JCM Accepted Manuscript Posted Online 27 January 2021
J Clin Microbiol doi:10.1128/JCM.02881-20

Performance of Saliva, Oropharyngeal Swabs, and Nasal Swabs for SARS-CoV-2 Molecular Detection: A Systematic Review and Meta-analysis

Rose A. Lee MD, MSPH[a,b], Joshua C. Herigon MD, MPH, MBI[a], Andrea Benedetti PhD[d], Nira R. Pollock MD, PhD[b, c], Claudia M. Denkinger MD, PhD[e]

[a]Division of Infectious Diseases, Department of Pediatrics, Boston Children's Hospital, Boston, USA
[b]Division of Infectious Diseases, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, USA
[c]Department of Laboratory Medicine, Boston Children's Hospital, Boston, USA
[d]Department of Medicine and of Epidemiology, Biostatistics & Occupational Health, McGill University, Montreal, Canada
[e]Division of Clinical Tropical Medicine, Center of Infectious Diseases, Heidelberg University Hospital, Heidelberg, Germany

Corresponding Author:
Nira Pollock, M.D., Ph.D., D(ABMM)
Associate Medical Director, Infectious Diseases Diagnostic Laboratory, Boston Children's Hospital,
Division of Infectious Diseases, Beth Israel Deaconess Medical Center,
Associate Professor of Pathology and Medicine, Harvard Medical School
Farley Building 8th floor, Room FA828
300 Longwood Ave 45 Boston, MA 02115
Phone: 857 -218 -5113 Fax: 617 -730 -0383
email: nira.pollock@childrens.harvard.edu

ABSTRACT:

Background: Nasopharyngeal (NP) swabs are considered the highest-yield sample for diagnostic testing for respiratory viruses, including SARS-CoV-2. The need to increase capacity for SARS-CoV-2 testing in a variety of settings, combined with shortages of sample collection supplies, have motivated a search for alternative sample types with high sensitivity. We systematically reviewed the literature to understand the performance of alternative sample types compared to NP swabs.

Methods: We systematically searched PubMed, Google Scholar, medRxiv, and bioRxiv (last retrieval October 1st, 2020) for comparative studies of alternative specimen types [saliva, oropharyngeal (OP), and nasal (NS) swabs] versus NP swabs for SARS-CoV-2 diagnosis using nucleic acid amplification testing (NAAT). A logistic-normal random-effects meta-analysis was performed to calculate % positive alternative-specimen, % positive NP, and % dual positives overall and in sub-groups. The QUADAS 2 tool was used to assess bias.

Results: From 1,253 unique citations, we identified 25 saliva, 11 NS, 6 OP, and 4 OP/NS studies meeting inclusion criteria. Three specimen types captured lower % positives [NS (82%, 95% CI: 73-90%), OP (84%, 95% CI: 57-100%), saliva (88%, 95% CI: 81 – 93%)] than NP swabs, while combined OP/NS matched NP performance (97%, 95% CI: 90-100%). Absence of RNA extraction (saliva) and utilization of a more sensitive NAAT (NS) substantially decreased alternative-specimen yield.

Conclusions: NP swabs remain the gold standard for diagnosis of SARS-CoV-2, although alternative specimens are promising. Much remains unknown about the impact of variations in specimen collection, processing protocols, and population (pediatric vs. adult, late vs. early in disease course) and head-to head studies of sampling strategies are urgently needed.

Introduction:

Testing for SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2, the etiologic agent of COVID-19), has preferentially utilized nasopharyngeal (NP) sampling with flocked swabs. This sampling method is presumed to have the highest diagnostic yield, as evidenced by its use as a reference method by the Food and Drug Administration (FDA).(1). However, NP sampling requires professional collection (a major limitation given the strained health care system in almost all parts of the world) and protective equipment which, similar to the flocked swabs themselves, is in short supply. In addition, NP sampling is uncomfortable, limiting patients' willingness to come forward for testing especially if asymptomatic.

If large scale testing of symptomatic and asymptomatic patients is to become a reality, innovation in sampling is as important as innovation in testing. Innovation in sampling requires consideration of sampling methods (e.g., choice of swab, choice of sample type) as well as self-sampling.

As of October 2020, interim guidance from the Centers for Disease Control and Prevention (CDC) (2) recommends upper respiratory tract testing with any of the following specimens: NP swab, NP wash/aspirate, nasal wash/aspirate, oropharyngeal (OP) swab, nasal mid-turbinate (MT) swab using a flocked tapered swab, an anterior nares (AN) nasal swab using a flocked or spun polyester swab, or a saliva specimen obtained by supervised self-collection. The FDA, in contrast, states that NP, OP, MT, and AN swab samples are appropriate for clinical testing, and that "more data are necessary to better understand the performance when using specific saliva collection devices or other specimen types for COVID-19 testing" (1).

Saliva is the subject of the largest body of research on alternative specimen types as well as the specimen of choice for numerous companies putting high-volume testing programs in place. Certain regions of the world, such as Hong Kong, have already adopted saliva in their mass screening protocols (3). The pathophysiological rationale for saliva sampling is based on the angiotensin-converting enzyme II (ACE-2) being the cellular receptor for SARS-CoV-2 (4), similar to SARS-CoV (5). High ACE-2 receptor expression in oral mucosa and salivary glands has been recently demonstrated (6, 7). Salivary gland duct epithelial cells previously were identified as a target for SARS-CoV in a rhesus macaque model (8). One study detected SARS-CoV-2 in saliva collected via expression directly from the salivary gland duct (9). These findings suggest that saliva may be a suitable and high-yield diagnostic sample for SARS-CoV-2 testing based on local viral replication, in addition to the possible mixing in saliva of lower and upper respiratory tract fluids that can carry virus. Saliva offers several advantages as it is a non-invasive sample type that can be self-collected, thus decreasing infectious risk to medical personnel, use of personal protective equipment, and reliance on equipment subject to supply shortages such as nasal, OP, or NP swabs.

OP (throat) swabs have been in wide use for SARS-CoV-2 diagnosis since the beginning of the pandemic (3). Consensus is lacking on both the best collection approach and nomenclature for oral specimens. The terms "oropharyngeal" and "throat" have been used in the literature, and therefore we describe and group them here as "oropharyngeal "or OP swabs. No comprehensive evaluations of their performance compared to other methods exist. OP swabs are less specialized than NP swabs and thus OP samples can be collected with a broader range of swab products. While the FDA recommends that OP swabs be collected by a health care professional (1), some have suggested that self-swabbing might be possible.

Journal of Clinical Microbiology

Nasal swabs (NS), another important alternative specimen type, also have the advantages of increased comfort and possible self-collection. They have been classified into two types anatomically. Nasal mid-turbinate (MT) swabs, also called deep nasal swabs, are defined by the CDC (2) as flocked/tapered swabs performed while tilting the patient's head back 70 degrees and inserting the swab less than one inch (about 2 cm into the nostril) until resistance is met at the turbinates before rotating the swab several times against the nasal wall. Anterior nares (AN) swabs are defined without a head tilt and inserting the entire swab at least 0.5 inch (1cm) inside the nostril (naris) and sampling the membrane by various methods, including rotating the swab in place or around the inside wall of the nostril multiple times, and/or leaving in place for 10 to 15 seconds. The current "lower nasal swab" protocol sanctioned by the FDA specifies swab insertion "until you feel a bit of resistance" and thus matches the MT depth defined by the CDC, though swab type is not specified (10). The CDC and FDA suggest swabbing both nares for sampling.

A large body of literature on the yield of these alternative sample types for nucleic acid amplification testing (NAAT) has been created but a comparative systematic summary of the relative performance of these alternative specimen types, compared to NP swabs, is missing and is needed to inform decision makers. Furthermore, the data for self-collection of these alternative specimens (1, 2) have not been systematically evaluated. Here, we aim to clarify the performance of alternative specimen types for diagnosis of SARS-CoV-2 by systematically reviewing and meta-analyzing the literature on this topic available through October 2020.

Methods:

This systematic review and meta-analysis is reported in accordance with PRISMA guidelines (see Supplementary File for the PRISMA checklist). The protocol for this work was registered in the International Prospective Register of Systematic Reviews (PROSPERO) (identifier: CRD42020214660).

Search strategy, information sources, and eligibility criteria

We performed a comprehensive search of the following databases (Pubmed/MEDLINE and Google Scholar) as well as the preprint servers medRxiv and bioRxiv to identify relevant studies from January 1st, 2020 until October 1st, 2020. Only English language articles were allowed. An example search strategy is provided in Supplementary Methods. Additional studies were retrieved by screening the reference lists of the included articles and from archives of the reviewers. We excluded papers from the same hospital with overlapping inclusion dates, to avoid including patients more than once and thus minimize bias in the data (11). Cross-sectional, case-control, and cohort studies and randomized controlled trials were included independent of number of specimens tested. Conference proceedings and abstracts were deemed ineligible. Participants of all age groups with presumed SARS-CoV-2 infections, in all settings, were included. We included only papers utilizing NAAT for SARS-CoV-2 detection.

Data Extraction

Two reviewers assessed all articles (R.L. and J.H.) independently and disagreements were resolved with input of a third investigator (N.R.P. or C.M.D.).

We compared alternative sampling to NP sampling. We extracted only data for positive NAAT results on at least one sample type and only when sampling methods being compared were performed

Journal of Clinical Microbiology

synchronously. If patients were tested serially over time and data could not be extracted for specific timepoints, we excluded the study. Asymptomatic, symptomatic, and unspecified symptomatology patient cohorts were included. If multiple different RT-PCR tests were performed on one sample within one study, we utilized the data from the best-performing assay (highest positive detection rate) for a sample type. If a study contributed data to more than one analysis (e.g., two different alternative sample types in one study, each compared to NP swab), it was considered as two or more datasets. If a combined NP/OP sample was used, and no data was available for NP sampling alone, we included the study, using NP/OP as the comparator. For each specimen type, in addition to data on test performance, we extracted data for factors likely to affect test performance as detailed below. Data on throat or gargle washes were not included in this meta-analysis (12, 13). We obtained the limit of detection (LOD) from studies by direct report within the study when available and otherwise by manufacturer claims (package insert if available, Supp. Table 1 lists how LOD was ascertained by study).

Our study retrieval process is depicted in Supplementary Figure 1. Records were organized using a reference manager (Zotero Version 5.0.89, George Mason University).

Assessment of methodological quality

The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool, a validated quality assessment tool for diagnostic studies, was used to assess the included studies' risk of bias (14). The four domains assessed for risk of bias and applicability include: 1) participant selection; 2) index test; 3) reference test; 4) flow and timing.

Data analysis

Under the assumption that 1) we do not know in advance which specimen type performs best for SARS-CoV-2 detection and 2) false positives are very infrequent for NAATs performed in qualified laboratories, we focused only on individuals positive for at least one sample type and report on agreement between samples rather than measures of sensitivity and specificity. Specifically, for each study, we included all individuals with paired specimens who had at least one positive specimen as the denominator, and calculated 1) % positive Alt [percent of individuals positive by the alternative specimen type (e.g., saliva, NS, or OP swab)], 2) % positive NP [percent of individuals positive by the NP specimen (or NP/OP in a minority of cases)], and 3) % dual positive (percent of individuals positive for both the alternative specimen and NP specimen).

We present results of the systematic review in forest plots. Meta-analyses were only performed when there were at least 4 primary studies and at least 20 patients per study. The % positive Alt, % positive NP, and % dual positive for pooled, overall, and in sub-groups were estimated using a logistic-normal random-effects model (via the 'metaprop_one' command with Stata (15)). We applied the Freeman-Tukey Double Arcsine transformation to stabilize variances and score 95% intervals were computed. Heterogeneity was measured by the inconsistency index (I^2), which describe the percentage of variation across studies due to heterogeneity rather than chance. To investigate possible contributors to heterogeneity, we present sub-group analyses by (1) site of sampling; (2) swab material (flocked, unflocked); (3) sampling procedure; (4) professional vs. self-sampling; (5) specimen processing/NAAT (including dilution, nucleic acid extraction procedure, and assay used); (6) populations (e.g., pediatric, asymptomatic); (7) symptom duration prior to testing. Due to heterogeneity in sub-groups and small sample size we report differences between subgroups descriptively but did not perform direct statistical

Journal of Clinical Microbiology

testing between sub-groups. The robustness of the meta-analysis to publication bias was assessed by the symmetry of funnel plots.

All analyses and graphs were performed using Stata 15.1 (StataCorp, Texas) and GraphPad 8.5 (Prism, SanDiego).

Results:

Our search yielded 1,253 unique citations, of which 25 were included in the analyses for saliva, 11 for NS, 6 for OP, and 4 for OP/NS swabs. Reasons for exclusions of studies are laid out in Supplementary Figure 1.

The studies we included in the meta-analysis overall had a moderate-high risk of bias according to QUADAS-2 (Supplementary Table 2,3) with studies on NS having a low to moderate risk (Supplementary Table 4). Many studies did not specify patient selection methodology (random or consecutive). Many studies were also comprised of cohorts of known positives (case-control) that were then re-tested with paired specimens and therefore at a high risk of a selection bias limiting applicability of results to the general screening population.

Saliva

In total, 25 studies (16–40) met our inclusion criteria to assess saliva as an alternative sample type for SARS-CoV-2 diagnosis. The studies cumulatively included 4,528 paired saliva and NP swab specimens, although two studies used a combined NP/OP swab as a comparator (24, 39). An additional 13 studies (3, 9, 41–51) described a performance estimate for saliva as a specimen type, but these studies were excluded from the meta-analysis as the saliva and NP samples were either not collected synchronously, or multiple paired specimens were taken from the same set of patients and the data could not be extracted for a unique patient/time point.

Across the 25 studies, we found that the % positive saliva [88%, 95% confidence interval (CI) 81 – 93%] was lower than the % positive NP (or NP/OP) although not substantially different [94% (95% CI 90 – 98%)]. The % dual positive was noticeably lower than either specimen type alone [79%, (95% CI 71 – 86%), Figure 1], indicating relatively poor agreement. Considerable heterogeneity was also detected (I^2 88.6%). This heterogeneity was likely attributable to the variation in study procedures and patient population between studies, as outlined in Supplementary Table 5. Notably, there were no head-to-head studies for any of the comparisons described below for differences in collection, processing, and populations.

Saliva collection protocols for included studies were assessed for differences with respect to 1) asking patients to cough or clear their throat before submission of sample (likely mixed sputum and saliva specimen or deep throat saliva specimen) or 2) requesting the patients submit “drool” or “spit”. While some authors (42) have hypothesized that capture of posterior oropharyngeal saliva or mixed sputum/lower respiratory specimen is important for diagnostic sensitivity, we did not find a considerable difference in performance, although % positive saliva was higher for studies (33, 34, 37, 40) that specified cough or deep throat saliva specimen vs. studies that did not specifically ask for this [94% (95% CI: 87-99%) vs. 86% (95% CI: 78-92%), Fig. 2]. For NP samples in these two groups, the % positive detection was similar [89% (95% CI: 60-100%) vs. 95% (95% CI: 93-97%)]. Notably, many studies

that supported the hypothesis of a coughing or deep throat saliva being better than drool/spit sample were excluded from the meta-analysis due to non-synchronous sample collection or repeat testing on the same set of patients (3, 41–43).

Other differences in sampling between studies included avoidance of eating, drinking, or brushing teeth before specimen submission (typically 30 minutes to 2 hours). Protocols that specified avoidance of eating, drinking, or brushing teeth prior to saliva collection (22, 23, 26, 31–34, 40) had higher % positive saliva, although the difference was not substantial [91% (95% CI 86 – 95%) vs. 86% (95% CI 79 – 92%), Fig. 2]. For NP sampling in the same two groups of studies, the difference was minimal [94% (95% CI: 80-100%) vs. 95% (95% CI: 92-97%)]. A few studies specifically requested morning submission (33, 40) in addition to avoidance of food/drink and similar saliva performance was identified between protocols that specified morning submission versus those that did not [92% positive saliva (95% CI: 88-96%) vs. 87% (95% CI: 79-93%), respectively]. However, for NP sampling in the two groups of protocols, the % positive NP was lower with morning submission [66% positive (95% CI: 60-72%) vs. 96% positive (95% CI: 94-97%)], but this was largely driven by one study (33) in which NP swabs performed poorly (this was an outpatient cohort of patients who were asymptomatic at the time of collection and had received their diagnosis at least one week prior).

While many studies specified self-collected saliva (16, 17, 20, 25–27, 29, 30, 33, 34, 37, 39, 40), some studies described supervised collection (22, 28, 31, 35, 38). The % positive saliva was higher for self-collection than supervised collection although the difference was not substantial [92% (95% CI: 86-96%) vs. 83% (95% CI: 60-98%), respectively, Fig. 2]. Notably, only two studies (26, 33) reported using RNase P as a control for human material sampling adequacy.

There were also substantial differences in specimen processing, including variable dilution of the saliva specimen prior to NAAT, and use of protocols that directly input saliva samples into the NAAT without nucleic acid extraction. We found that even in the absence of dilution for saliva, the % positive could not match that of NP swabs. Studies utilizing undiluted saliva specimens (22, 26) had similar % positive saliva to studies utilizing diluted saliva (16, 17, 23–25, 27–32, 35, 37, 40) [Fig. 2, 92% (95% CI: 86-97%) vs. 89% (95% CI: 81-95%)]. All NP swabs were eluted in viral transport media, and performance was similar in these two groups of studies [97% (95% CI: 91-100%) for undiluted saliva vs. 94% (95% CI: 92-96%) for diluted saliva]. In studies using diluted saliva, there was wide variation in dilution methods, with many groups not specifying the degree of dilution.

Studies that did not use a nucleic acid extraction step (19, 28) but instead directly input the saliva specimen into the amplification assay without any pre-processing showed substantially lower % positive saliva than studies that had an extraction step [60% (95% CI: 49%-70%) vs. 89% (95% CI: 83%-94%), Fig. 2]. In contrast, % positive NP swab was not different between the two groups of studies [89% (95% CI: 81-95%) vs. 95% (95% CI: 90-98%)], with all except one study (28) using a nucleic acid extraction for the NP swab sample (28).

While the majority of studies utilized reverse transcription polymerase chain reaction (RT-PCR) assays, one study (19) used reverse transcription loop-mediated isothermal amplification (RT-LAMP), and two studies used transcription mediated amplification (TMA) (35, 36). NAATs utilized also differed over ten-fold in their limits of detection between studies, which has been previously described to affect diagnostic positivity (52). We found slightly higher, but not substantially different, percentages of positive detection in studies with a more sensitive/lower [<1000 copies/milliliter (cp/mL)] limit of

Journal of Clinical Microbiology

detection (LOD) (20, 21, 24, 34, 39) compared to studies with LOD ≥ 1000 cp/mL (23, 26, 27, 31–33, 51) [90% (95% CI: 79-97%) vs. 87% (95% CI: 77 – 95%)]. The differences were similar for NP swabs [100% (95% CI: 97-100%) vs. 91% (95%CI: 71-100%)].

We also assessed if patient symptomatology could explain variable diagnostic performance between saliva and NP sampling. We found that only a few studies provided a direct comparison between asymptomatic (16, 19, 33) and symptomatic patient data (17, 18, 21–23, 25, 27, 28, 34, 35, 37, 39, 40) that could be parsed and extracted for analysis. We found that % positive saliva was similar between asymptomatic and symptomatic patients [87% (95% CI: 70-98%) vs. 88% (95% CI: 79-95%) respectively]. The difference between % positive NP in asymptomatic vs. symptomatic patients was much larger [symptomatic 96% (95% CI: 93-99%) vs. asymptomatic 73% (95% CI: 47%-93%, Fig. 2)]. These findings, however, were driven by one study (33) with superior saliva performance in asymptomatic patients [% positive saliva 93% (95% CI: 88-97%) vs. % positive NP 52% (95% CI: 44-60%)]. As discussed above, this study (33) included an outpatient cohort who had received their diagnosis at least one week prior. Notably, although the patient population is described as "asymptomatic," it is unclear if this was just at the time of collection and they had symptoms closer to their initial diagnosis.

Another important question that we were not able to adequately address from the literature is the performance of saliva in pediatric populations. There are two studies (51, 53) that evaluate the performance of saliva in children, both of which showed worse performance compared to NP swabs [8/11 children positive by NP swab were positive by saliva in one study (51), and in the other study 53% of the children detected by NP swab were also positive by saliva (53)]. However, the timing of saliva collection versus NP swab collection in both studies was unclear, results for asymptomatic vs. symptomatic children were not clearly distinguished, and sample processing methods were not clearly described.

Invalid test results were also not consistently reported across studies. Viscosity of saliva was highlighted in some studies as increasing errors in automated pipetting steps, necessitating dilution or biochemical pre-treatment of samples (26, 35, 44).

Given that the fluctuation of viral load over time may differ between saliva and NP samples, and the timepoints patients present themselves for diagnosis may vary, we also assessed differences in diagnostic performance at different timepoints throughout the illness. Sub-group analysis of 6 studies with extractable data (25, 27, 28, 36, 39, 51) found that % positive saliva was overall lower >7 days after symptom onset [74% (95%CI: 62-85%)] compared to ≤ 7 days [89% (95% CI: 73-99%)], which was also observed for NP swabs [91% (95% CI: 82-98%) vs. 99% (95% CI: 90-100%), respectively].

OP swab

We identified six studies that assessed OP vs. NP swabs (54–59) and were suitable for meta-analysis. Given the paucity of data, subgroup analyses to assess differences in collection procedure, sample processing, and patient symptomatology were limited (Supplementary Figure 2). We found that % positive OP swab was similar to % positive NP swab [84% (95% CI: 57-100%) vs. 88% (95% CI: 73-98%)] although % dual positive was only 68% (95% CI: 36-93%) suggesting limited agreement (Figure 3). Notably, the % positive NP estimate was unusually low in this group of studies, which is largely driven by one study (56) with a large gap in performance between % positive OP and % positive NP [86% (95% CI: 65-97%) vs. 41% (95% CI: 21-64%) respectively] that was not observed in other studies. This study was

Journal of Clinical Microbiology

unique in that paired samples were taken near the end of a hospitalization (unclear duration of symptoms) in a cohort comprised of known positives from prior NP swab RT-PCR positive patients to help determine discharge eligibility.

Another outlier study in this dataset was a cohort of symptomatic patients (55) in a Wuhan hospital early on in the pandemic with unusually poor performance of OP swabs [36% (95% CI: 25-47%)] in comparison to NP swabs [88% (95% CI: 79-94%)]. There were not sufficient data in the OP swab meta-analysis to assess the impact of symptomatology or duration of symptoms on percent positives.

Only 3 studies specified flocking of swabs (54, 57, 59), and in all 3 cases unflocked oropharyngeal swabs were compared to flocked NP swabs. The percent positive detection rate for these 3 studies was similar between OP and NP sampling [96% (95% CI: 88-100%) vs. 97% (95% CI: 90-100%)].

There were two studies that used healthcare-worker collected oral swabs (54, 57) and the positive detection rate for OP vs. NP sampling was similar [97% (95% CI: 89 – 100%) vs. 96% (95% CI: 87- 100%) respectively, Fig. 3]. One study used unobserved self-collected OP swabs and reported similar performance between sample types but in a limited sample set of only 12 positive patients [92% (95% CI: 62-100%) for OP vs. 100 (95% CI: 74-100%) for NP)]. The 3 other studies (55, 56, 58) in this meta-analysis did not specify self- vs. healthcare-worker collection. There was one study (60) that compared self- vs. lab-technician collected OP swabs and found that only 14/24 total positives were detected by self-collection versus 22/24 total positives for lab-technician-collected. This study was excluded from this meta-analysis, however, as the pairing of OP swabs to NP swabs was unclear.

Three studies that assessed oral swab specimens were excluded from this meta-analysis as they were not oropharyngeal. There was one pediatric study of buccal swabs in Singapore (61) where children underwent daily NP and buccal swabs; 9/11 children with SARS-CoV-2 detected by NP swab also at some point had positive buccal swabs. One study that described sampling of the anterior 2/3rd of the dorsum of the tongue (62) found similar positive detection rates from “tongue” swabs (46/51 total positives) and NP swabs (49/51 total positives). Another study asked patients to cough prior to sampling oral fluid in cheeks, gums, hard palate, and tongue (63), and reported that detection was slightly higher (26/29 total positives) than NP swabs (23/29 total positives).

AN and MT Swabs

We identified 11 studies using either AN or MT swab that could be combined for pooled meta-analysis (32, 35, 54, 59, 62–68) with NP swabs as the reference sample type, although one study (59) used a combined NP/OP swab for reference. We found that the % positive NS (either AN or MT) (82%, 95% CI: 73– 90%), was substantially lower than % positive NP swab (98%, 95% CI: 96 – 100%) as well as % dual positive (79%, 95% CI: 69 – 88%), suggestive of limited agreement between the two sample types (Fig. 4). Considerable heterogeneity was again detected between studies (I^2 = 87%), which we attributed to differences in study procedures. Accordingly, we assessed performance by collection protocol, self-collection/supervision/healthcare-worker-collection, sample processing, and patient symptoms (Supplementary Table 6). Notably there were no head-to-head studies for any of the comparisons.

The pooled estimates of % positive NS via AN (35, 62) versus MT swabs (using the depth of insertion to classify AN vs. MT sampling, per CDC description) were similar (54, 59, 63, 66–68) [90% (95% CI: 84-94%) vs. 84% (95% CI: 65 – 97%), Fig. 5]. % positive NP was similar between groups [99% (95% CI: 95-

100%) vs. 97% (95% CI: 92-100%) respectively]. There were only two studies that assessed AN swabs, and notably, one of the studies (62) also compared AN to MT sampling (AN performed before MT). Performance was similar between AN and MT swabs (48 out of 51 total positives for AN vs. 50 out of 52 total positives for MT). In this study, both the AN and MT swabs were collected from both nares, and the AN swab was unflocked, whereas the MT swab was flocked.

We found that a more sensitive assay (LOD <1000 copies/mL) (64, 65, 68) resulted in worse performance of NS in comparison to assays with LOD ≥ 1000 copies/mL (32, 54, 59, 63, 66, 67) [61% (95% CI: 40-79%) vs. 85% (95% CI: 82-91%)], while % positive NP were similar (99% (95% CI: 93-100%) vs. 97% (95% CI: 92-100%) respectively). This may reflect lower viral burden in the mid-turbinate/anterior nares region than the nasopharynx resulting in lower performance in comparison to NP swabs that is only evident when using a highly sensitive (LOD < 1000 copies/mL) assay. Notably, the discordant paired samples described in the studies also were found to have lower viral loads than concordant pairs.

Two of the studies (64, 68) with the worst NS performance used a more sensitive assay and also compared unflocked NS to flocked NP swabs (positive NS detection 48-56%). There were 3 other studies (35, 54, 59), however, that also used unflocked NS compared to flocked NP swabs and reported higher detection (83-100%). Ultimately, NS studies that specified that both NS and NP swabs were flocked (63, 66) had a substantially higher % positive NS than studies where an unflocked NS was compared to a flocked NP swab (35, 54, 59, 64, 68), although confidence intervals were wide and overlapping (90% (95% CI: 81-97%) vs. 77% (95% CI 55-93%), Fig. 5). This finding may have been partially driven by the two studies using the lower LOD assay. % positive NP between these two groups were similar [97% (95% CI: 90-100%) vs. 97% (95% CI: 94-99%)]. One study (62) collected unflocked AN swabs, flocked MT swabs, and unflocked NP swabs for comparison and reported similar performance of all three specimens as described above. However, use of a non-flocked swab for the NP sampling may have decreased its sensitivity and artificially increased the sensitivity of MT and AN sampling.

Only 4 studies (35, 59, 62, 68) specified that NS were performed before NP swabs and although % positive NS was higher in comparison to studies that did not specify the swab order, there was not a substantial difference [85% (95% CI: 58-100%) vs. 81% (95% CI: 74-88%), Fig. 5]. % positive NP was the same regardless of NS order [98% (95% CI: 95-100%)]. Surprisingly, NS specimens collected from both nares (35, 54, 62, 64, 67, 68) seemed to perform worse in comparison to swabs collected from a single nostril (59, 63, 66), although the difference was not substantial [77% (95% CI: 59-91%) vs. 93% (59-91%), Fig. 5]. % positive NP was again similar in these two scenarios [95% (95% CI: 95-99%) for both nares vs. 97% (95% CI: 83-100%) for one nare]. Notably, this finding may again have been driven by the two studies (64, 68) utilizing a more sensitive < 1000 cp/mL NAAT assay (both with poor detection by NS and performed with swabs collected in both nares), a factor which we described previously as associated with a lower NS detection rate.

We also found that unsupervised self-collected NS specimens (35, 59, 62) had higher percent positives in comparison to swabs collected by healthcare workers (54, 64, 67, 68) [93%, 95% CI: 85-98% vs. 68%, 95% CI: 47-86%, Fig. 5]. In only one study (63), the patient's self-collection was supervised (92%, 95% CI: 74-99%). Professional NP sampling performance was similar between the NS self-collection and HCW-collection groups [99% (95% CI: 96-100%) vs. 96% (95%CI: 92-99%)]. Notably, the same two studies (64, 68) showing the worst NS performance (and using more sensitive assays) were in the healthcare-worker

Journal of Clinical Microbiology

collected group. Additionally, only two studies (62, 63) reported use of RNase P as a human sampling control.

We could not find any studies that specifically assessed only asymptomatic patients with NS, although multiple studies used mixed populations and a few studies specified that all of the cohort was symptomatic. Studies of only-symptomatic patients (35, 62, 64–67) had a similar % positive NS rate to studies of mixed or unspecified patients [82% (95% CI: 73-91%) vs. 83% (95% CI: 62-97%), Fig. 5]. % positive NP was similar between groups [99% (95% CI: 98-100%) vs. 96% (95% CI: 88-100%)]. There were two studies (63, 67) with extractable data to assess the impact of symptom duration prior to testing, and a lower yield after 7 days was found although the difference was not substantial [76% (95% CI: 60-87%) for > 7 days symptoms vs. 88% (95% CI: 74-95%) for ≤ 7 days, Fig. 5]. % positive NP was similar in both groups [100% (95% CI: 86-100%) vs. 100% (98-100%)].

Again, data were limited on pediatric populations when using nasal swabs. There are two pediatric studies of nasal samples; one study (69) described NS to be outperforming OP swabs in 56 paired samples from 11 pediatric patients, with Ct values lower in NS versus OP for all 11 first paired samples. This study had to be excluded due to repeat sampling, as we were not able to extract unique patient data from different timepoints. The other study (70) was not a nasal swab study, but describes NP aspirates to be outperforming NP swabs for detection (% positive for NP aspirates was 88% in comparison to 51% for NP swabs), though methods details provided were minimal and repeat sampling on patients occurred.

Combined OP/NS as a specimen type

Four studies (59, 71–73) evaluated combined oropharyngeal and nasal swabs in comparison to NP swabs (Fig. 6). Three of these studies (71–73) specified that a single swab was used for collection of an OP/NS sample, whereas in one study it was unclear if two separate swabs were used for OP and NS sampling and the results compiled (59). Two of the studies (59, 73) used MT swab depth, while one used AN sampling (72), and one was unspecified (71). Three of the studies were healthcare-collected swabs, in which the OP sampling was performed prior to NS (71–73), and only two of these studies specified swabbing both nares (72, 73). Two studies used flocked swabs (71, 73), while the others used unflocked swabs. The LOD of the assay in copies/mL was only available in one study (59) and > 1000 cp/mL. Two of the studies (59, 71) found that the percent positive detection of the combined swab specimen was greater than percent positive detection for the reference NP swabs. Pooled detection estimates were similarly high between the combined swabs and NP swabs with the same % positive estimate [97% (95% CI: 90 – 100%)] although agreement between the two methods was less [90% dual positive (95% CI: 84 – 96%)].

Assessment of publication bias

Visual inspection of a funnel plot of the study data versus standard error shows substantial asymmetry and therefore suggest publication bias for all alternative sample types skewed towards publication of positive findings (i.e. strong performance of alternative specimens) (Supplementary Fig. 3).

Discussion:

This systematic review and meta-analysis synthesizes a large number of studies comparing alternative sample types to NP swab for SARS-CoV-2 detection by NAAT. While all 3 sample types independently

seemed to capture lower % positives [nasal swabs (82%, 95% CI: 73-90%), OP swabs (84%, 95% CI: 57-100%) and saliva (88%, 95% CI: 81 – 93%)] in comparison to NP swabs, combined OP/nasal swabs in 4 studies interestingly had the same % positive detection rate as NP swabs (97%, 95% CI: 90-100%) (Fig. 6).

For saliva specimens, we found slightly lower performance compared to NP swabs overall [88% (95% CI: 81-93%) vs. 94% (95% CI: 90-98%)], and only the absence of nucleic acid extraction resulted in a substantially lower rate of detection. One study (44) had to be excluded from the meta-analysis due to non-synchronous collection of NP and saliva samples. This study is of interest nevertheless, as it tested different RT-PCR platforms on the same set of saliva specimens using 3 different extraction-free commercial RT-PCR assays against a standard RT-PCR assay with extraction and reported that 79, 81, and 52 specimens were detected, respectively, out of a total of 84 positive specimens detected on the standard assay, indicating that choice of extraction-free assay matters.

Self-collection, coughing or deep throat saliva, and avoiding food, drink, or toothbrushing resulted in >5% increased positive saliva detection rates, although the difference was not substantial. Collection of saliva greater than 7 days after symptom onset also resulted in >10% lower detection, although the difference was not substantial. Viscosity has been described qualitatively in multiple studies as a challenge in utilization of saliva as a specimen type, but invalid rates were not documented in many studies; we note that invalid rates are a critical parameter and should be consistently reported.

We found that OP swabs seemed to perform similarly to saliva and NP swabs, but these estimates were highly affected by one study (56) where samples were collected near the end of a hospitalization for discharge purposes and the % positive NP was unusually low in comparison to the rest of the literature. Overall, the data argue for provider-facilitated collection and against self-collection of this sample type.

For NS, the literature to date supports that they perform worse than NP swabs although these findings were largely driven by two studies (64, 68). There was no substantial difference between AN and MT swab detection, and we note that AN/MT swab collection protocols may have overlapped in practice; we defined AN vs. MT based on depth of insertion, as the swab type was not always specified. We hypothesize that the difference in NS to NP performance was driven by use of a particularly sensitive (< 1000 cp/mL LOD) assay for two of the studies. We observed that the discordant samples (NP+/NS-) typically had low viral loads on the NP sample. While this may reflect differences in viral burden anatomically between NP and AN/MT sampling, the clinical significance of this difference remains to be determined.

We hypothesize that nasal swab performance is likely highly dependent on collection procedure, which has in turn evolved over time and a substantial amount of data remains unpublished (personal communication Nira Pollock). Self-collection and single-nare sampling trended towards higher although not substantially different % positive detection, although these results were largely driven by the two studies with poor NS performance (64, 68). Detection was also lower after > 7 days of symptoms although this difference was not substantial. There remain unresolved questions on the best performing swab material (spun polyester, foam, rayon) and sample transport (buffer, dry swab) that could not be addressed in this study due to limited data (many studies did not report the swab product or elution details). We also still do not fully understand the impact of flocking on this specimen type. While studies using flocked nasal swabs had a slightly higher % positive detection in comparison to unflocked swab

Journal of Clinical Microbiology

studies, the unflocked swab group included two studies using a lower LOD assay, and the 3 studies that used a less sensitive assay had near 90% detection (similar to the flocked group).

OP/NS samples had surprisingly high performance compared to NP swabs; while this sample type likely requires an operator for collection, additional studies are warranted to understand acceptability in adult and pediatric patients.

Overall, much remains unknown about the impact on diagnostic sensitivity of variations in specimen collection and processing protocols, and performance in specific key sub-populations (asymptomatic vs. symptomatic, pediatric vs. adult, late vs. early testing from symptom onset). While we assess aggregate data from different studies to gain insight into these variables, a limitation of this meta-analysis is that true comparison is precluded in the absence of head-to-head studies. Furthermore, while there are trends we observe in our subgroup analyses, these findings may be population-related and should be interpreted with caution. Timing of sampling from symptom onset was also quite variable (collection occurred within days to weeks in some studies), and was inconsistently reported, which likely had a major impact on diagnostic performance given decreasing viral load over time. Head-to-head studies are urgently needed of flocked vs. unflocked swabs (and specialized vs. unspecialized swabs for MT collection), collected at different times in disease and with different sampling methods, and also in important subpopulations (e.g. children), to resolve the persistent uncertainty.

We note that the reporting quality of studies was low, STARD guidelines (74) were not consistently followed, and study bias was considered moderate to high on QUADAS 2. Lastly, in this study we chose to report the % positive alternative-specimen, % positive comparator-specimen, and % dual positives instead of the positive percent agreement (PPA). This decision was motivated by our presumption regarding the low rate of false-positives using NAAT, and the potential for an alternative to yield more positive results than the comparator NP, which would otherwise not be considered.

In summary, while alternative specimens (particularly saliva and OP/NS samples) show promise, we find that the literature to date suggests that NP swabs are indeed the gold standard in comparison to alternative specimen types (saliva, OP swab, NS). We identify self-collected nasal swabs and saliva to have similar performance to healthcare-worker obtained specimens, which is also supported by a head-to-head comparison of self-collected AN versus professionally-collected NP swabs utilizing an antigen rapid test for detection (75). We reiterate that the LOD of any assay will impact detection and centers should be aware of the increased possibility of false negatives with any sample type when using a less sensitive assay. Given the promising results of combined oropharyngeal and nasal swab studies, more studies on alternative specimen combinations would be useful. Lastly, we encourage future studies to provide more clarity about exact details of collection procedures, specific swab shape and materials used, sample processing methods (dilution, extraction, storage, transport), and NAAT assay utilized (including LOD), allowing the field to clearly define the tradeoffs required to sufficiently bring SARS-CoV-2 testing to scale.

Financial Support: none

Potential Conflicts of Interests: All authors have no conflicts of interest to declare.

509

510

511

512

513

514

515

516

517 References:

518 1. Food and Drug Administration, Center for Devices and Radiological Health. FAQs on
519 Testing for SARS-CoV-2. https://www.fda.gov/medical-devices/coronavirus-covid-19-and-
520 medical-devices/faqs-testing-sars-cov-2. Accessed on October 12, 2020.

521 2. Centers for Disease Control and Prevention. Information for Laboratories about Coronavirus
522 (COVID-19). https://www.cdc.gov/coronavirus/2019-ncov/lab/guidelines-clinical-
523 specimens.html. Accessed on October 4, 2020.

524 3. Wong SCY, Tse H, Siu HK, Kwong TS, Chu MY, Yau FYS, Cheung IYY, Tse CWS, Poon
525 KC, Cheung KC, Wu TC, Chan JWM, Cheuk W, Lung DC. 2020. Posterior Oropharyngeal
526 Saliva for the Detection of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-
527 CoV-2). Clinical Infectious Diseases https://doi.org/10.1093/cid/ciaa797.

528 4. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-
529 L, Chen H-D, Chen J, Luo Y, Guo H, Jiang R-D, Liu M-Q, Chen Y, Shen X-R, Wang X,
530 Zheng X-S, Zhao K, Chen Q-J, Deng F, Liu L-L, Yan B, Zhan F-X, Wang Y-Y, Xiao G-F,
531 Shi Z-L. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat
532 origin. Nature 579:270–273.

533 5. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, Somasundaran M, Sullivan JL,
534 Luzuriaga K, Greenough TC, Choe H, Farzan M. 2003. Angiotensin-converting enzyme 2
535 is a functional receptor for the SARS coronavirus. Nature 426:450–454.

6. Xu J, Li Y, Gan F, Du Y, Yao Y. 2020. Salivary Glands: Potential Reservoirs for COVID-19 Asymptomatic Infection. Journal of Dental Research https://doi.org/10.1177/0022034520918518.

7. Xu H, Zhong L, Deng J, Peng J, Dan H, Zeng X, Li T, Chen Q. 2020. High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. International Journal of Oral Science 12:1–5.

8. Liu L, Wei Q, Alvarez X, Wang H, Du Y, Zhu H, Jiang H, Zhou J, Lam P, Zhang L, Lackner A, Qin C, Chen Z. 2011. Epithelial cells lining salivary gland ducts are early target cells of severe acute respiratory syndrome coronavirus infection in the upper respiratory tracts of rhesus macaques. J Virol 85:4025–4030.

9. Chen L, Zhao J, Peng J, Li X, Deng X, Geng Z, Shen Z, Guo F, Zhang Q, Jin Y, Wang L, Wang S. 2020. Detection of SARS-CoV-2 in Saliva and Characterization of Oral Symptoms in COVID-19 Patients. Cell Prolif https://doi.org/10.1111/cpr.12923.

10. HealthPulse. Nasal Swab Instructions. https://healthpulsenow.org/collect. Accessed on Ocotber 29, 2020.

11. Bauchner H, Golub RM, Zylke J. 2020. Editorial Concern—Possible Reporting of the Same Patients With COVID-19 in Different Reports. JAMA 323:1256.

12. Mittal A, Gupta A, Kumar S, Surjit M, Singh B, Soneja M, Soni KD, Khan AR, Singh K, Naik S, Kumar A, Aggarwal R, Nischal N, Sinha S, Trikha A, Wig N. 2020. Gargle lavage as a viable alternative to swab for detection of SARS-CoV-2. Indian J Med Res 152:77–81.

556 13. Wang W-K, Chen S-Y, Liu I-J, Chen Y-C, Chen H-L, Yang C-F, Chen P-J, Yeh S-H, Kao
557 C-L, Huang L-M, Hsueh P-R, Wang J-T, Sheng W-H, Fang C-T, Hung C-C, Hsieh S-M, Su
558 C-P, Chiang W-C, Yang J-Y, Lin J-H, Hsieh S-C, Hu H-P, Chiang Y-P, Wang J-T, Yang P-
559 C, Chang S-C. Detection of SARS-associated Coronavirus in Throat Wash and Saliva in
560 Early Diagnosis. Emerging Infectious Diseases Journal
561 https://doi.org/10.3201/eid1007.031113.

562 14. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang
563 MMG, Sterne JAC, Bossuyt PMM. 2011. QUADAS-2: A Revised Tool for the Quality
564 Assessment of Diagnostic Accuracy Studies. Annals of Internal Medicine 155:529–536.

565 15. Nyaga VN, Arbyn M, Aerts M. 2014. Metaprop: a Stata command to perform meta-analysis
566 of binomial data. Archives of Public Health 72:39.

567 16. Yokota I, Shane PY, Okada K, Unoki Y, Yang Y, Inao T, Sakamaki K, Iwasaki S,
568 Hayasaka K, Sugita J, Nishida M, Fujisawa S, Teshima T. 2020. Mass screening of
569 asymptomatic persons for SARS-CoV-2 using saliva. Clinical Infectious Diseases
570 https://doi.org/10.1093/cid/ciaa1388.

571 17. Yokota I, Hattori T, Shane PY, Konno S, Nagasaka A, Takeyabu K, Fujisawa S, Nishida
572 M, Teshima T. 2020. Equivalent SARS-CoV-2 viral loads between nasopharyngeal swab
573 and saliva in symptomatic patients. medRxiv https://doi.org/10.1101/2020.09.01.20186254.

574 18. Bhattacharya DD, Parai DD, Rout UK, Nanda RR, Kanungo DS, Dash DGC, Palo DSK,
575 Giri DS, Choudhary HR, Kshatri DJS, Turuk DJ, Mishra DB, Dash DS, Pati DS. 2020.

576 Saliva as a potential clinical specimen for diagnosis of SARS-CoV-2. medRxiv
577 https://doi.org/10.1101/2020.09.11.20192591.

578 19. L'Helgouach N, Champigneux P, Santos-Schneider F, Molina L, Espeut J, Alali M,
579 Baptiste J, Cardeur L, Dubuc B, Foulongne V, Galtier F, Makinson A, Marin G, Picot M-C,
580 Prieux-Lejeune A, Quenot M, Checa-Robles FJ, Salvetat N, Vetter D, Reynes J, Molina F.
581 2020. EasyCOV : LAMP based rapid detection of SARS-CoV-2 in saliva. medRxiv
582 https://doi.org/10.1101/2020.05.30.20117291.

583 18. Rutgers University. Accelerated Emergency Use Authorization (EUA) Summary SARS-
584 CoV-2 Assay (Rutgers Clinical Genomics Laboratory).
585 https://www.fda.gov/media/136875/download. Accessed on October 2nd, 2020.

586 21. SoRelle J, Mahimmainathan L, McCormick-Baw C, Cavuoti D, Lee F, Bararia A, Thomas
587 A, Sarode R, Clark AE, Muthukumar A. 2020. Evaluation of symptomatic patient saliva as
588 a sample type for the Abbott ID NOW COVID-19 assay. medRxiv
589 https://doi.org/10.1101/2020.06.01.20119198.

590 22. McCormick-Baw C, Morgan K, Gaffney D, Cazares Y, Jaworski K, Byrd A, Molberg K,
591 Cavuoti D. 2020. Saliva as an Alternate Specimen Source for Detection of SARS-CoV-2 in
592 Symptomatic Patients Using Cepheid Xpert Xpress SARS-CoV-2. J Clin Microbiol
593 https://doi.org/10.1128/JCM.01109-20.

594 23. Becker D, Sandoval E, Amin A, Hoff PD, Diets A, Leonetti N, Lim YW, Elliott C, Laurent
595 L, Grzymski J, Lu J. 2020. Saliva is less sensitive than nasopharyngeal swabs for COVID-

19 detection in the community setting. medRxiv https://doi.org/10.1101/2020.05.11.20092338.

24. Pasomsub E, Watcharananan SP, Boonyawat K, Janchompoo P, Wongtabtim G, Suksuwan W, Sungkanuparph S, Phuphuakrat A. 2020. Saliva sample as a non-invasive specimen for the diagnosis of coronavirus disease 2019: a cross-sectional study. Clinical Microbiology and Infection https://doi.org/10.1016/j.cmi.2020.05.001.

25. Iwasaki S, Fujisawa S, Nakakubo S, Kamada K, Yamashita Y, Fukumoto T, Sato K, Oguri S, Taki K, Senjo H, Sugita J, Hayasaka K, Konno S, Nishida M, Teshima T. 2020. Comparison of SARS-CoV-2 detection in nasopharyngeal swab and saliva. J Infect 81:e145–e147.

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

27. Jamal AJ, Mozafarihashjin M, Coomes E, Powis J, Li AX, Paterson A, Anceva-Sami S, Barati S, Crowl G, Faheem A, Farooqi L, Khan S, Prost K, Poutanen S, Taylor M, Yip L, Zhong XZ, McGeer AJ, Mubareka S, Investigators TIBDNC-19, Coleman BL, Chen D, Farshait N, Gold W, Kandel CE, Katz K, Kozak R, Mazzulli T, Muller M, Opavsky A, Ostrowski M, Plevneshi A, Rau N, Ricciuto D, Richardson D, Rose D, Sales V, Walmsley S. 2020. Sensitivity of Nasopharyngeal Swabs and Saliva for the Detection of Severe Acute Respiratory Syndrome Coronavirus 2. Clin Infect Dis https://doi.org/10.1093/cid/ciaa848.

28. Dogan OA, Kose B, Agaoglu NB, Yildiz J, Alkurt G, Demirkol YK, Irvem A, Dinler-Doganay G, Doganay L. 2020. Does sampling saliva increase detection of SARS-CoV-2 by

RT-PCR? Comparing saliva with oro-nasopharyngeal swabs. medRxiv https://doi.org/10.1101/2020.07.26.20158618.

29. Williams E, Bond K, Zhang B, Putland M, Williamson DA. 2020. Saliva as a Noninvasive Specimen for Detection of SARS-CoV-2. Journal of Clinical Microbiology https://doi.org/10.1128/JCM.00776-20.

30. Skolimowska K, Rayment M, Jones R, Madona P, Moore LSP, Randell P. 2020. Non-invasive saliva specimens for the diagnosis of COVID-19: caution in mild outpatient cohorts with low prevalence. Clinical Microbiology and Infection https://doi.org/10.1016/j.cmi.2020.07.015.

31. Miller M, Jansen M, Bisignano A, Mahoney S, Wechsberg C, Albanese N, Castillo L, Farinas P, Lazarin GA, Jaremko M. 2020. Validation of a Self-administrable, Saliva-based RT-qPCR Test Detecting SARS-CoV-2. medRxiv https://doi.org/10.1101/2020.06.05.20122721.

32. Griesemer SB, Van Slyke G, Ehrbar D, Strle K, Yildirim T, Centurioni DA, Walsh AC, Chang AK, Waxman MJ, St. George K. 2020. Evaluation of specimen types and saliva stabilization solutions for SARS-CoV-2 testing. medRxiv https://doi.org/10.1101/2020.06.16.20133041.

33. Rao M, Rashid FA, Sabri FSAH, Jamil NN, Zain R, Hashim R, Amran F, Kok HT, Samad MAA, Ahmad N. 2020. Comparing nasopharyngeal swab and early morning saliva for the identification of SARS-CoV-2. Clinical Infectious Diseases https://doi.org/10.1093/cid/ciaa1156.

638 34. Zheng S, Yu F, Fan J, Zou Q, Xie G, Yang X. Saliva as a Diagnostic Specimen for SARS-
639 CoV-2 by a PCR-Based Assay: A Diagnostic Validity Study. SSRN
640 https://doi.org/10.2139/ssrn.3543605.

641 35. Hanson KE, Barker AP, Hillyard DR, Gilmore N, Barrett JW, Orlandi RR, Shakir SM.
642 2020. Self-Collected Anterior Nasal and Saliva Specimens versus Healthcare Worker-
643 Collected Nasopharyngeal Swabs for the Molecular Detection of SARS-CoV-2. Journal of
644 Clinical Microbiology https://doi.org/10.1128/JCM.01824-20.

645 36. Migueres M, Mengelle C, Dimeglio C, Didier A, Alvarez M, Delobel P, Mansuy J-M,
646 Izopet J. 2020. Saliva sampling for diagnosing SARS-CoV-2 infections in symptomatic
647 patients and asymptomatic carriers. J Clin Virol 130:104580.

648 37. Otto MP, Darles C, Valero E, Benner P, Dutasta F, Janvier F. 2020. Posterior
649 oropharyngeal salivafor the detection of SARS-CoV-2. Clin Infect Dis
650 https://doi.org/10.1093/cid/ciaa1181.

651 38. COVISAL Guyane, Nacher M, Demar M. 2020. Prospective comparison of saliva and
652 nasopharyngeal swab sampling for mass screening for COVID-19. medRxiv
653 https://doi.org/10.1101/2020.09.23.20150961.

654 39. Byrne RL, Kay GA, Kontogianni K, Brown L, Collins AM, Cuevas LE, Ferreira D, Fraser
655 AJ, Garrod G, Hill H, Menzies S, Mitsi E, Owen SI, Williams CT, Hyder-Wright A, Adams
656 ER, Cubas-Atienzar AI. 2020. Saliva offers a sensitive, specific and non-invasive
657 alternative to upper respiratory swabs for SARS-CoV-2 diagnosis. medRxiv
658 https://doi.org/10.1101/2020.07.09.20149534.

40. Chen JH-K, Yip CC-Y, Poon RW-S, Chan K-H, Cheng VC-C, Hung IF-N, Chan JF-W, Yuen K-Y, To KK-W. 2020. Evaluating the use of posterior oropharyngeal saliva in a point-of-care assay for the detection of SARS-CoV-2. Emerging Microbes & Infections 9:1356–1359.

41. To KK-W, Tsang OT-Y, Yip CC-Y, Chan K-H, Wu T-C, Chan JM-C, Leung W-S, Chik TS-H, Choi CY-C, Kandamby DH, Lung DC, Tam AR, Poon RW-S, Fung AY-F, Hung IF-N, Cheng VC-C, Chan JF-W, Yuen K-Y. 2020. Consistent Detection of 2019 Novel Coronavirus in Saliva. Clin Infect Dis 71:841–843.

42. To KK-W, Tsang OT-Y, Leung W-S, Tam AR, Wu T-C, Lung DC, Yip CC-Y, Cai J-P, Chan JM-C, Chik TS-H, Lau DP-L, Choi CY-C, Chen L-L, Chan W-M, Chan K-H, Ip JD, Ng AC-K, Poon RW-S, Luo C-T, Cheng VC-C, Chan JF-W, Hung IF-N, Chen Z, Chen H, Yuen K-Y. 2020. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. The Lancet Infectious Diseases 20:565–574.

43. Leung EC, Chow VC, Lee MK, Lai RW. Deep throat saliva as an alternative diagnostic specimen type for the detection of SARS-CoV-2. Journal of Medical Virology https://doi.org/10.1002/jmv.26258.

44. Nagura-Ikeda M, Imai K, Tabata S, Miyoshi K, Murahara N, Mizuno T, Horiuchi M, Kato K, Imoto Y, Iwata M, Mimura S, Ito T, Tamura K, Kato Y. 2020. Clinical Evaluation of Self-Collected Saliva by Quantitative Reverse Transcription-PCR (RT-qPCR), Direct RT-qPCR, Reverse Transcription–Loop-Mediated Isothermal Amplification, and a Rapid

Journal of Clinical Microbiology

680 Antigen Test To Diagnose COVID-19. Journal of Clinical Microbiology
681 https://doi.org/10.1128/JCM.01438-20.

682 45. Kim SE, Lee JY, Lee A, Kim S, Park KH, Jung SI, Kang SJ, Oh TH, Kim UJ, Lee SY, Kee
683 SJ, Jang HC. 2020. Viral Load Kinetics of SARS-CoV-2 Infection in Saliva in Korean
684 Patients: a Prospective Multi-center Comparative Study. J Korean Med Sci 35:e287.

685 46. Azzi L, Carcano G, Gianfagna F, Grossi P, Gasperina DD, Genoni A, Fasano M, Sessa F,
686 Tettamanti L, Carinci F, Maurino V, Rossi A, Tagliabue A, Baj A. 2020. Saliva is a reliable
687 tool to detect SARS-CoV-2. Journal of Infection 81:e45–e50.

688 47. Fang Z, Zhang Y, Hang C, Ai J, Li S, Zhang W. 2020. Comparisons of viral shedding time
689 of SARS-CoV-2 of different samples in ICU and non-ICU patients. J Infect 81:147–178.

690 48. Vogels CBF, Brackney D, Wang J, Kalinich CC, Ott I, Kudo E, Lu P, Venkataraman A,
691 Tokuyama M, Moore AJ, Muenker MC, Casanovas-Massana A, Fournier J, Bermejo S,
692 Campbell M, Datta R, Nelson A, Yale IMPACT Research Team, Dela Cruz C, Farhadian S,
693 Ko A, Iwasaki A, Hui P, Liu C, Wyllie AL, Grubaugh N. 2020. SalivaDirect: Simple and
694 sensitive molecular diagnostic test for SARS-CoV-2 surveillance. medRxiv
695 https://doi.org/10.1101/2020.08.03.20167791.

696 49. Wyllie AL, Fournier J, Casanovas-Massana A, Campbell M, Tokuyama M, Vijayakumar P,
697 Warren JL, Geng B, Muenker MC, Moore AJ, Vogels CBF, Petrone ME, Ott IM, Lu P,
698 Venkataraman A, Lu-Culligan A, Klein J, Earnest R, Simonov M, Datta R, Handoko R,
699 Naushad N, Sewanan LR, Valdez J, White EB, Lapidus S, Kalinich CC, Jiang X, Kim DJ,
700 Kudo E, Linehan M, Mao T, Moriyama M, Oh JE, Park A, Silva J, Song E, Takahashi T,

Taura M, Weizman O-E, Wong P, Yang Y, Bermejo S, Odio CD, Omer SB, Dela Cruz CS, Farhadian S, Martinello RA, Iwasaki A, Grubaugh ND, Ko AI. 2020. Saliva or Nasopharyngeal Swab Specimens for Detection of SARS-CoV-2. New England Journal of Medicine 383:1283–1286.

50. Teo AKJ, Choudhury Y, Tan IB, Cher CY, Chew SH, Wan ZY, Cheng LTE, Oon LLE, Tan MH, Chan KS, Hsu LY. 2020. Validation of Saliva and Self-Administered Nasal Swabs for COVID-19 Testing. medRxiv https://doi.org/10.1101/2020.08.13.20173807.

51. Han MS, Seong M-W, Kim N, Shin S, Cho SI, Park H, Kim TS, Park SS, Choi EH. Early Release - Viral RNA Load in Mildly Symptomatic and Asymptomatic Children with COVID-19, Seoul - Volume 26, Number 10—October 2020 - Emerging Infectious Diseases journal - CDC. Emerging Infectious Diseases Journal https://doi.org/10.3201/eid2610.202449.

52. Arnaout R, Lee RA, Lee GR, Callahan C, Yen CF, Smith KP, Arora R, Kirby JE. 2020. SARS-CoV2 Testing: The Limit of Detection Matters. bioRxiv https://doi.org/10.1101/2020.06.02.131144.

53. Chong CY, Kam K-Q, Li J, Maiwald M, Loo LH, Nadua KD, Tan NWH, Yung CF, Thoon KC. 2020. Saliva is not a useful diagnostic specimen in children with Coronavirus Disease 2019. Clinical Infectious Diseases https://doi.org/10.1093/cid/ciaa1376.

54. Berenger BM, Fonseca K, Schneider AR, Hu J, Zelyas N. 2020. Sensitivity of Nasopharyngeal, Nasal and Throat Swab for the Detection of SARS-CoV-2. medRxiv https://doi.org/10.1101/2020.05.05.20084889.

722 55. Wang X, Tan L, Wang X, Liu W, Lu Y, Cheng L, Sun Z. 2020. Comparison of
723 nasopharyngeal and oropharyngeal swabs for SARS-CoV-2 detection in 353 patients
724 received tests with both specimens simultaneously. Int J Infect Dis 94:107–109.

725 56. Yu C, Li L, Tuersun Y, Zhao X, Feng Q, Zhang T, Tay FR, Ma J. 2020. Oropharyngeal
726 Secretion as Alternative for SARS-CoV-2 Detection. J Dent Res 99:1199–1205.

727 57. Calame A, Mazza L, Renzoni A, Kaiser L, Schibler M. 2020. Sensitivity of
728 nasopharyngeal, oropharyngeal and nasal washes specimens for SARS-CoV-2 detection in
729 the setting of sampling device shortage. medRxiv
730 https://doi.org/10.1101/2020.08.01.20166397.

731 58. Patel MR, Carroll D, Ussery E, Whitham H, Elkins CA, Noble-Wang J, Rasheed JK, Lu X,
732 Lindstrom S, Bowen V, Waller J, Armstrong G, Gerber S, Brooks JT. Performance of
733 Oropharyngeal Swab Testing Compared With Nasopharyngeal Swab Testing for Diagnosis
734 of Coronavirus Disease 2019—United States, January 2020–February 2020. Clin Infect Dis
735 https://doi.org/10.1093/cid/ciaa759.

736 59. Wehrhahn MC, Robson J, Brown S, Bursle E, Byrne S, New D, Chong S, Newcombe JP,
737 Siversten T, Hadlow N. 2020. Self-collection: An appropriate alternative during the SARS-
738 CoV-2 pandemic. J Clin Virol 128:104417.

739 60. Abdollahi A, Shakoori A, Khoshnevis H, Arabzadeh M, Dehghan Manshadi SA,
740 Mohammadnejad E, Ghasemi D, Safari Aboksari M, Alizadeh S, Mehrtash V, Eftekhar-
741 Javadi A, Safaei M. 2020. Comparison of Patient-collected and Lab Technician-collected

742 Nasopharyngeal and Oropharyngeal Swabs for Detection of COVID-19 by RT-PCR. Iran J
743 Pathol 15:313–319.

744 61. Kam K, Yung CF, Maiwald M, Chong CY, Soong HY, Loo LH, Tan NWH, Li J, Nadua
745 KD, Thoon KC. 2020. Clinical Utility of Buccal Swabs for Severe Acute Respiratory
746 Syndrome Coronavirus 2 Detection in Coronavirus Disease 2019–Infected Children.
747 Journal of Pediatric Infectious Diseases Society 9:370–372.

748 62. Tu Y-P, Jennings R, Hart B, Cangelosi GA, Wood RC, Wehber K, Verma P, Vojta D,
749 Berke EM. 2020. Swabs Collected by Patients or Health Care Workers for SARS-CoV-2
750 Testing. New England Journal of Medicine 383:494–496.

751 63. Kojima N, Turner F, Slepnev V, Bacelar A, Deming L, Kodeboyina S, Klausner JD. 2020.
752 Self-Collected Oral Fluid and Nasal Swabs Demonstrate Comparable Sensitivity to
753 Clinician Collected Nasopharyngeal Swabs for Covid-19 Detection. medRxiv
754 https://doi.org/10.1101/2020.04.11.20062372.

755 64. Basu A, Zinger T, Inglima K, Woo K, Atie O, Yurasits L, See B, Aguero-Rosenfeld ME.
756 2020. Performance of Abbott ID Now COVID-19 Rapid Nucleic Acid Amplification Test
757 Using Nasopharyngeal Swabs Transported in Viral Transport Media and Dry Nasal Swabs
758 in a New York City Academic Institution. J Clin Microbiol
759 https://doi.org/10.1128/JCM.01136-20.

760 65. Harrington A, Cox B, Snowdon J, Bakst J, Ley E, Grajales P, Maggiore J, Kahn S. 2020.
761 Comparison of Abbott ID Now and Abbott m2000 Methods for the Detection of SARS-

762 CoV-2 from Nasopharyngeal and Nasal Swabs from Symptomatic Patients. J Clin
763 Microbiol https://doi.org/10.1128/JCM.00798-20.

764 66. Péré H, Podglajen I, Wack M, Flamarion E, Mirault T, Goudot G, Hauw-Berlemont C, Le
765 L, Caudron E, Carrabin S, Rodary J, Ribeyre T, Bélec L, Veyer D. 2020. Nasal Swab
766 Sampling for SARS-CoV-2: a Convenient Alternative in Times of Nasopharyngeal Swab
767 Shortage. Journal of Clinical Microbiology https://doi.org/10.1128/JCM.00721-20.

768 67. Pinninti S, Trieu C, Pati SK, Latting M, Cooper J, Seleme MC, Boppana S, Arora N, Britt
769 WJ, Boppana SB. 2020. Comparing Nasopharyngeal and Mid-Turbinate Nasal Swab
770 Testing for the Identification of SARS-CoV-2. Clin Infect Dis
771 https://doi.org/10.1093/cid/ciaa882.

772 68. Callahan C, Lee R, Lee G, Zulauf KE, Kirby JE, Arnaout R. 2020. Nasal-Swab Testing
773 Misses Patients with Low SARS-CoV-2 Viral Loads. medRxiv
774 https://doi.org/10.1101/2020.06.12.20128736.

775 69. Palmas G, Moriondo M, Trapani S, Ricci S, Calistri E, Pisano L, Perferi G, Galli L,
776 Venturini E, Indolfi G, Azzari C. 2020. Nasal Swab as Preferred Clinical Specimen for
777 COVID-19 Testing in Children. The Pediatric Infectious Disease Journal 39:e267.

778 70. Capecchi E, Di Pietro GM, Luconi E. 2020. Is Nasopharyngeal Swab Comparable With
779 Nasopharyngeal Aspirate to Detect SARS-CoV-2 in Children? The Pediatric Infectious
780 Disease Journal 39:e288.

781 71. Vlek ALM, Wesselius TS, Achterberg R, Thijsen SFT. 2020. Combined throat/nasal swab
782 sampling for SARS-CoV-2 is equivalent to nasopharyngeal sampling. Eur J Clin Microbiol
783 Infect Dis https://doi.org/10.1007/s10096-020-03972-y.

784 72. LeBlanc JJ, Heinstein C, MacDonald J, Pettipas J, Hatchette TF, Patriquin G. 2020. A
785 combined oropharyngeal/nares swab is a suitable alternative to nasopharyngeal swabs for
786 the detection of SARS-CoV-2. J Clin Virol https://doi.org/10.1016/j.jcv.2020.104442.

787 73. Desmet T, Paepe PD, Boelens J, Coorevits L, Padalko E, Vandendriessche S, Leroux-Roels
788 I, Aerssens A, Callens S, Braeckel V, Malfait T, Vermassen F, Verhasselt B. Combined
789 oropharyngeal/nasal swab is equivalent to nasopharyngeal sampling for SARS-CoV-2
790 diagnostic PCR. medRxiv https://doi.org/10.1101/2020.06.05.20123745.

791 74. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, Lijmer JG, Moher
792 D, Rennie D, de Vet HCW, Kressel HY, Rifai N, Golub RM, Altman DG, Hooft L,
793 Korevaar DA, Cohen JF, STARD Group. 2015. STARD 2015: an updated list of essential
794 items for reporting diagnostic accuracy studies. BMJ 351:h5527.

795 75. Lindner AK, Nikolai O, Kausch F, Wintel M, Hommes F, Gertler M, Krüger LJ, Gaeddert
796 M, Tobian F, Lainati F, Köppel L, Seybold J, Corman VM, Drosten C, Hofmann J, Sacks
797 JA, Mockenhaupt FP, Denkinger CM. 2020. Head-to-head comparison of SARS-CoV-2
798 antigen-detecting rapid test with self-collected anterior nasal swab versus professional-
799 collected nasopharyngeal swab. medRxiv https://doi.org/10.1101/2020.10.26.20219600.

800
801
802

803 Figure 1: Summary forest plot of individual studies assessing saliva

804 Figure 2: Summary forest plot of sub-group data from saliva sampling for different clinical populations,
805 and collection as well as processing procedures.

806 Fig. 3: Summary forest plot of individual studies assessing oropharyngeal swabs

807 Fig. 4: Summary forest plot of individual studies assessing nasal swabs

808 Fig. 5: Summary forest plot of sub-group data from nasal swab sampling for different clinical
809 populations, and collection as well as processing procedures.

810 Fig. 6: Summary forest plot of individual studies assessing oropharyngeal/nasal swab sampling

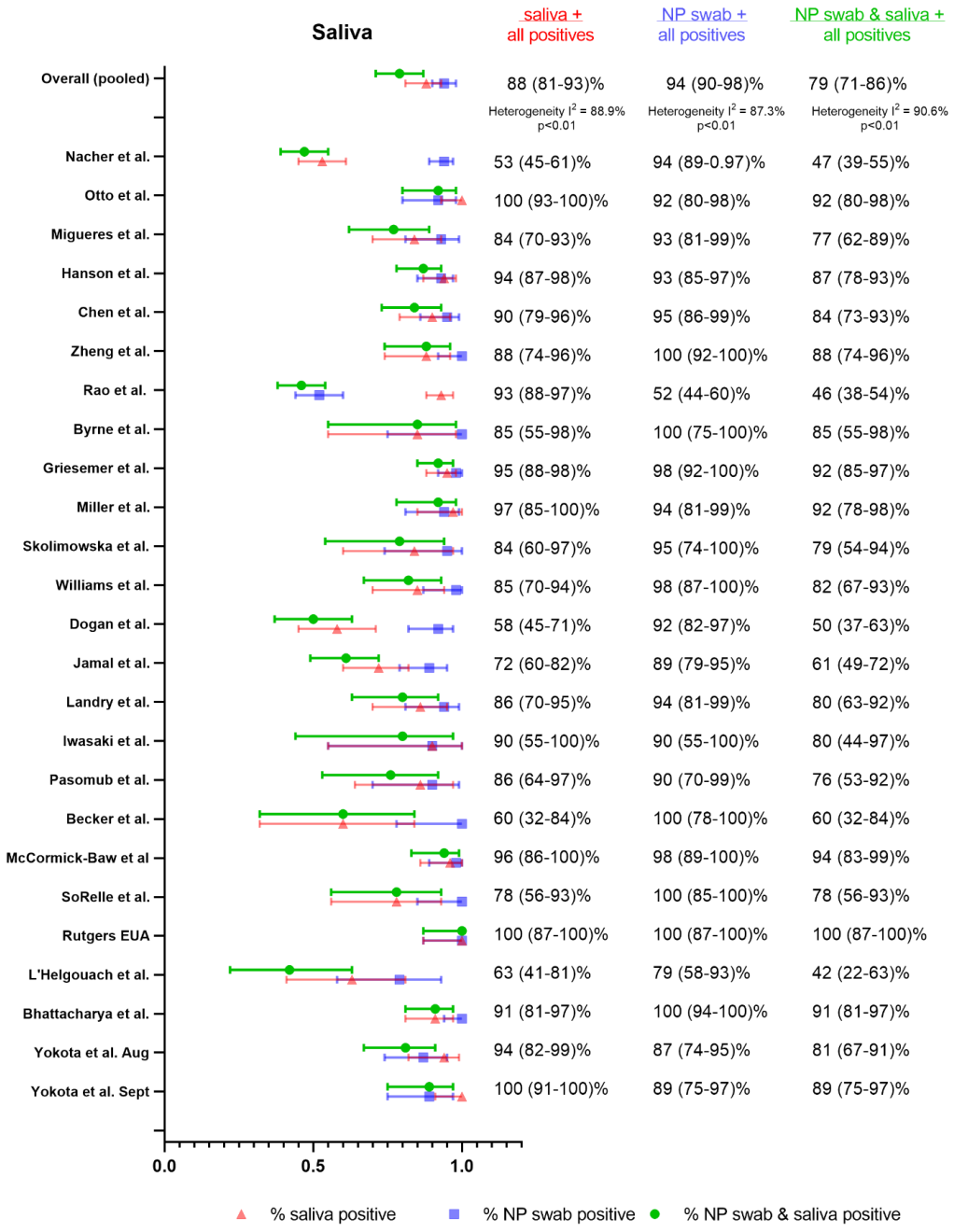
811

812

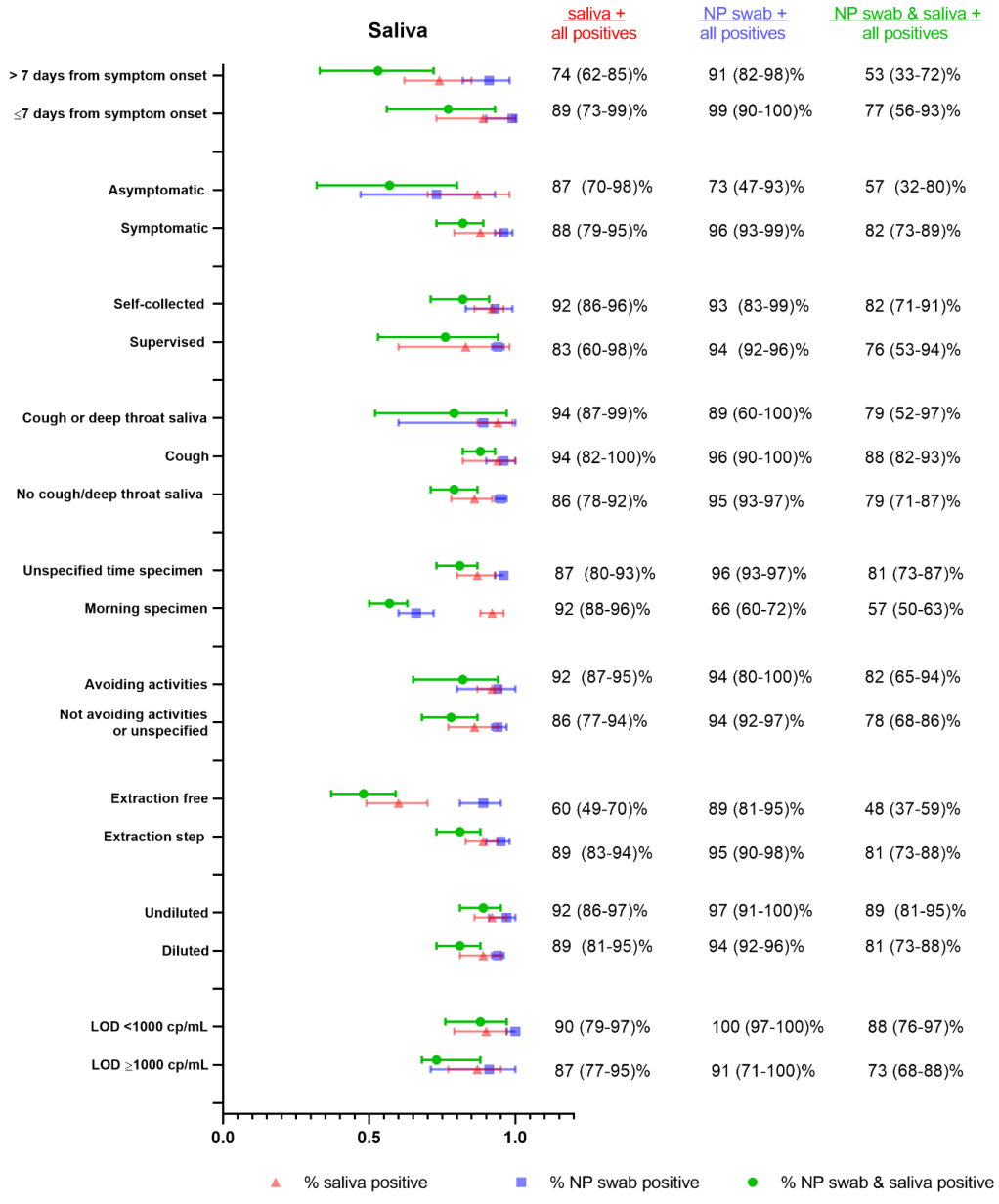
813

814

815



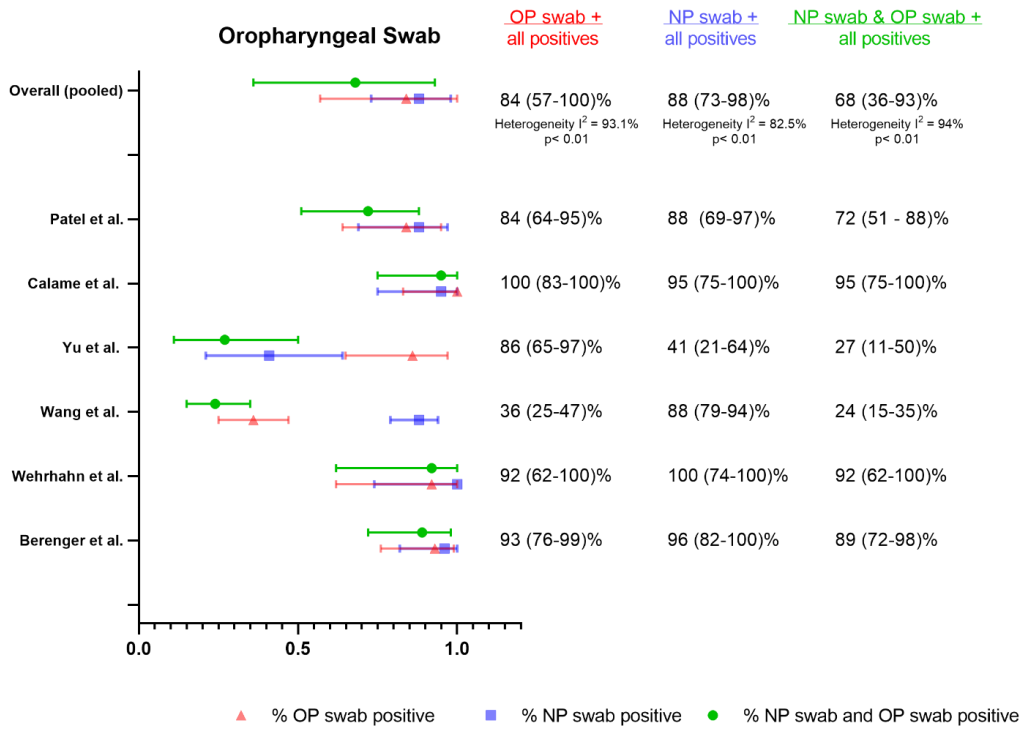
816

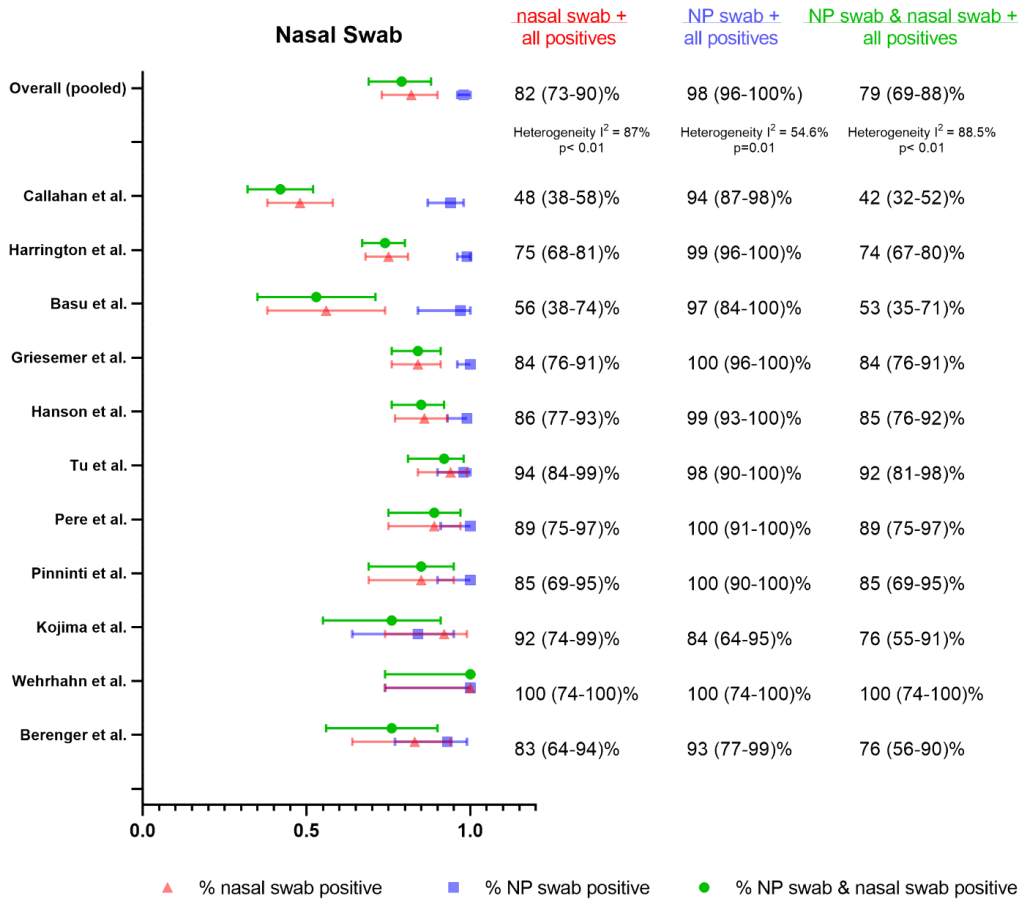


817

818

819





833

834

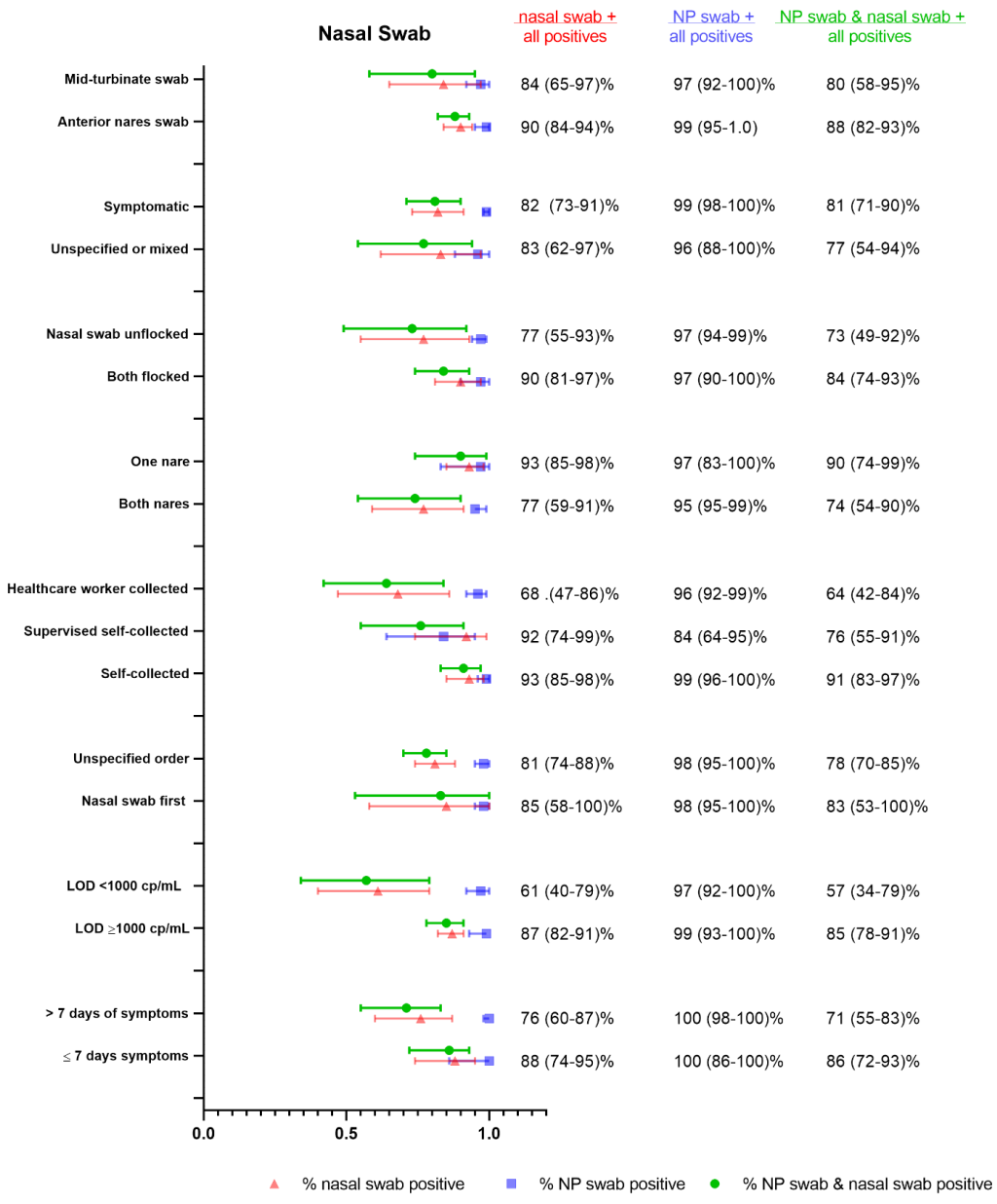
835

836

837

838

839



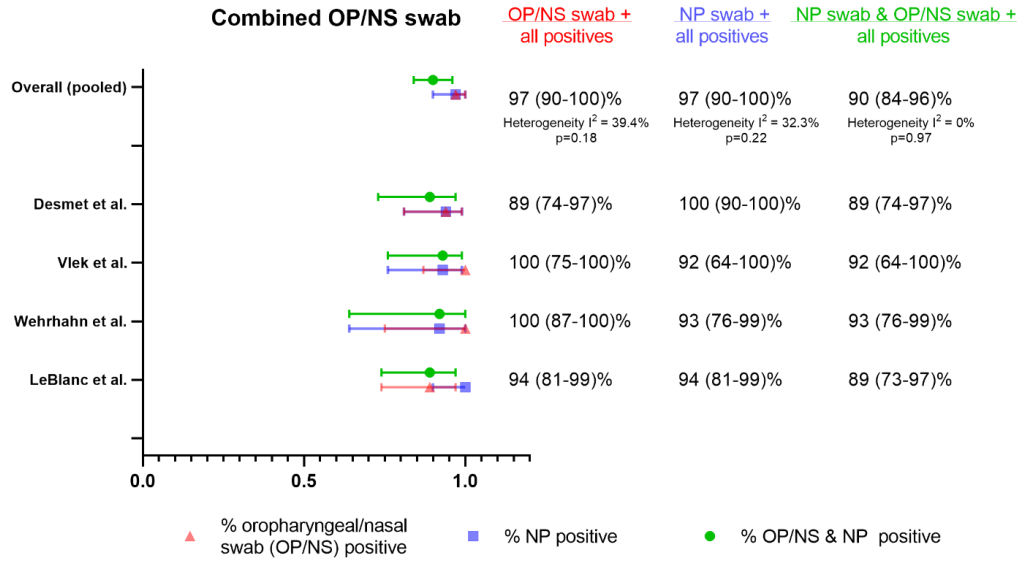
840

841

842

843

844



845

846

847

848

849

850