the disease with a negative test result. Comparably high specificities, ranging from 97% to 99%, were observed among all four clinical specimen collection methods (figure 2).

PPV represents the probability that a patient truly has the disease after having a positive test result. The highest PPV was given by pooled nasal and throat (97%, 95% CI 90–100) and nasal swabs (96%, 87–100), followed by saliva (93%, 88–97) and throat swabs (75%, 45–96; figure 3). NPV represents the probability that a patient truly does not have the disease after a negative test result. NPVs were generally comparable among different
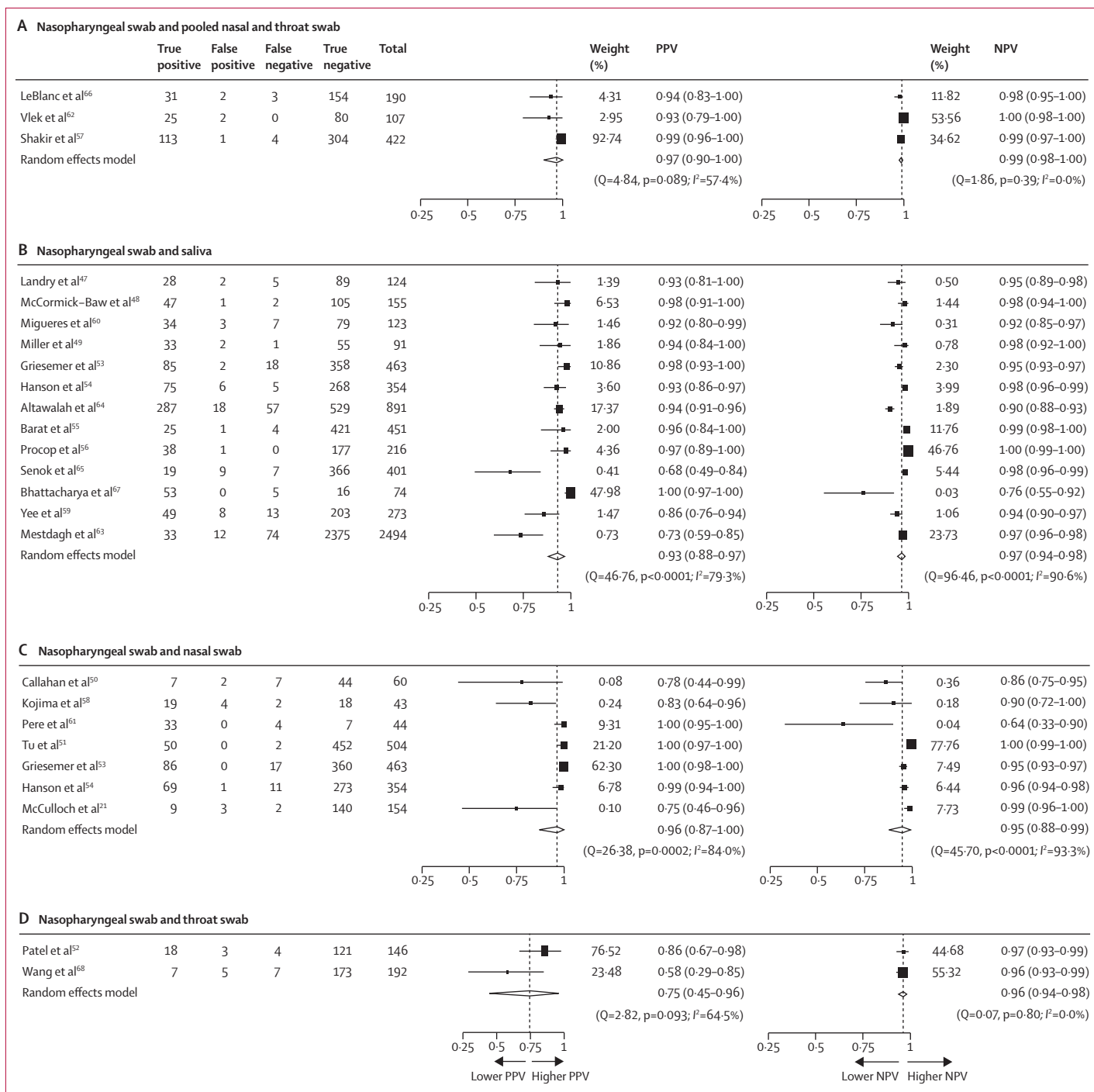
Figure 3: Meta-analysis of PPV and NPV, using nasopharyngeal swab as reference standard
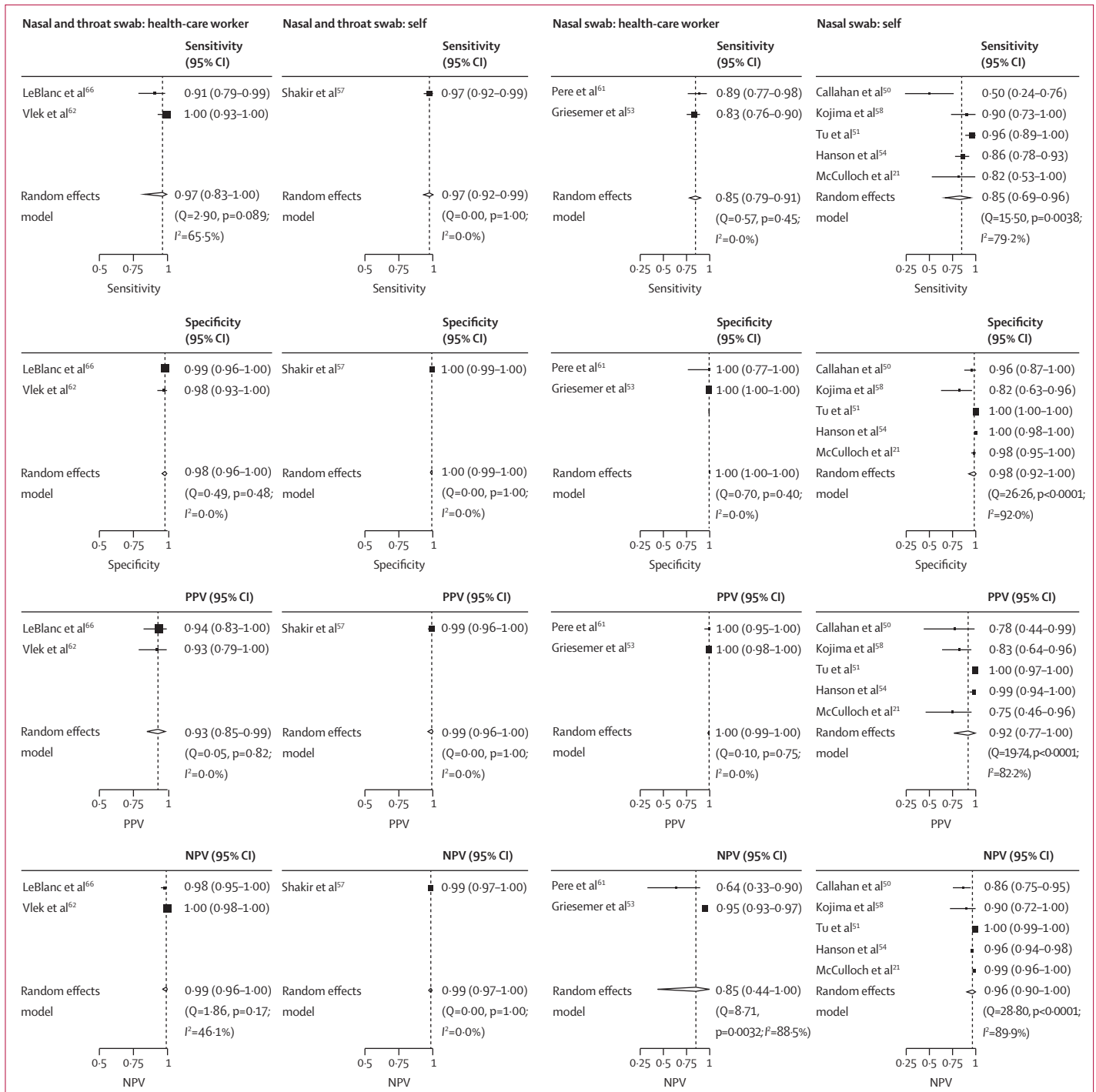Forest plots of PPV and NPV. Squares (proportional to the sample size, disease prevalence, and heterogeneity) represent point estimates. NPV=negative predictive value. PPV=positive predictive value.

clinical specimens, with highest value of 99% (95% CI 98–100) given by pooled nasal and throat swabs, followed by saliva (97%, 95% CI 94–98), throat swabs (96%, 94–98), and nasal swabs (95%, 88–99; figure 3).

No significant heterogeneity was observed in the analysis of pooled nasal and throat swab (I^2 of different

diagnostic estimates ranged between 0% and 57%) and throat swabs (I^2 of different diagnostic estimates ranged between 0% and 74%; appendix pp 7–10). Sensitivity, specificity, NPV, and PPV from studies on saliva and nasal swabs were more heterogeneous (I^2 ranged between 69% and 93%; appendix pp 7–10), with some particularly low

Figure 4: **Meta-analysis of the sensitivity, specificity, PPV, and NPV of health-care worker-collected and self-collected pooled nasal and throat swab and nasal swab**
Forest plots of sensitivity, specificity, positive predictive value, and negative predictive value. Squares (proportional to the weight in random effect models, accounted by sample size, disease prevalence, and heterogeneity) represent point estimates. NPV=negative predictive value. PPV=positive predictive value.

values from studies with a smaller inverse-variance weighting.[21,61,65,68] Generally no significant differences of the diagnostic performance indicators was found on stratified analyses for pooled nasal and throat swab. For

saliva, a lower sensitivity was observed for studies with disease prevalence of less than 10% (64% *vs* 90%), studies done outside the USA (74% *vs* 91%), studies involving asymptomatic individuals in the samples (76% *vs* 94%),

preprints (79% *vs* 89%), or studies that tested for one target gene in the PCR assay (75% *vs* 90%; appendix pp 7–10). A lower PPV in saliva was found for studies with a disease prevalence of less than 10% (81% *vs* 95%; appendix pp 7–10). For nasal swabs, a lower sensitivity was observed for studies that included samples from asymptomatic individuals (78% *vs* 90%) and in preprints (70% *vs* 91%; appendix pp 7–10). A lower PPV in nasal swabs was found for studies with a disease prevalence of less than 10% (75% *vs* 98%) and a lower NPV for studies done outside the USA (64% *vs* 96%; appendix pp 7–10). Heterogeneity generally remained moderate to substantial ($I^2 \geq 30\%$) in most stratified analyses, as defined according to the Cochrane Handbook for Systematic Reviews of Interventions.[70]

For pooled nasal and throat swabs, stratified comparison of samples collected by health-care workers (two studies) and self-collected samples (one study) showed no statistical difference in sensitivity, specificity, and NPV between the two approaches, and a slightly higher PPV for self-collected swabs (93% *vs* 99%; figure 4). For nasal swabs, all four indicators were similar between samples collected by health-care workers from two studies and self-collected samples from five studies (figure 4). Stratified comparison of collection personnel was not done for saliva and throat swabs, because all 13 studies on saliva were by self-collection and both studies on throat swabs were by collection by health-care workers. Scatterplots of the performance indicators against the range of disease prevalence of individual studies (from 4·3% to 84·1%) indicated that sensitivites and specificities were relatively stable across different disease prevalences (appendix p 2). However, PPVs showed an increasing trend for higher prevalence and a decreasing trend for lower prevalence (appendix p 2), whereas NPVs showed a decreasing trend for higher prevalence and an increasing trend for lower prevalence (appendix p 2).

Quality assessment showed that all included studies were of good or acceptable quality and low risk of bias (appendix p 3). Because six of the 23 studies were published on platforms without a formal peer-review process, a sensitivity analysis excluding these six articles was done. This analysis gave a similar result to the full analysis, with pooled nasal and throat swabs giving the highest sensitivity and PPV and a generally high specificity and NPV for all sample types (appendix pp 4–6).

Discussion

To our knowledge, this is the first systematic review and meta-analysis examining alternative specimen collection methods for SARS-CoV-2 RT-PCR testing, and reports the pooled analysis of sensitivity, specificity, PPV, and NPV to inform a comprehensive evaluation of the relative diagnostic performance of different sampling approaches for the diagnosis of SARS-CoV-2 infection among suspected cases. By comparing all the performance indicators of alternative specimen collection methods to the same reference gold standard (ie, nasopharyngeal swabs) and including only studies with standardised RT-PCR testing procedures, our review minimised potential bias from variation of testing techniques and allowed for a scientifically valid assessment of different sampling approaches.

Our findings showed that pooled nasal and throat swabs offered the best diagnostic performance, with high sensitivity (97%), specificity (99%), PPV (97%), and NPV (99%), making it the best alternative option for accurate laboratory testing. This result is compatible with that found by Lee and colleagues,[30] who also report a high percentage positive of combined oropharyngeal and nasal swabs, comparable to that of nasopharyngeal swabs. We also found that saliva and nasal swabs gave similar and very good diagnostic performance, with moderate sensitivities and high specificities, PPVs, and NPVs. This result is compatible with the moderate sensitivity observed in two previous reviews,[32,33] but contrasts with the lower percentage positive of saliva and nasal swabs in Lee et al.[30] These three alternative specimens should represent clinically acceptable alternatives to nasopharyngeal swabs for diagnosis of SARS-CoV-2 infection in the ambulatory care setting. Throat swabs gave a much lower sensitivity and PPV than nasopharyngeal swabs, a finding similar to that of Mohammadi and colleagues,[3] indicating that it is a worse specimen collection method and should not be recommended for the diagnosis of SARS-CoV-2 infection.

In a traditional screening setting, false positive results are associated with the concern of unnecessary invasive confirmatory tests and its associated risk and stress. In the context of the COVID-19 epidemic, cases labelled as positive might be relabelled as non-cases after they are followed up by relevant public health agencies and verified by further confirmatory testing. However, timely outbreak management and re-testing might only be possible when the number of false positive cases is not large.[69] Thus, there is a real risk of wrongful hospital admission or isolation with further unnecessary risk exposure for false-positive individuals or their close contacts. This can be particularly problematic when using throat swabs as the initial diagnostic specimen, given its much lower PPV of 75%. However, the likelihood of false positivity should be much lower for pooled nasal and throat swabs, nasal swabs, and saliva, as indicated by their high and comparable specificity (99%) and PPVs (93–97%).

By contrast, false negativity has more severe implications in the evolving COVID-19 pandemic, as cases not detected by the test would not be isolated and followed up with contact tracing and could seed further community transmission and infection. Although the high sensitivity (97%) of the pooled nasal and throat swab indicated its good detection power, the moderate sensitivities of saliva (85%) and nasal (86%) swabs suggest potential risk missing 14–15% of infected cases

on average. The much lower and substandard sensitivity of 68% for throat swabs suggests that they could missing almost a third (32%) of infected cases, giving a false negative result. Besides affecting accurate patient diagnosis, inadequate sensitivity to detect SARS-CoV-2 in a positive specimen with low viral load could also hinder the effectiveness of mass testing programmes of high-risk target groups, such as health-care workers,[71] or the entire population.[72]

Although generally regarded as the reference sample for SARS-CoV-2 testing in many countries, nasopharyngeal swabs have numerous limitations that have hindered their efficient and widespread use in ambulatory care settings, including their technical difficulty,[7] procedural discomfort, risk exposure implication, and the resulting expertise and facilities constraints.[8] The comparable diagnostic accuracy of alternative specimen collection methods, as shown in our findings, has practical benefits in clinical practice. Compared with nasopharyngeal swabs, pooled nasal and throat swabs, saliva, and nasal swabs are much less invasive and technically easier to collect.[73] The reduced procedural discomfort might help to prevent the triggering of gag reflexes, coughing, and sneezing and reduce the associated exposure risk for the health-care workers. The reduced requirement for trained health-care workers, high-level personal protective equipment, and negative-pressure facilities for collection of these alternative specimens will allow for their allocation to other competing needs in resource-constrained settings. The relative procedural simplicity could also allow for self-collection by patients or their relatives in different community settings.[74] Similar self-collection approaches have been adopted for the testing of influenza virus infection for diagnostic and surveillance purposes in various settings, with proven validity and acceptability.[75–77] In our analysis, the comparable performance profiles of self-collected and health-care worker-collected pooled nasal and throat swabs and nasal swabs, and the generally good performance of self-collected saliva, supported self-collection as a viable option and indicated that it was not associated with any significant impairment of diagnostic accuracy for the diagnosis of SARS-CoV-2 infection. The feasibility, accessibility, and acceptability of the self-collection for testing might help to facilitate the scaling up of SARS-CoV-2 testing in communities, with lower resource requirements and occupational exposure risk for health-care workers.

This review has several limitations. First, substantial heterogeneity was observed in several of the diagnostic performance indicators in studies on saliva and nasal swabs, which varied in terms of the disease prevalence, study location, symptom status of the study sample, and number of candidate genes tested. For example, studies on pooled nose and throat swabs were low in number, and those on nasal swabs collected by health-care workers were underrepresented compared with self-collected nasal swabs, which might have limited our power

to assess their diagnostic performance. A stratified comparison of throat swab collection by health-care workers and self-collection was not possible owing to the absence of studies examining its self-collection. Because no study has compared the use of gargle and nasopharyngeal swabs, it was also not possible for us to examine the comparative performance of gargle. Second, publication bias and selective reporting might have resulted in overestimation of some of the performance indicators we examined if studies with null or negative diagnostic performance were less likely to be published.[78] Although only a small number of relevant primary studies are available, our extended search of multiple literature databases and the inclusion of preprints should have helped to minimise the risk of publication bias in our review. Third, geographical coverage was skewed, with most studies of saliva and nasal swabs done in the USA and only a few from in European, Mediterranean, or Asian countries. Fourth, because the heterogeneity remained high in most stratified analyses, factors contributing to the residual heterogeneity might not have been accounted for and affected our observed results. Detailed exploration of the adjusted impact of these factors identified in subgroups analyses by use of a multivariate meta-regression approach was not possible with the small sample size and reduced power due to a small number of available studies. Finally, our study was primarily focused on the diagnosis of patients presenting with symptoms or history of exposure risk in an ambulatory care setting. Although this represented the most common clinical health care seeking scenario, our findings might have reduced generalisability to other settings with different disease prevalence or symptom profiles, such as specialist hospital wards or tertiary referral centres where infection is more likely, mass screening of individuals without symptoms, or testing an entire population with close to zero prevalence. Because our results indicated that the reported sensitivity and PPV were lower in studies done in settings with lower disease prevalence and that including asymptomatic individuals in the sample might contribute to heterogeneity, further prospective study is warranted to examine the effect of alternative sampling approaches on diagnostic performance when used in these settings.

In summary, in this review of all relevant published studies, we synthesised the pooled estimates of the diagnostic performance of different sampling approaches and found that, compared with the gold standard of nasopharyngeal swabs, pooled nasal and throat swabs offered the best diagnostic performance of the alternative sampling approaches for diagnosis of SARS-CoV-2 infection. Self-collection of pooled nasal and throat swabs appeared to be a viable option and did not associate with any significant impairment of diagnostic accuracy. Our results could inform evidence-based clinical practice, including the choice of suitable alternative sampling

approaches for the diagnosis of SARS-CoV-2 infection in the community to enable efficient downstream clinical and public health management, especially in situations where nasopharyngeal swabs are not practically feasible due either to reduced manpower and expertise, lack of protective facilities, or overloaded diagnostic testing need. With the added advantages of being less invasive and technically demanding, these alternative sampling approaches would help to boost acceptability and accessibility and facilitate the efficient scaling up SARS-CoV-2 testing in a community setting. Additionally, our pooled analysis result also provides a framework for the proper interpretation of testing results using different samples.

References
1 WHO. Weekly epidemiological update on COVID-19. March 22, 2021. https://www.who.int/publications/m/item/weekly-operational-update-on-covid-19---22-march-2021 (accessed March 23, 2020).
2 García LF. Immune Response, inflammation, and the clinical spectrum of COVID-19. *Front Immunol* 2020; **11:** 1441.
3 Mohammadi A, Esmaeilzadeh E, Li Y, Bosch RJ, Li JZ. SARS-CoV-2 detection in different respiratory sites: a systematic review and meta-analysis. *EBioMedicine* 2020; **59:** 102903.
4 Cao W, Li T. COVID-19: towards understanding of pathogenesis. *Cell Res* 2020; **30:** 367–69.
5 WHO. Laboratory testing for coronavirus disease (COVID-19) in suspected human cases: interim guidance, 19 March 2020. Geneva: World Health Organization, 2020.
6 Centers for Disease Control and Prevention. Interim Guidelines for Collecting, Handling, and Testing Clinical Specimens for COVID-19. July 8, 2020. Centers for Disease Control and Prevention, 2020. https://www.cdc.gov/coronavirus/2019-ncov/lab/guidelines-clinical-specimens.html (accessed Sept 22 2020).
7 Marty FM, Chen K, Verrill KA. How to obtain a nasopharyngeal swab specimen. *N Engl J Med* 2020; **382:** e76.
8 Karligkiotis A, Arosio A, Castelnuovo P. How to obtain a nasopharyngeal swab specimen. *N Engl J Med* 2020; **383:** e14.
9 Pettit SD, Jerome KR, Rouquié D, et al. 'All In': a pragmatic framework for COVID-19 testing and action on a global scale. *EMBO Mol Med* 2020; **12:** e12634.
10 Valentine-Graves M, Hall E, Guest JL, et al. At-home self-collection of saliva, oropharyngeal swabs and dried blood spots for SARS-CoV-2 diagnosis and serology: Post-collection acceptability of specimen collection process and patient confidence in specimens. *PLoS One* 2020; **15:** e0236775.
11 Tajima Y, Suda Y, Yano K. A case report of SARS-CoV-2 confirmed in saliva specimens up to 37 days after onset: Proposal of saliva specimens for COVID-19 diagnosis and virus monitoring. *J Infect Chemother* 2020; **26:** 1086–89.
12 Pinninti S, Trieu C, Pati SK, et al. Comparing nasopharyngeal and mid-turbinate nasal swab testing for the identification of SARS-CoV-2. *Clin Infect Dis* 2020; published online June 29. https://doi.org/10.1093/cid/ciaa882.
13 Harrington A, Cox B, Snowdon J, et al. Comparison of Abbott ID Now and Abbott m2000 methods for the detection of SARS-CoV-2 from nasopharyngeal and nasal swabs from symptomatic patients. *J Clin Microbiol* 2020; **58:** e00798-20.
14 Rhoads DD, Cherian SS, Roman K, Stempak LM, Schmotzer CL, Sadri N. Comparison of Abbott ID Now, DiaSorin Simplexa, and CDC FDA emergency use authorization methods for the detection of SARS-CoV-2 from nasopharyngeal and nasal swabs from individuals diagnosed with COVID-19. *J Clin Microbiol* 2020; **58:** e00760-20.
15 Nachtigall FM, Pereira A, Trofymchuk OS, Santos LS. Detection of SARS-CoV-2 in nasal swabs using MALDI-MS. *Nat Biotechnol* 2020; **38:** 1168–73.
16 Palmas G, Moriondo M, Trapani S, et al. Nasal swab as preferred clinical specimen for COVID-19 testing in children. *Pediatr Infect Dis J* 2020; **39:** e267–70.
17 Basu A, Zinger T, Inglima K, et al. Performance of Abbott ID Now COVID-19 rapid nucleic acid amplification test using nasopharyngeal swabs transported in viral transport media and dry nasal swabs in a New York City academic institution. *J Clin Microbiol* 2020; **58:** e01136-20.
18 Petruzzi G, De Virgilio A, Pichi B, et al. COVID-19: Nasal and oropharyngeal swab. *Head Neck* 2020; **42:** 1303–04.
19 Berenger BM, Fonseca K, Schneider AR, Hu J, Zelyas N. Sensitivity of nasopharyngeal, nasal and throat swab for the detection of SARS-CoV-2. *medRxiv* 2020; published online May 8. https://doi.org/10.1101/2020.05.05.20084889 (preprint).
20 Guest JL, Sullivan PS, Valentine-Graves M, et al. Suitability and sufficiency of telehealth clinician-observed, participant-collected samples for SARS-CoV-2 testing: the iCollect cohort pilot study. *JMIR Public Health Surveill* 2020; **6:** e19731.
21 McCulloch DJ, Kim AE, Wilcox NC, et al. Comparison of unsupervised home self-collected midnasal swabs with clinician-collected nasopharyngeal swabs for detection of SARS-CoV-2 infection. *JAMA Netw Open* 2020; **3:** e2016382.
22 Xu CLH, Raval M, Schnall JA, Kwong JC, Holmes NE. Duration of respiratory and gastrointestinal viral shedding in children with SARS-CoV-2: a systematic review and synthesis of data. *Pediatr Infect Dis J* 2020; **39:** e249–56.
23 Fakheran O, Dehghannejad M, Khademi A. Saliva as a diagnostic specimen for detection of SARS-CoV-2 in suspected patients: a scoping review. *Infect Dis Poverty* 2020; **9:** 100.
24 Weiss A, Jellingsø M, Sommer MOA. Spatial and temporal dynamics of SARS-CoV-2 in COVID-19 patients: a systematic review and meta-analysis. *EBioMedicine* 2020; **58:** 102916.
25 Arevalo-Rodriguez I, Buitrago-Garcia D, Simancas-Racines D, et al. False-negative results of initial RT-PCR assays for COVID-19: a systematic review. *medRxiv* 2020; published online Aug 13. https://doi.org/10.1101/2020.04.16.20066787 (preprint).
26 Zhu J, Guo J, Xu Y, Chen X. Viral dynamics of SARS-CoV-2 in saliva from infected patients. *J Infect* 2020; **81:** e48–50.
27 Bwire GM, Majigo MV, Njiro BJ, Mawazo A. Detection profile of SARS-CoV-2 using RT-PCR in different types of clinical specimens: a systematic review and meta-analysis. *J Med Virol* 2021; **93:** 719–25.
28 Böger B, Fachi MM, Vilhena RO, Cobre AF, Tonin FS, Pontarolo R. Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19. *Am J Infect Control* 2021; **49:** 21–29.
29 Mawaddah A, Gendeh HS, Lum SG, Marina MB. Upper respiratory tract sampling in COVID-19. *Malays J Pathol* 2020; **42:** 23–35.
30 Lee RA, Herigon JC, Benedetti A, Pollock NR, Denkinger CM. Performance of saliva, oropharyngeal swabs, and nasal swabs for SARS-CoV-2 molecular detection: a systematic review and meta-analysis. *medRxiv* 2020; published online Nov 13. https://doi.org/10.1101/2020.11.12.20230748 (preprint).

31 Czumbel LM, Kiss S, Farkas N, et al. Saliva as a candidate for COVID-19 diagnostic testing: a meta-analysis. *Front Med (Lausanne)* 2020; **7:** 465.

32 Butler-Laporte G, Lawandi A, Schiller I, et al. 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* 2021; **181:** 353–60.

33 Bastos ML, Perlman-Arrow S, Menzies D, Campbell JR. 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* 2021; published online Jan 12. https://doi.org/10.7326/M20-6569.

34 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 Patol Oral Cir Bucal* 2021; published online Feb 20. https://doi.org/10.4317/medoral.24424.

35 Nasiri K, Dimitrova A. Comparing saliva and nasopharyngeal swab specimens in the detection of COVID-19: a systematic review and meta-analysis. *J Dent Sci* 2021; published online Jan 29. https://doi.org/10.1016/j.jds.2021.01.010.

36 Riccò M, Ranzieri S, Peruzzi S, 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. *Acta Biomed* 2020; **91:** e2020025.

37 Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; **155:** 529–36.

38 Ma L-L, Wang Y-Y, Yang Z-H, Huang D, Weng H, Zeng X-T. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? *Mil Med Res* 2020; **7:** 7.

39 Veroniki AA, Jackson D, Bender R, et al. Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis. *Res Synth Methods* 2019; **10:** 23–43.

40 Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med* 1999; **18:** 2693–708.

41 Langan D, Higgins JPT, Jackson D, et al. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Res Synth Methods* 2019; **10:** 83–98.

42 Miller JJ. The inverse of the Freeman–Tukey double arcsine transformation. *Am Stat* 1978; **32:** 138.

43 Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327:** 557–60.

44 Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; **21:** 1539–58.

47 Landry ML, Criscuolo J, Peaper DR. Challenges in use of saliva for detection of SARS CoV-2 RNA in symptomatic outpatients. *J Clin Virol* 2020; **130:** 104567.

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

49 Miller M, Jansen M, Bisignano A, et al. Validation of a self-administrable, saliva-based RT-qPCR test detecting SARS-CoV-2. *medRxiv* 2020; published online June 9. https://doi.org/10.1101/2020.06.05.20122721 (preprint).

50 Callahan C, Lee RA, Lee GR, Zulauf K, Kirby JE, Arnaout R. Nasal-swab testing misses patients with low SARS-CoV-2 viral loads. *medRxiv* 2020; published online June 14. https://doi.org/10.1101/2020.06.12.20128736 (preprint).

51 Tu Y-P, Jennings R, Hart B, et al. Swabs collected by patients or health care workers for SARS-CoV-2 testing. *N Engl J Med* 2020; **383:** 494–96.

52 Patel MR, Carroll D, Ussery E, et al. Performance of oropharyngeal swab testing compared with nasopharyngeal swab testing for diagnosis of coronavirus disease 2019—United States, January 2020–February 2020. *Clin Infect Dis* 2021; **72:** 403–10.

53 Griesemer SB, Van Slyke G, Ehrbar D, et al. Evaluation of specimen types and saliva stabilization solutions for SARS-CoV-2 testing. *medRxiv* 2020; published online June 18. https://doi.org/10.1101/2020.06.16.20133041 (preprint).

54 Hanson KE, Barker AP, Hillyard DR, et al. Self-collected anterior nasal and saliva specimens versus health care worker-collected nasopharyngeal swabs for the molecular detection of SARS-CoV-2. *J Clin Microbiol* 2020; **58:** e01824-20.

55 Barat B, Das S, De Giorgi V, et al. Pooled saliva specimens for SARS-CoV-2 testing. *J Clin Microbiol* 2021; **59:** e02486-20.

56 Procop GW, Shrestha NK, Vogel S, et al. A direct comparison of enhanced saliva to nasopharyngeal swab for the detection of SARS-CoV-2 in symptomatic patients. *J Clin Microbiol* 2020; **58:** e01946-20.

57 Shakir SM, Barker AP, Hillyard DR, et al. Combined self-collected anterior nasal and oropharyngeal specimens versus provider-collected nasopharyngeal swabs for the detection of SARS-CoV-2. *J Clin Microbiol* 2020; **59:** 21.

58 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. *Clin Infect Dis* 2020; published online Oct 19. https://doi.org/10.1093/cid/ciaa1589.

59 Yee R, Truong T, Pannaraj PS, et al. Saliva is a promising alternative specimen for the detection of SARS-CoV-2 in children and adults. *medRxiv* 2020; published online Oct 27. https://doi.org/10.1101/2020.10.25.20219055 (preprint).

60 Migueres M, Mengelle C, Dimeglio C, et al. Saliva sampling for diagnosing SARS-CoV-2 infections in symptomatic patients and asymptomatic carriers. *J Clin Virol* 2020; **130:** 104580.

61 Péré H, Podglajen I, Wack M, et al. Nasal swab sampling for SARS-CoV-2: a convenient alternative in times of nasopharyngeal swab shortage. *J Clin Microbiol* 2020; **58:** e00721-20.

62 Vlek ALM, Wesselius TS, Achterberg R, Thijsen SFT. Combined throat/nasal swab sampling for SARS-CoV-2 is equivalent to nasopharyngeal sampling. *Eur J Clin Microbiol Infect Dis* 2020; published online July 14. https://doi.org/10.1007/s10096-020-03972-y.

63 Mestdagh P, Gillard M, Arbyn M, et al. Evaluation of saliva sampling procedures for SARS-CoV-2 diagnostics reveals differential sensitivity and association with viral load. *medRxiv* 2020; published online Oct 13. https://doi.org/10.1101/2020.10.06.20207902 (preprint).

64 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. *J Clin Virol* 2020; **132:** 104652.

65 Senok A, Alsuwaidi H, Atrah Y, et al. Saliva as an alternative specimen for molecular COVID-19 testing in community settings and population-based screening. *Infect Drug Resist* 2020; **13:** 3393–99.

66 LeBlanc JJ, Heinstein C, MacDonald J, Pettipas J, Hatchette TF, Patriquin G. A combined oropharyngeal/nares swab is a suitable alternative to nasopharyngeal swabs for the detection of SARS-CoV-2. *J Clin Virol* 2020; **128:** 104442.

67 Bhattacharya DD, Parai DD, Rout UK, et al. Saliva as a potential clinical specimen for diagnosis of SARS-CoV-2. *medRxiv* 2020; published online Sept 11. https://doi.org/10.1101/2020.09.11.20192591 (preprint).

68 Wang X, Tan L, Wang X, et al. Comparison of nasopharyngeal and oropharyngeal swabs for SARS-CoV-2 detection in 353 patients received tests with both specimens simultaneously. *Int J Infect Dis* 2020; **94:** 107–09.

69 Ramdas K, Darzi A, Jain S. 'Test, re-test, re-test': using inaccurate tests to greatly increase the accuracy of COVID-19 testing. *Nat Med* 2020; **26:** 810–11.

70 Deeks JJ, Higgins JPT, Altman DG. Cochrane handbook for systematic reviews of interventions. Version 6.2. 2021. https://training.cochrane.org/handbook/current/chapter-10#section-10-10 (accessed March 23, 2021).

71 Han Y, Yang Q, Liu Y, et al. Feasibility study of mixing throat swab samples for severe acute respiratory syndrome coronavirus-2 screening. *Virol Sin* 2020; **35:** 830–32.

72 Schmidt M, Hoehl S, Berger A, et al. Novel multiple swab method enables high efficiency in SARS-CoV-2 screenings without loss of sensitivity for screening of a complete population. *Transfusion* 2020; published online July 6. https://doi.org/10.1111/trf.15973.

73   Martinez RM. Clinical samples for SARS-CoV-2 detection: review of the early literature. *Clin Microbiol Newsl* 2020; **42:** 121–27.

74   Goyal S, Prasert K, Praphasiri P, et al. The acceptability and validity of self-collected nasal swabs for detection of influenza virus infection among older adults in Thailand. *Influenza Other Respir Viruses* 2017; **11:** 412–17.

75   Ip DKM, Schutten M, Fang VJ, et al. Validation of self-swab for virologic confirmation of influenza virus infections in a community setting. *J Infect Dis* 2012; **205:** 631–34.

76   Sueki A, Matsuda K, Yamaguchi A, et al. Evaluation of saliva as diagnostic materials for influenza virus infection by PCR-based assays. *Clin Chim Acta* 2016; **453:** 71–74.

77   To KK, Lu L, Yip CC, et al. Additional molecular testing of saliva specimens improves the detection of respiratory viruses. *Emerg Microbes Infect* 2017; **6:** e49.

78   Song F, Khan KS, Dinnes J, Sutton AJ. Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. *Int J Epidemiol* 2002; **31:** 88–95.