Use of Pooled Testing for the Diagnosis, Screening and Surveillance of COVID-19

Rapid Review

Prepared by:
Health Technology Assessment Unit
Policy Planning and Evaluation Team

Use of Pooled Testing for the Diagnosis, Screening and Surveillance of COVID-19

Rapid Review

Service Line	Rapid Review
Version	4
Version date	06 November 2020
Report Length	152 Pages
Prepared by	Deinzel R. Uezono Anne Julienne M. Genuino, RPh, MSc Sabrina Grace T. Aguinaldo, RPh Lara Alyssa B. Liban, RPh April Dominique M. Ocampo Jeanne Pauline G. Panopio, RPh Mark Joshua Nathaniel A. Ramos Patrick Wincy C. Reyes

This page is intentionally left blank.

Table of Contents

List of Tables

List of Figures

1. CONTEXT AND POLICY ISSUES

In early 2020, the World Health Organization (WHO) declared severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causing novel coronavirus disease 2019 (COVID-19) as a global pandemic affecting more than 189 countries and regions. There are at least 38, 925, 204 cases and 1, 098, 378 deaths worldwide as of 16 October, 2020 (Johns Hopkins Coronavirus Resource Center, 2020). In the Philippines, COVID-19 has affected over 351, 750 cases with 6, 531 deaths as of 16 October, 2020 (DOH, 2020). To date, treatment remains unknown. (Dong, Du & Gardner, 2020; DOH, 2020)

Currently, real-time reverse-transcription polymerase chain reaction (RT-PCR) assay is the recommended test to confirm COVID-19 infection. It is a qualitative detection of nucleic acid from SARS-CoV-2 in nasopharyngeal or oropharyngeal samples, sputum, bronchoalveolar lavage fluid, and other bodily fluids, including feces and blood collected from suspected cases (Philippine Society for Microbiology and Infectious Disease, 2020). The viral genes targeted so far include the N, E, S, ORF1ab and RdRP gene. Positive results indicate the presence of SARS-CoV-2 RNA and must be clinically correlated with patient history as well as other diagnostic information to determine patient infection status. (World Health Organization, 2020)

Last May 2020, the Health Technology Assessment Unit (HTAU) published a rapid review on SARS-COV-2 RT PCR Test Kits for the diagnosis of COVID-19 which served as evidentiary basis for the recommendation of the Health Technology Assessment Council (HTAC). Recognizing that the performance of the test is varied across brands, the HTAC has recommended a set of minimum specifications and technical requirements to guide the DOH and its accredited laboratories in making available RT-PCR test kits and in ensuring consistency and reliability of test results during the COVID-19 pandemic. (Guidance Document on the Technical Requirements for SARS-CoV-2 RT-PCR Test Kits Used in the Diagnosis of COVID-19 Cases). Currently, RT-PCR remains as the gold standard for diagnosing COVID-19. (DOH DM 2020-0258)

As the number of COVID-19 cases increase, availability of diagnostic kits and reagents emerged as a major bottleneck in the laboratory testing of SARS-CoV-2 (Khodare, 2020). Current strategies involve testing only symptomatic individuals. However, evolving strategies for testing worldwide now also include testing asymptomatic individuals with pertinent contact history to curb the spread of infection in the community (Praharaj, 2020). Large scale population screening for COVID-19 infection is generally considered a necessary part of an exit strategy from the coronavirus lockdown (Wacharapluesadee, 2020).

In response to the growing demand for testing capacity, the recently issued DOH Department Memorandum (DM) 2020-0439 *Omnibus Testing Guidelines on Prevention, Detection, Isolation, Treatment and Reintegration Strategies for COVID-19* has included pooled testing among the currently recommended testing options. The issuance, however, stated that pooled testing strategies are currently being evaluated and validated; hence, these guidelines shall be further amended as new developments ensue from new studies and pilot implementation. While the guidelines mentioned that pooled testing may be used for surveillance testing of asymptomatic workers, it emphasized that such methodology may only be used once results of on-going pilot testing are positive and favorable, based on the recommendations of experts.

In light of this, the Philippine Department of Health (DOH) sought the expertise of the HTAC in appraising the use of pooled testing for different use cases in testing for COVID-19. Pooling the diagnostic tests has been applied in other infectious diseases and is especially attractive as it requires

no additional training, equipment, or materials (Gupta, 2020). The most common pool testing approach involves combining aliquots of individual specimens into one pool and testing the pool. If the pool is negative, all results within the pool are considered negative. If the pool is positive, individual specimens from that pool are then tested to determine which specimen(s) are positive (Bateman, 2020)

As such, this rapid review was conducted to search, appraise and synthesize currently existing synthesized relevant information and evidence on the regulatory approval and validation requirements, performance characteristics, existing recommendatory testing guidelines and assessment findings, as well as resource requirements on implementing pooled testing for the following use cases in COVID-19: (1) Diagnosis; (2) Screening; and (3) Surveillance. Following definitions set by the DOH DM 2020-0439, these use cases are defined as follows:

Table 1. Definitions of use cases

Use Case	Definition
Diagnosis	Diagnostic testing/ testing for diagnosis looks for presence of COVID-19 at the individual level and is performed when there is a particular reason to suspect that an individual may be infected (i.e. manifestation of symptoms or known history of exposure). Diagnostic testing intends to diagnose an infection in patients suspected for COVID-19 by their healthcare provider, such as in symptomatic individuals, individuals who have had recent exposure, and individuals who are in high-risk group such as healthcare providers with known exposure. In these guidelines, this shall be applied to close contacts and suspect cases identified after symptoms-based screening
Screening	Screening testing/ testing for screening intends to identify infected individuals prior to development of symptoms or those infected individuals without signs of symptoms who may be contagious, so measures can be taken to prevent them from infecting others. This includes broad screening of asymptomatic individuals without known exposure and then deciding on the next courses of action based on individual test results. In these guidelines, this shall be applied to travelers from high prevalence areas.
Surveillance	Testing for surveillance is primarily used to obtain information at a population level, rather than an individual level. Surveillance testing may be random sampling of a certain percentage of a specific population, to (1) monitor for increasing or decreasing prevalence, and (2) determine the effects of community interventions such as physical distancing at the population level. In these guidelines, these shall be applied to front liners and essential workers.

2. POLICY AND RESEARCH QUESTIONS

POLICY QUESTION

What use case and for which population should the Philippine DOH consider the use of pooled testing for COVID-19?

RESEARCH QUESTIONS

1. **Regulatory Approval**

   What are the approved uses of pooled testing by regulatory agencies in other countries?

   What are the validation testing requirements of pooled testing of regulatory agencies in other countries?

What are the performance standards of pooled testing by regulatory agencies in other countries?

2. **Performance Characteristics:**

   What is the accuracy of pooled testing in the diagnosis, screening, and surveillance of COVID-19 as compared to unpooled RT-PCR?

3. **Global guidelines and position on use of pooled testing**

   Which countries have implemented testing strategies using pooled testing for diagnosis, screening, and surveillance of COVID-19?

   What is the current position of HTA agencies regarding the use of pooled testing for diagnosis, screening, and surveillance of COVID-19?

4. **Resource requirements**

   What are the resource requirements needed to implement pooled testing?

3. KEY FINDINGS

REGULATORY STANDARDS

Eleven regulatory agencies were searched for guidelines on the regulation and authorization of the use of pooled testing. Only the United States Food and Drug Administration (US FDA) has established guidelines on pooled testing and on how to validate the test when requesting for an Emergency Use Authorization (EUA). In the EUA diagnostic template, patients suspected of COVID-19 infection and individuals without symptoms or other reasons to suspect the said infection are the targeted population for pooled testing. The US FDA believes that the use of pooling in certain SARS-CoV-2 tests (i.e., screening and diagnostic tests) can be an option and be authorized granting that there exist proper mitigations and validations, thereby, ensuring its proper implementation. Furthermore, on the use of SARS-CoV-2 diagnostic test for surveillance purposes (e.g., determining the prevalence of acute infections in a population), the US FDA explicitly stated that it does not generally regulate the use of a test for this purpose or use case. Nonetheless, it was mentioned by US FDA that if surveillance testing is performed by a non-Clinical Laboratory Improvement Amendments (CLIA) certified laboratory, a confirmatory test on the detected positive individuals should be performed by a CLIA-certified laboratory.

As for the validation testing requirements, if commercial test kit manufacturers would like to include sample pooling and swab pooling to their authorized uses, they must submit an EUA request together with the data from their clinical validation studies with the following guidelines:

- The reference standard used is the same EUA-authorized assay RT-PCR kit for which the added indication is being requested for.
- If the RT-PCR has not been previously granted an EUA, the reference standard should be a comparator assay that has an established high sensitivity with an internationally recognized standard or the FDA SARS-CoV-2 reference panel.
- The index test must have at least 20 positive individual samples for an EUA RT-PCR kit and at least 30 positive individual samples for a new RT-PCR kit. Each positive n-sample pool shall have 1 positive sample and n-1 negative samples. An equal number of negative pools must also be tested.

- The pool size should be chosen in the context of the positivity rate and percent of weak positives in the intended test population and the sensitivity of the test. However, the US FDA recommends to test developers that they begin validating their test kits using a pool size of 5.

The US FDA also requires two additional studies for swab pooling to evaluate the inhibition observed when (1) the sample has high concentrations of swab specimen (e.g., mucin) and (2) when the sample has a high concentration of viral load.

The table below summarizes the performance standards set by the US FDA for the validation studies required when requesting an EUA for pooled testing:

Performance standard domain	Sample/Media Pooling	Swab Pooling
Performance for pooled testing vs individual testing	• ≥85% PPA • Individual weak positive samples with high cycle threshold (Ct) values (viral loads close to LoD of the assay) can be accurately detected when pooled with n-1 negative samples	
Interference of swab specimens	N/A	• 95% agreement with expected results and • <5% invalid rate
Interference of viral load	N/A	• 100% positivity rate or • ≤5% invalid rate

Of the ten RT-PCR kits that the US FDA has granted the authorization to be used for pooled testing, only one brand, cobas SARS-CoV-2, is approved by the Philippine Food and Drug Administration (FDA). However, this is only used for individual testing in the Philippines as regulatory standards for pooled testing in the country do not exist yet.

DIAGNOSTIC PERFORMANCE

- A total of 20 studies were included in the qualitative synthesis. Nineteen studies were primary diagnostic accuracy studies and one was a clinical trial protocol. The 18 primary studies were included for quantitative synthesis. Eight studies looked at pooled testing for screening, two studies explored pooled testing for surveillance, five were for diagnosis and three did not mention a particular use case. Majority of the studies did not describe the study population used and were laboratory-based simulations rather than field validation. However, one data set from one study mentioned use of pooled testing among asymptomatic healthcare personnel, employees of essential industries, and residents and employees of nursing homes while one study mentioned use of pooled testing among asymptomatic residents of a nursing home.
- The sensitivity of pooled testing greatly varies when the unit of analysis is by pool (sensitivity point estimates ranging from 25 –100%) wherein the number of pools tested is small, compared to when the analysis is by individual (sensitivity point estimates ranging from 92 –100%). The overall pooled sensitivity of pooled testing analyzed by pool was found to be 87%, (95% CI: 81-91, I2=71%) while pooled sensitivity analyzed by individual was 97% (95% CI: 95-99, I2=0%). The overall pooled sensitivity analyzed by individual suggests that pooled testing has a good rate of correctly identifying COVID-19 positive individuals and consequently, a low rate of false negatives. Meanwhile, the pooled sensitivity analyzed by pools should be interpreted with caution due to high heterogeneity.
- On the other hand, the specificity of pooled testing was consistently high ranging from 97% to 100%. The overall pooled specificity of pooled testing analyzed by pool [98.9% (95% CI: 89-96,

I^2=12%)] and analyzed by individual [99.99%, CI: 98.9-100, I^2=0%)] implies that the test has a very low rate of both false positive pools and false positive individuals.

- Given the substantial heterogeneity present in the pooling of sensitivity, we performed subgroup analysis to assess the impact of several factors (e.g., pool size, brand of index test, CT value threshold, presence of symptoms, onset of symptoms, use case, specimen, and prevalence of disease) that may have served as sources of heterogeneity across the studies. The results from the subgroup analysis suggest that:
    - Based on low to moderate quality of evidence from 18 studies, pooled testing showed higher sensitivity estimates when pool sizes used are smaller. This finding was very intuitive because the more specimens in the same pool, the more diluted the shared reagent becomes losing the power of the test.
    - Based on low to moderate quality of evidence from 18 studies, the sensitivity varies greatly from one brand to another possibly because of the different processes and protocols each one takes, as well as the different criteria set by their respective manufacturers.
    - Based on low and moderate quality evidence from 2 studies, an increase in the CT value of positive samples in a pool decreases the sensitivity estimate. However, this needs further investigation since the number of studies analyzed for this variable is small.
    - The sensitivity of pooled testing for screening (6 studies, low to moderate validity) was the highest but was closely followed by those that indicated diagnosis (4 studies, low to moderate validity) as its use case. Though, both use cases were found to have substantial variation within the included studies.
    - Sensitivity estimates for symptomatic individuals, although based only on one moderate quality study, were higher compared to the asymptomatic (2 studies, low and moderate validity) and unspecified population (13 studies, low to moderate validity). Caution must also be taken in the interpretation of this analysis since very few studies had available information regarding the clinical characteristics of the sample.
    - Pooled testing that used saliva had the highest sensitivity but was comparable to those that collected test material using nasopharyngeal specimen. However, it should be noted that only one moderate quality study used saliva as its specimen. Those that used mixed samples of nasal and oral specimen also had an acceptable sensitivity.
    - Few studies indicated the prevalence of the disease where the sample was obtained so our numerical results for studies with low prevalence still warrant further investigation.
- The Philippine study by Lo et al (2020) had a lower sensitivity in both units of analysis than the overall pooled sensitivity estimate of our quantitative synthesis. The reported sensitivity estimates of the study for analysis by pools were 83% (95% CI: 67-94), 72% (95% CI: 55-86), and 67% (95% CIL 49-81) while the sensitivity estimate for analysis by individuals were reported to be at 83% (95% CI: 52-98), 58% (95% CI: 28-85), and 50% (95% CI: 21-79). In contrast, the specificity of the local study with individuals as unit of analysis (100%, 95% CI: 99-100) is comparable to the overall specificity estimate of the included studies in the meta-analysis. Furthermore, the study of Lo et al showed consistent trend with the observations in one subgroup analysis which indicates that lower pool sizes were seen to have higher sensitivity estimates.
- As for the quality of these studies on diagnostic performance, our critical appraisal shows that eight studies had high risk of bias and eleven had moderate risk of bias. Some factors that have affected the validity of these studies include non-independence of the definition, performance, and interpretation of the index and reference test.
- Based on low to moderate quality evidence, the use of pooled testing for COVID-19 shows high specificity but varied sensitivity. Further, the prevalence of the population to which pooled testing can be applied remains unclear.

GUIDELINE RECOMMENDATIONS

Of the fourteen guidelines reviewed, only four (United States Centers for Disease Control (US CDC), European Centers for Disease Control (ECDC), Public Health Ontario (PHO), and Philippine Department of Health) have existing guidelines on pooled testing for COVID-19. US CDC recommendations had the widest scope of use which is for diagnosis, screening, and surveillance. Both the ECDC and the Philippine DOH recommend its use only for screening or surveillance. We also note that two guidelines explicitly stated that they do not recommend pooled testing for diagnosis. Public Health Ontario (PHO) recommends its use for diagnosis and surveillance.

To further retrieve information on the use of pooled testing among different countries specifically on pilot implementation efforts or plans which are not usually reported in guidelines, the reviewers conducted a targeted search for news articles from different ministries of health websites as well as independent news agencies. The United Kingdom, Singapore, Malaysia, Korea, and Vietnam have mentioned the use of pooled testing as screening tests in different settings and populations. Indonesia, Thailand, and China have mentioned the use of pooled testing for surveillance under different circumstances.

Diagnosis

Two guidelines (US CDC and PHO) currently recommend the use of pooled testing for diagnosis of COVID-19 infection.

- As a diagnostic tool, the US CDC currently recommends pooled testing among patients with symptoms or recent exposure or to determine the resolution of infection.
- PHO uses pooled testing for a portion of specimens submitted to its laboratory from assessment centers. PHO did not further elaborate on the target population nor on the characteristics of population being tested for diagnosis
- US CDC and PHO both recommend a two-stage specimen pooling strategy in which samples are pooled together, and if a pooled test result is negative, then all specimens within the pool can be presumed negative with the single test; whereas if a pooled test result is positive or indeterminate, all the specimens within the pool need to be retested individually.

ECDC and the Philippine DOH explicitly stated in their guidelines that they do not recommend the use of pooled testing for diagnosis. The Philippine DOH does not recommend it for certain populations which include individuals with symptoms (regardless of severity), recovered patients, and close contacts of positive individuals.On the other hand, ECDC does not recommend it due to the possibility of error.

Screening

Three guidelines (US CDC, ECDC, and Philippine DOH) currently recommend the use of pooled testing in screening infected individuals without, or have not yet developed symptoms who may be contagious so that measures can be taken to prevent further transmission.

- The US CDC does not specify the target population of this use case.
- The ECDC specifically recommends it for mild and asymptomatic patients.
- The Philippine DOH recommends the use of pooled testing in screening the following populations: inbound travellers, overseas Filipino workers (deployment and returning), and locally stranded individuals.

In addition to testing guidelines, several news articles from countries like the United Kingdom, Singapore, Malaysia, Korea, and Vietnam have mentioned the use of pooled testing as screening tests in different settings and populations.

- The United Kingdom is introducing the use of COVID-19 screening using pooled testing in universities to help prevent outbreaks and allow campuses to stay open (Mahase, 2020).

- Singapore on the other hand, allows the use of pooled testing in migrant worker dormitories and nursing facilities, with their ministry of health recommending its use in testing sub-populations with very low prevalence rates of COVID-19, or for mass screening purposes. (Singapore MOH, 2020a) (Singapore MOH, 2020b) (Sin.Y., 2020) (Sun.D., 2020)
- Malaysia recommends the use of pooled testing in mass testing groups with high risk of infection such as members of the Kuching church where the biggest clusters occurred due to mass gatherings. (Choong.J., 2020)
- Korea also recommends using pooled testing in local clusters at a higher risk of acquiring COVID-19 infection using pool sizes of 10 (Sung-sun, 2020)
- Vietnam uses pooled testing on returnees from Da Nang population, which is considered to be the outbreak epicentre in Vietnam by using pool size of three to five individuals per laboratory test. (Kiet, 2020) (WHO, 2020)

Surveillance

Across the three use cases, pooled testing is most commonly used for surveillance purposes, based on the guidelines reviewed. All four guidelines (US CDC, ECDC, PHO, and Philippine DOH) recommend the use of pooled testing for surveillance.

- The US CDC recommends the use of pooled testing to monitor a community- or population-level occurrence, such as an infectious disease outbreak or to characterize the occurrence once detected, as well as to look at incidence and prevalence of occurrence.
- The ECDC recommends the use of pooled testing in determining prevalence of disease in the community or to enhance testing of mild and asymptomatic patients;
- PHO recommends it in testing asymptomatic patients especially during outbreak investigations.
- The Philippine DOH recommends the use of pooled testing in the surveillance of the following populations: healthcare workers and all workers in health facilities, essential workers including market vendors, transport workers, frontline government workers, and other economy workers.

In addition to testing guidelines, several news articles from countries like Indonesia, Thailand, and China have mentioned the use of pooled testing for surveillance under different circumstances.

- Indonesia plans to conduct pooled testing with pool sizes of five each in eight provinces that have been hardest hit by the coronavirus. A thousand samples will be taken from these provinces using a multi-step random sampling. (Sutrisno.B, 2020)
- Thailand also employs pooled testing using saliva samples to accelerate testing 100,000 persons in targeted groups including health and medical professionals, prison inmates, drivers for public buses, and migrant workers. (WHO, 2020)
- China used pooled testing on the entirety of Wuhan population as mass, indiscriminate testing, using pool sizes of five to ten individuals in one laboratory test (BBC, 2020)

We note that of the guidelines reviewed, the WHO and Australia do not have any guidelines regarding pooled testing, nor are there any articles which cite the use of this strategy.

HTA REVIEW RECOMMENDATIONS

We searched for existing reviews from 10 HTA agencies. Of these, there was no existing review with recommendations specific to pooled testing. Moreover, none of the HTA agencies included in our search have published any information regarding any ongoing studies on the use of pooled testing in COVID-19.

RESOURCE REQUIREMENTS

Originally, we had planned to look at the resources needed to implement pooled testing. However, among the nineteen studies that were reviewed, nine studies included only information on the resources saved when using their proposed pooling strategy at a certain prevalence and positivity rate. Three additional studies were also obtained to give us more information on the reduction in resource requirements when pooled testing strategies are implemented. The studies expressed resources saved in terms of money amounts of costs or in terms of specific resource requirements such as the number of tests saved, reduction in testing hours, and changes in human resources required.

Five studies relayed cost savings in terms of money as an advantage of pooled testing over individual testing. It should be noted that not all declared whether their calculations included costs saved in compensation of personnel, cost saved in utilities, costs saved in test kits, and other costs saved due to the decrease in tests required when pooled testing is employed so their estimates can still vary.

For the studies specifying reduction in number of tests saved when implementing pooled testing in comparison to the standard individual PCR testing, they show that given low positivity rates, using a pooled testing strategy may reduce the number of tests to be conducted individually from 62 – 87%, consequently increasing the capacity of testing for COVID –19. In terms of testing hours, studies show that pooled testing, when done using a strategic pool size among a population with low prevalence, will save the laboratory hours of processing time even when positive pools have to be deconvoluted and tested individually. . As with the changes in personnel required, the one study concluded that when pooled testing is employed, less laboratory staff would be required but there will be a small increase in the need for clerical staff.

In the study by Lo et al (2020) conducted in the Philippine setting, savings were reported in terms of the number of test kits saved when employing multi-stage Dorfman pooled testing. They found that for a population with a positivity rate of 3%, employing pooling strategies can reduce the number of tests required by 69-83%. The study also showed that test savings is a function of the positivity rate or prevalence rate by calculating the average savings per batches of 100 samples using simulations. The calculations showed that as the positivity/prevalence rate increases, the number of test savings decreases. In terms of testing hours, the study found that pooled testing strategies would increase the turnaround time since more than one batch run would be required to release the results of a positive or negative sample.

4. METHODOLOGY

4.1. Literature Search Methods

Six reviewers performed a targeted search on relevant evidence and information on performance standards and validation testing requirements by selected regulatory bodies, testing guidelines from selected countries, and positions or assessment recommendations from selected HTA agencies regarding the use of pooled testing in diagnosis, screening and surveillance and resource requirements. Below are the targeted sources reviewed:

Table 2. List of Countries, Agencies, and Databases Searched

Regulatory standards, validation requirements, and resource requirements	11 Regulatory Agencies - Health Canada, Japan Pharmaceutical and Medical Devices Agency (PMDA), UK Medicines and Healthcare Products Regulatory Agency (MHRA), US Food and

	Drug Administration (US FDA), and the Philippine Food and Drug Administration (PH FDA), Australia Therapeutic Goods Authority (TGA), European Medicine Agency (EMA), The French Agency for Food, Environmental and Occupational Health and Safety (ANSES), Germany Federal Institute for Drugs and Medical Devices (BfArM), The Pharmaceutical Service Ministry of Health Republic of Italy, Swissmedic Switzerland
National Testing Guidelines	14 National Testing Guidelines – US Center for Disease Control (US CDC), Japan Ministry of Health Labor and Welfare, South Korea Ministry of Health and Welfare, Vietnam Ministry of Health, United Kingdom National Health Service (UK NHS), Center for Disease Network Australia, Malaysia Ministry of Health, China Center for Disease Control, Philippines Department of Health (PPH DOH), Public Health Canada, Singapore Ministry of Health, Indonesia Ministry of Health, Thailand Ministry of Health and the World Health Organization (WHO), European Centers for Disease Control
HTA Agency Reviews	10 HTA agencies - EUnetHTA, US Agency for Healthcare Research and Quality (AHRQ), UK National Institute for Health and Care Excellence (NICE), Australia Medical Services Advisory Committee (MSAC), Canadian Agency for Drugs and Technologies for Health (CADTH), China National Health Economics Institute (NHEI), Indonesian Health Technology Assessment Committee (InaHTAC), Malaysian Health Technology Assessment Section (MAHTAS), Singapore Agency for Care Effectiveness (ACE), South Korea National Evidence-based Healthcare Collaborating Agency (NECA)

For the evidence on diagnostic performance as well as resource requirements, two reviewers performed a literature search of relevant studies published from inception to October 5, 2020 via Pubmed and LOVE database. The search terms used were pooled testing and its related terms as well as COVID-19 and other related terms as defined by the supplementary concept in the MeSH database. We also searched for the PSMID Rapid Evidence Reviews on COVID-19 Management for additional reviews through their website, as well as clinicaltrials.gov and the World Health Organization International Clinical Trials Registry Platform for ongoing/future clinical trials on pooled testing. No filters/ restriction on study type, language and publication date were applied.

Microsoft Word and Google translate were used for direct English translation of contents obtained from both the targeted search and systematic literature search which were not originally written in the English language. On the other hand, Endnote version X9.3.3 was used to organize list of references.

4.2. Selection Criteria and Methods

For the review of the performance standards and validation requirements by selected regulatory bodies, country guidelines, the positions/recommendations from HTA agencies, and the resource requirements, a total of six reviewers screened the documents which were included in our review.

For the review of evidence on the diagnostic performance of pooled testing, seven researchers independently screened study titles and abstracts. Disagreements were resolved by consensus first, then by consulting a third review author. Relevant titles and abstracts were then subjected to independent full-text screening by the same seven authors to identify the final list of studies for inclusion in our rapid review. Disagreements were resolved by consensus first, then by consulting a third review author. Title and abstract screening were conducted using Rayyan QCRI while Microsoft Excel was used for full-text screening.

The references from the systematic search were evaluated against the following inclusion and exclusion criteria:

Table 3 . Inclusion criteria for considering studies for review

Population	Any human population (patient samples) tested for COVID-19 disease, applied for use cases diagnosis, screening and surveillance
Intervention/ Exposure	RT-PCR Pooled testing
Comparator	individual RT-PCR testing or use of previously identified positive or negative SARS-CoV-2 samples
Outcomes	*Primary Outcomes:* Sensitivity; specificity; positive predictive value [PPV]; negative predictive value [NPV]; likelihood ratios; number of true positives, false negatives, true negatives, and false positives *Secondary Outcomes:* Resources saved (e.g. tests/reagent, time, cost)
Study Designs	Systematic reviews (SRs), meta-analyses, primary diagnostic accuracy studies which are externally conducted by an independent laboratory

Exclusion criteria
Articles were excluded if they were preprints or modelling studies on pooled testing for COVID-19.

4.3. Data extraction and Management

The following information were obtained for each reference type:

Table 4. Extracted Information (Testing Guidelines, Regulatory Agencies, and HTA Reports)

Regulatory approval	• Country of origin/Name of regulatory agency • Approved use cases of pooled testing • Performance standards • Validation requirements
Testing guidelines	• Country of origin • Originating agency of the guidelines • Use cases of pooled testing • Target population of pooled testing • Recommendation

Reviews from HTA agencies	• Country of Origin/Name of HTA agency • Use case of pooled testing • Target population • Conclusion/Recommendation

Meanwhile, for the studies on diagnostic accuracy, seven reviewers extracted the following information from the included studies using a pre-defined data extraction form. Disagreements were resolved by consensus first, then by consulting a third review author.

Table 5. Extracted Information (Diagnostic Accuracy Studies and Resource Requirements

Studies on Diagnostic accuracy and Resource requirement	• Country of origin • Number of samples analyzed • Characteristics of patients (e.g. presence of symptoms, onset of disease, severity of disease, co-morbidity, age, special population, etc.) • Index test (brand and manufacturer) • Reference standard (brand and manufacturer) • Specimen used • Mode of sample collection (i.e. prospective, retrospective) • Use Case • Prevalence of COVID-19 from population where samples were takenPool size • Number of true positive, false negative, true negative, false positive • Accuracy measures (sensitivity, specificity, PPV, NPV, LRs, etc.) • Cost savings • Resources saved • Declared conflict of interest • Funding source

4.4. Assessment of Methodological Quality

Information from studies obtained for the clinical accuracy studies were extracted and summarized for key data domains mentioned above using a standard data extraction tool. The Evaluation of Articles on Diagnosis (Dans et al, 2017) tool was used to evaluate the quality of the included clinical accuracy studies. Based on the internal validity domain of the tool, the following cutoff scores were used by the reviewers to identify low, moderate, and high validity studies:

Table 6. Cutoff Scores for Internal Validity

Number of satisfied criteria under internal validity domain	Classification of Internal Validity
1	Low
2-3	Moderate
4	high

4.5. Data Synthesis

Sensitivity was used to measure the true positive rate while specificity was used to measure the true negative rate of pooled testing as compared to individual RT-PCR testing or status of the sample used. In addition, we computed for values such as positive predictive value, negative predictive value, positive likelihood ratio, and negative likelihood ratio. Two units of analysis were used in the study: (1)pools, and (2) individuals. We define the statistics used for the two units of analysis as well as the formula in the table below:

Table 7. Parameters and Units of Analysis used in the review

Statistic	Pool	Individual
Sensitivity	Proportion of **pools** with the target condition in whom the test is positive Computed by the following formula: $\frac{\#\ of\ TP\ pools}{\#\ of\ TP\ pools + \#\ of\ FN\ pools}$	Proportion of **individuals** with the target condition in whom the test is positive Computed by the following formula: $\frac{\#\ of\ TP\ patients}{\#\ of\ TP\ patients + \#\ of\ FN\ patients}$
Specificity	Proportion of **pools** without the target condition in whom the test is negative Computed by the following formula: $\frac{\#\ of\ TN\ pools}{\#\ of\ TN\ pools + \#\ of\ FP\ pools}$	Proportion of **individuals** without the target condition in whom the test is negative Computed by the following formula: $\frac{\#\ of\ TN\ patients}{\#\ of\ TN\ patients + \#\ of\ FP\ patients}$
Positive predictive value	Proportion of **pools** with a positive test who have the target condition. Computed by the following formula: $\frac{\#\ of\ TP\ pools}{\#\ of\ TP\ pools + \#\ of\ FP\ pools}$	Proportion of **individuals** with a positive test who have the target condition. Computed by the following formula: $\frac{\#\ of\ TP\ patients}{\#\ of\ TP\ patients + \#\ of\ FP\ patients}$
Negative predictive value	Proportion of **pools** with a negative test who do not have the target condition. Computed by the following formula: $\frac{\#\ of\ TN\ pools}{\#\ of\ TN\ pools + \#\ of\ FN\ pools}$	Proportion of **individuals** with a positive test who have the target condition. Computed by the following formula: $\frac{\#\ of\ TN\ patients}{\#\ of\ TN\ patients + \#\ of\ FN\ patients}$
Positive likelihood ratio	Probability that a **pool** with the target condition tested positive (true positive) divided by the probability that a **pool** without the target condition tested positive (false positive) Computed by the following formula: $\frac{Sensitivity}{1 - Specificity}$	Probability that an **individual** with the target condition tested positive (true positive) divided by the probability that an **individual** without the target condition tested positive (false positive) Computed by the following formula: $\frac{Sensitivity}{1 - Specificity}$

Negative likelihood ratio	Probability that a **pool** with the target condition tested negative (false negative) divided by the probability that a **pool** without the target condition tested negative (true negative) Computed by the following formula: $\frac{1 - Sensitivity}{Specificity}$	Probability that an individual with the target condition tested negative (false negative) divided by the probability that an **individual** without the target condition tested negative (true negative) Computed by the following formula: $\frac{1 - Sensitivity}{Specificity}$

We obtained the true positive, false negative, true negatives, and false positive data from included studies. Whenever possible, we tried to reconstruct the 2x2 table if sufficient data was available. The 95% confidence intervals for the measures mentioned above were computed as well. For sensitivity and specificity, we used Clopper-Pearson confidence intervals while for positive and negative predictive values, we used the standard logit confidence intervals given by Mercado et al. (2007). All computations for the point estimates and confidence intervals of sensitivity and specificity were conducted using an online calculator (https://www.medcalc.org/calc/diagnostic_test.php). Meanwhile, confidence intervals for likelihood ratios were computed using the log method as described by Altman et al (2000) using the *madad* function in R.

A univariate meta-analysis was performed to pool the sensitivity and specificity from published studies using the *metaprop* function in R. In general, we used an inverse variance method with DerSimonian-Laird estimator for between-study variance, Jackson method for confidence interval of τ^2 and τ, logit transformation, and normal approximation confidence interval for individual studies. For data sets where majority report to have extremely common (i.e. *p=1)* or extremely rare (i.e. *p*=0) events of interest (i.e. proportion of true positive or true negative), the arcsine transformation was used instead of the logit transformation due to its limitations, one of which is its incapability of stabilizing variance. In addition, the logit transformation uses a continuity correction factor for extreme proportions due to the logits and sampling variances being undefined and applies these even to studies that have no such problems, resulting in a more biased result (Wang, 2018). Further, the use of the arcsine transformation is supported by the recommendations of Schwarzer et al. (2019) and Trikalinos, Trow, & Schmid (2013).

The I^2 statistic was used to measure heterogeneity between studies. I^2 statistic values of 0-30%, 31-50%, 51-75%, and 76-100% indicate insignificant heterogeneity, moderate heterogeneity, substantial heterogeneity, and considerable heterogeneity, respectively. These ranges and interpretations were adapted from the Cochrane Handbook for Systematic Reviews of Interventions (Higgins et. al., 2019). Sources of heterogeneity were further explored by doing subgroup analysis based on identified variables when sufficient data were available. These included pool size, presence of symptoms, prevalence of disease where sample was obtained, onset of symptom of disease (for symptomatic), days since first positive PCR test or days since contact with a confirmed case (for asymptomatic), brand of index test, use case, specimen type, and Ct values.

5. SUMMARY OF EVIDENCE

5.1. Regulatory Standards on Pooled Testing

From the 11 regulatory agencies listed and reviewed, only the US Food and Drug Administration (US FDA) has authorized the use and has established its validation testing requirements and performance standards for pooled RT-PCR testing.

5.1.1. Authorization of RT-PCR pooled testing for COVID-19

In the EUA diagnostic template, patients suspected of COVID-19 infection and individuals without symptoms or other reasons to suspect the said infection are the targeted population for pooled RT-PCR testing. According to US FDA (2020), the use of sample pooling in certain SARS-CoV-2 tests (i.e., screening and diagnostic tests) can be an option and authorized granting that there exist proper mitigations for reduction of analytical sensitivity and validations of the test, thereby, ensuring its proper implementation. Furthermore, on the use of SARS-CoV-2 diagnostic test for surveillance purposes (e.g., determining the prevalence of acute infections in a population), the US FDA explicitly stated that it does not generally regulate the use of a test for this purpose or use case. Even though the US FDA does not regulate surveillance testing, the CDC suggests that laboratories should use the US FDA authorized assay and test system (CDC, 2020). Nonetheless, it was mentioned by US FDA that if surveillance testing is performed by a non-Clinical Laboratory Improvement Amendments (CLIA) certified laboratory, a confirmatory test on the detected positive individuals should be performed by a CLIA-certified laboratory.

Further, the US FDA advises that sample pooling works best in areas where the COVID-19 infection rate is low or in the low risk population, however, there was no operational definition of infection or prevalence rate of the disease. As for the method of specimen pooling, the EUA templates laid out two (2) approaches: sample/media pooling and swab pooling. *Sample/media pooling* is a Dorfman approach to testing wherein a single pool consists of aliquots sample of transport media (each transport media has a specific volume which contains a single swab sample). For this method, a negative result treats all the samples as negative while a positive result denotes that at least one sample in the pool is positive and in the process, the positive pool is then tested individually. On the other hand, *swab pooling* refers to the multiple inclusion of individual swabs into a single volume of transport media (US FDA, 2020). Considering that swab pooling is a dense collection of swab specimens from numerous individuals, it is expected that there will be a high viral load in the pool, if there are multiple positive swabs present, where inhibition effects may be observed. Anent this, the US FDA emphasizes the collection of another specimen since deconvolving a positive swab is not possible.

Currently, the US FDA has registered and authorized ten (10) molecular diagnostic tests for SARS-CoV-2 for pooled testing, namely: *Panther Fusion SARS-CoV-2 Assay, Quest SARS-CoV-2 rRT-PCR, LabCorp COVID-19 RT-PCR Test, Verily COVID-19 RT-PCR Test, BayCare SARS-CoV-2 RT PCR Assay, Poplar SARS-CoV-2 TMA Pooling assay, UCSD RC SARS-CoV-2 Assay, Viracor SARS-CoV-2 assay, Aptima SARS-CoV-2 assay and cobas SARS-CoV-2* (US FDA, 2020). Of these diagnostic kits approved by the US FDA for pooled testing, only *cobas SARS-CoV-2* test is included in the Philippine FDA list of authorized COVID-19 PCR test kits (as of 03 November 2020) but used for individual testing only as regulatory standards for pooled testing in the Philippines do not exist yet.

5.1.2. Validation requirements for COVID-19 pooled testing

To establish the performance of the RT-PCR kit when used for pooled testing, the US FDA requires the manufacturer to conduct clinical validation studies in the intended population and using their proposed pooling strategy.

The US FDA has issued separate guidelines for validation of pooled testing for sample/media pooling and swab pooling. For sample/media pooling, the US FDA requires only one validation study that establishes the performance of pooled testing compared to individual testing. The parameters of the study depend on whether the RT-PCR kit intended for pooled testing has been previously granted an EUA or not. For swab pooling, the US FDA requires three validation studies: 1) a study to establish the performance of pooled testing compared to individual testing; 2) a study to establish the performance of the RT-PCR kit when multiple swab specimens interfere with the reading of results; 3) a study to evaluate the effect of high viral concentrations on assay performance.

The US FDA recommends choosing a pool size in consideration of the positivity rate of the test population, the sensitivity of the RT-PCR test and the percent of weak positive samples (defined by the US FDA as those having Ct values close to the limit of detection of the assay) in the test population. They also advise the developers to start their clinical validation studies using a pool size of 5. A test authorized by the US FDA for pooled testing with pool size n may also be used for pooled testing with a pool size below n.

5.1.2.1. Validation of pooled testing by sample/media pooling

The validation requirements of the US FDA to include pooled testing by sample/media pooling in the authorized use cases for an RT-PCR kit (i.e., screening and diagnosis) depends on whether the kit has been granted an EUA by the US FDA or not. The table below shows the study parameters for previously authorized RT-PCR kits and new RT-PCR kits.

Table 8. Validation Testing Requirements for Sample Pooling

Parameters	For an RT-PCR kit that has been previously authorized	For a new RT-PCR test that has not been previously granted an EUA
Clinical study requirement	Yes	Yes
Comparator	Individual testing using the RT-PCR kit of the developer	Individual testing using a comparator assay that has an established high sensitivity with an internationally recognized standard or the FDA SARS-CoV-2 reference panel
Index test	Pooled testing using the RT-PCR kit of the developer	Pooled testing using the RT-PCR kit of the developer
Sample size	At least 20 positive samples and an appropriate number of negative samples, depending on the proposed pool size	At least 30 positive samples and an appropriate number of negative samples, depending on the proposed pool size

Composition of n-sample pools (pool size = n)	At least 20 positive pools: 1 positive sample and n-1 negative samples per pool At least 20 negative pools: n negative samples per pool	At least 30 positive pools: 1 positive sample and n-1 negative samples per pool At least 30 negative pools: n negative samples per pool

5.1.2.2. Validation of pooled testing by swab pooling

If the test developer intends to use their RT-PCR test kit for swab pooling, they must conduct three validation studies. The first study is done to establish the performance of the test kit for pooled testing compared to individual testing. The guidelines for this validation study remain the same as described above for sample/media pooling. The two other studies described below evaluate the effect of inhibition due to 1) high concentrations of swab specimens and 2) high concentrations of viral load in a single volume of transport media. The swab pooling approach results in large amounts of swab specimen such as mucin, as specified by the US FDA, in minimal viral transport media making the sample subjected to RT-PCR very concentrated. The effects of inhibition due to high concentrations of swab specimen and high concentrations of the virus must be investigated by conducting these two validation studies. These two validation studies will be conducted using the RT-PCR kit for which swab pooling is being validated , whether it has previously been granted an EUA or not.

To establish the performance of the RT-PCR kit when the samples have high concentrations of swab specimens (e.g., mucin), the US FDA requires testing a negative pool of n-swab samples in the minimum volume of transport media spiked with positive sample at a concentration of 2-3x the LoD of the assay. For this study, the US FDA requires testing at least 3 pools and a total of at least 20 replicates composed of equal number of replicates from each pool sample. Each pool sample to be tested should contain the maximum number of swabs recommended by the test developer for pooling.

To establish the performance of the RT-PCR kit in testing samples with an unexpectedly high viral titer, the validation study required by the US FDA involves spiking a single negative sample with at least three times the expected viral load in a single positive swab. Ten replicates shall be tested using RT-PCR.

5.1.3. Performance standards for COVID-19 pooled testing

The US FDA has issued separate sets of performance standards for (1) For sample/media pooling; and (2) For pooled testing by swab pooling. The table below summarizes the required performance standards for the different validation studies conducted for sample/media pooling and swab pooling.

Table 9. Performance Standards for Sample Pooling and Swab Pooling

Performance standard domain	Sample/Media Pooling	Swab Pooling
Performance for pooled testing vs individual testing	• ≥85% PPA	

	• Individual weak positive samples with high Ct values (viral loads close to LoD of the assay) can be accurately detected when pooled with n-1 negative samples	
Interference of swab specimens	N/A	• 95% agreement with expected results and <5% invalid rate
Interference of viral load	N/A	• 100% positivity rate or • ≤5% invalid rate

5.2. Performance Characteristics

5.2.1. Quantity and Characteristics of Included Studies (Completed Studies)

Figure 1 illustrates the PRISMA flowchart for the study selection. Records identified through Pubmed and LOVE database yielded a total of 260 records. In addition, we obtained 12 records of ongoing/future trials from clinicaltrials.gov and the WHO International Clinical Trials Registry Platform. Furthermore, we obtained one rapid review from the website of the Philippine Society for Microbiology and Infectious Diseases and one unpublished local study by Lo et al. (2020) but the study has already been submitted for publication. Overall, we were able to obtain 274 records for deduplication, which was reduced to 251 after removal of duplicates. Of the 251 articles, 177 were excluded leaving 74 references for full text assessment. Among the 74 eligible references, a total of 54 were excluded leaving 20 articles for inclusion in data synthesis.

Of the 54 excluded references from full-text review, 2 studies did not include the intervention of interest, 4 did not have the comparator of interest, 11 did not report the outcomes of interest, 29 were not the study design of interest (e.g., case reports, narrative reviews), and 8 were preprint articles. Note that there was 1 rapid review retrieved which was used to map additional primary clinical studies that may have not been captured by our search. Upon review, the relevant studies in the rapid review that fit our inclusion criteria have been included in our analysis.

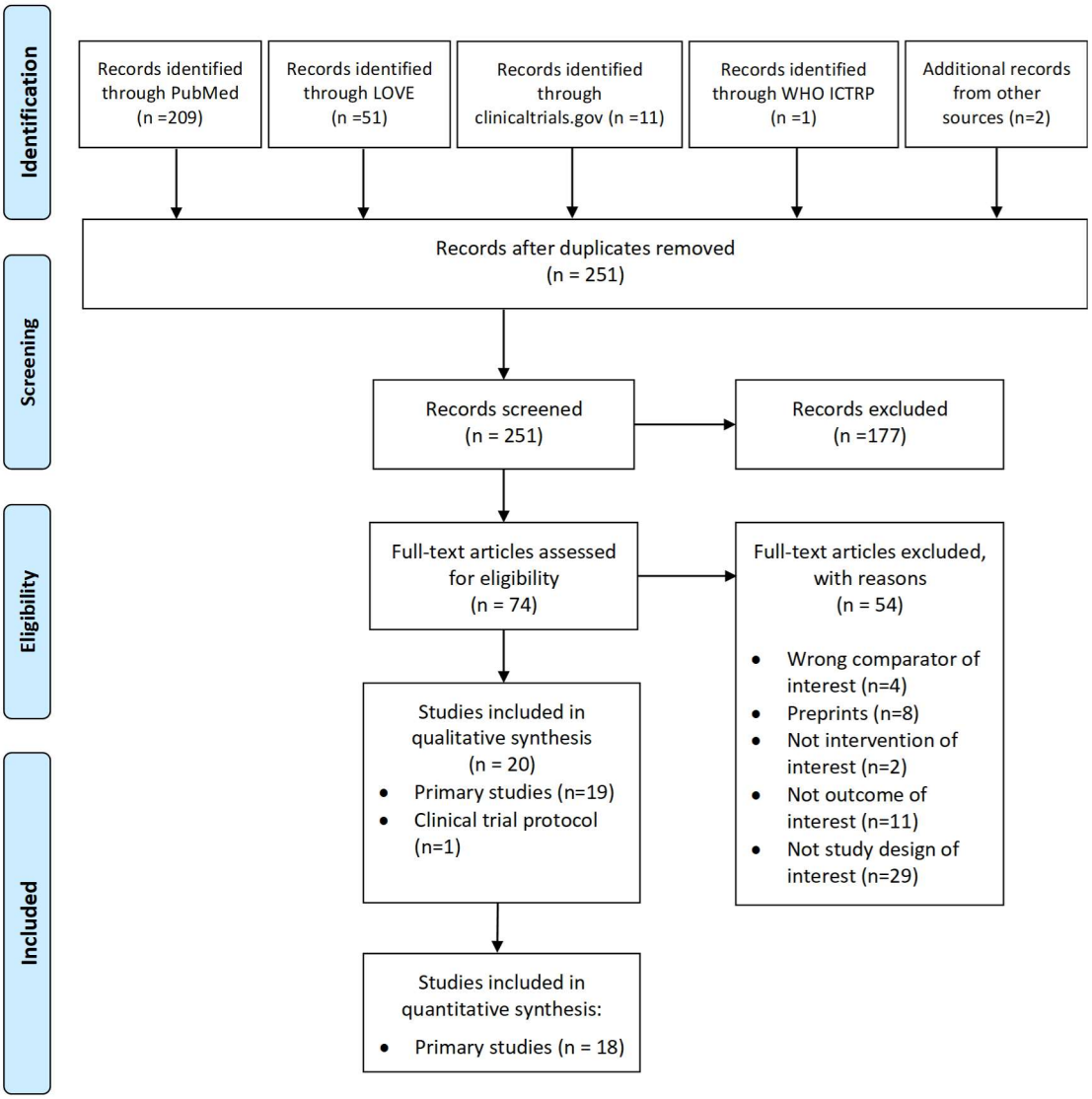

Figure 1. PRISMA flow diagram of studies

All 20 studies were included for qualitative synthesis in this rapid review with 19 studies being primary diagnostic accuracy studies and one being a clinical trial protocol. Meanwhile, only the 18 primary studies were included in the quantitative synthesis. Of the 18 primary studies, 11 were conducted in Asia, 2 were conducted in Europe, 5 were conducted in North America, and 1 was conducted in South America. For the clinical trial protocol, the study will be conducted in Asia, particularly in India and further details will be discussed in the section on characteristics of ongoing/future trials. The other primary study which was a field validation study of pooled testing was conducted in Asia particularly in the Philippines in two phases, using previously tested samples and samples collected from employees in a local supermarket. In addition, this study looked at the performance of two (5-1), three (10-5-1), and four (20-10-5-1) stage Dorfman pooling. Results of this local study were compared to the meta-analysis conducted for the 18 international primary studies.

In terms of the primary studies, eight (8) looked into the use of pooled testing for screening, five (5) looked into the use of pooled testing in the diagnosis of COVID-19, two (2) studies explored

the use of pooled testing in COVID-19 surveillance, and three (3) studies did not mention a particular use case for pooled testing that was being explored by their study. Furthermore, all 18 studies utilized a cross-sectional design to determine diagnostic accuracy of pooled testing as compared to individualized RT-PCR testing.

In terms of population characteristics, some tested populations were symptomatic patients, persons under investigation, and asymptomatic individuals. For the asymptomatic individuals, one data set from one study mentioned use of pooled testing among asymptomatic healthcare personnel, employees of essential industries, and residents and employees of nursing homes, one study mentioned use of pooled testing among asymptomatic residents of a nursing home. Moreover, all of the studies (18) did not characterize the full details of their population of interest such as age, comorbidities, and other risk factors. Out of the 18 studies, six did not provide any details on population characteristics, referred to as "Unspecified" in Appendix 1. In addition, majority of the studies were conducted as laboratory-based simulations of pooled testing rather than field validation.

Three of the studies conducted multiple procedures to validate pooled testing. Two of them conducted 3 validation procedures while one conducted 2 procedures in their research. While majority of the studies performed evaluation in the context of the two-stage pooled testing, four studies explored other methods, too. One of the studies validated matrix pooling which aims to find the conjunction of positive pools to identify the positive sample; one study mimicked pooling by diluting each positive sample with viral transport media instead of pooling positive and negative samples; another studied ratio pooling by using different combinations of pooled negative viral transport media (VTM), VTM from a positive sample with high viral load, and positive VTM with low viral load; and another study validated the Pooling-Based Efficient SARS-CoV-2 Testing (P-BEST) method which aims to identify the positive samples within a single round of RT-PCR.

Of the 19 studies appraised, 18 studies used RT-PCR as a reference standard while the remaining one has an unconfirmed reference standard test. In addition, few studies reported specific specimen used which included nasopharyngeal and/or oropharyngeal swabs and saliva. Majority did not define the specimen used in pooled testing.

Pool sizes studied ranged from as low as 2 samples per pool to as large as 50 samples per pool. Although the positivity rate of the population is an important factor to consider in choosing the pool size (US FDA, 2020), only two of the studies specified the prevalence of COVID-19 in the population where they obtained their samples which were reported to be 4.8% and 9%. Majority of the outcomes that were reported are Ct values difference and data on true positives, false negatives, true negatives and false positives from which measures such as sensitivity, specificity, PPV, NPV, and LRs were obtained.

The characteristics of included studies can be seen in Appendix 1.

5.2.3. Diagnostic Accuracy Findings

5.2.3.1. Unit of analysis: Pools

5.2.3.1.1. Overall pooling

We performed an overall pooling for sensitivity from 15 studies, and for specificity from 10 studies which reported data for these performance measures.

Sensitivity

Meta-analysis was conducted for 15 studies and the overall pooled sensitivity of pooled testing using pools as unit of analysis was 87% (95% CI: 81- 91). The point estimates of sensitivity from the studies ranged from 25 to 100%. The I^2 statistic was 71% which suggests that there is a substantial heterogeneity between the included studies. However, it is important to note that heterogeneity is typically expected for reviews of diagnostic test accuracy.

Meanwhile, the local study by Lo et. al (2020) where pooling by dilution was done to simulate pools of 5, 10, and 20 had an overall sensitivity of 83% (95% CI: 67-94), 72% (95% CI: 55-86), 67% (95% CI: 49-81), respectively.

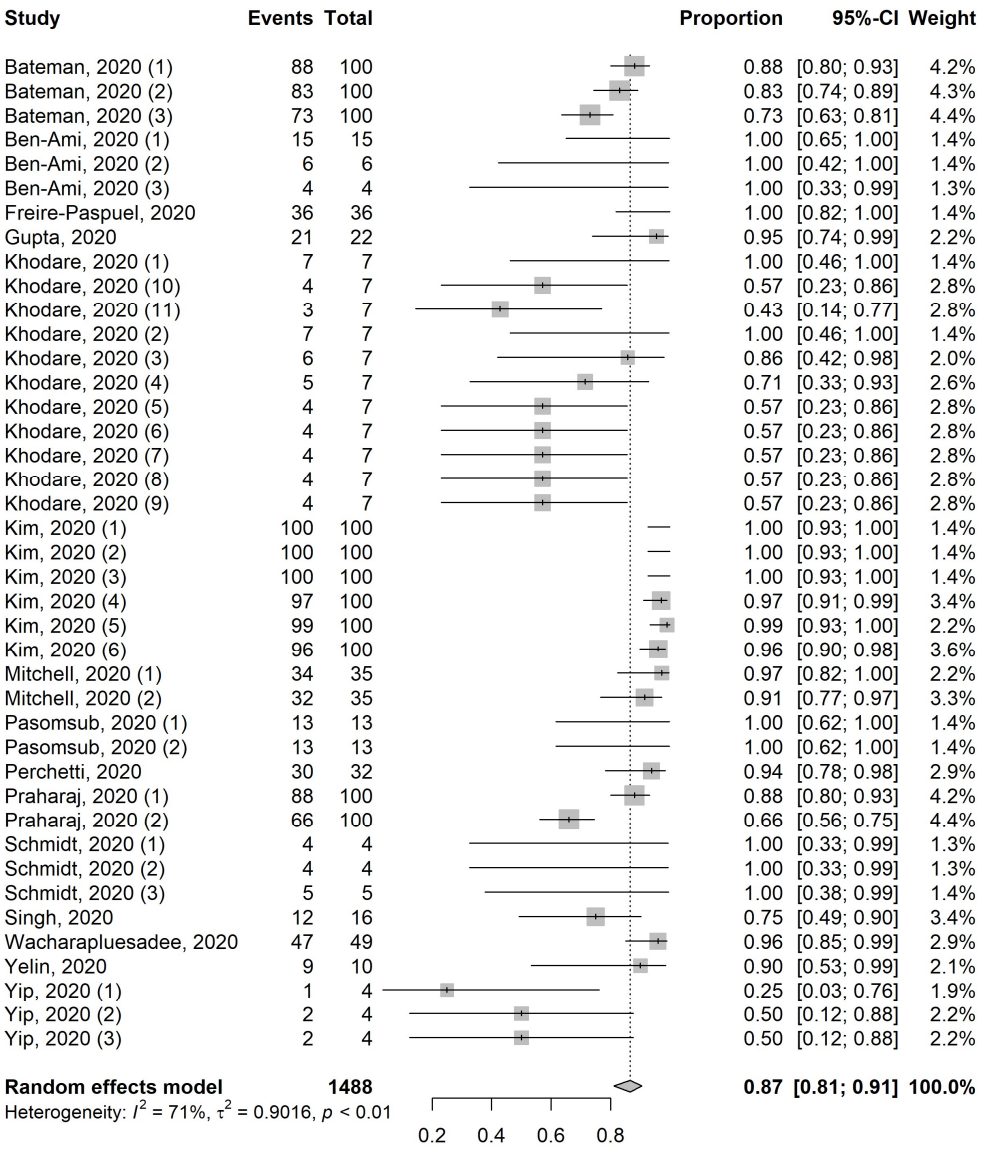

Figure 2. Sensitivity of Pooled Testing, Unit of Analysis: Pools

Specificity

The overall pooled specificity of pooled testing from 10 studies was estimated to be 99.6% (95% CI: 98.9–100). This implies that there is a low rate of false positives. The I^2 statistic was 0% which suggests that there is negligible heterogeneity between the included studies. Since the first phase of the experiment of Lo et al. (2020) used positive specimens only, there is no comparison that can be done for specificity.

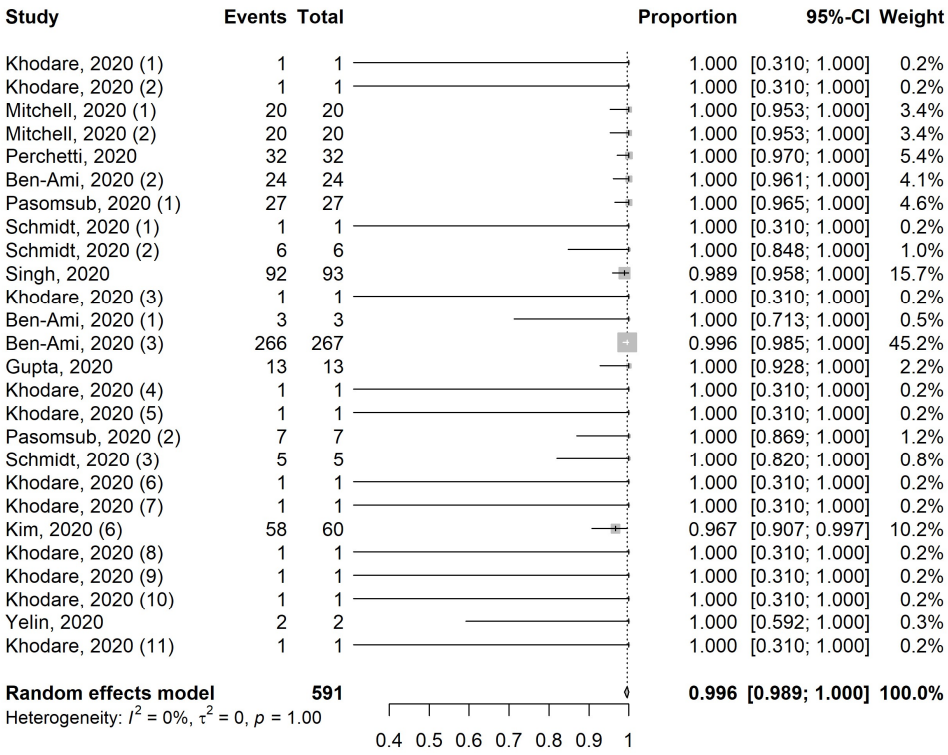

Figure 3. Specificity of Pooled Testing, Unit of Analysis: Pool

The PPV, NPV, +LRs and -LRs are computed for each data set from each study. The point estimates for these measures can be seen in Appendix 2A.

5.2.3.1.2 Subgroup analysis

Subgroup analysis was performed for sensitivity estimates since there is a substantial heterogeneity among the included studies in the overall pooling. Pre-identified factors explored in this review are pool size, brand of index test, CT value threshold, use case, prevalence of disease, and index specimen.

By Pool Size

Generally, pooled testing showed higher pooled sensitivity estimates when pool sizes used are smaller. Subgroups that used a pool size of at most 10 had pooled sensitivity estimates higher than the overall pooled sensitivity (i.e. 87%). Among the subgroups with at least two studies, two-sample pool had the highest pooled sensitivity at 98% (95% CI: 82 - 100) which contains the study of Khodare et al. (2020), and Kim et al. (2020). In contrast, subgroups utilizing 12 or more samples in pooled testing had lower pooled sensitivity estimates compared to the overall pooled sensitivity estimate. Most

notably, the subgroup of 50-sample pool had the lowest pooled sensitivity at 60% (95% CI: 38 - 69) among those that contain multiple studies. These include a part of the results of Bateman et al. (2020), and several results from Yip et al. (2020).

Consistently, the local study by Lo et al (2020) showed that smaller pool sizes lead to higher sensitivity estimates. The pool size of 5 showed highest sensitivity with point estimate of 83% (95% CI: 67-94) followed by pool sizes 10, and 20, with sensitivity point estimates of 72% (95% CI: 55-86) and 67% (95% CI: 49-81), respectively.

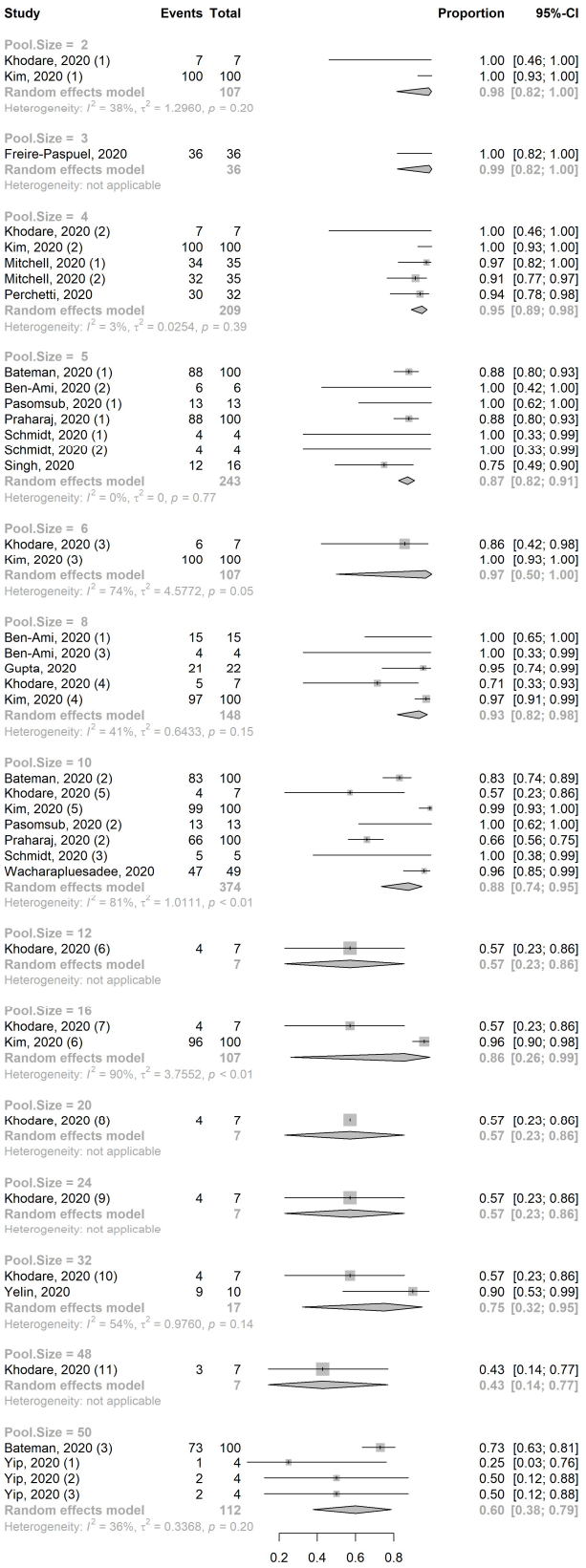

Figure 4. Sensitivity of Pooled Testing by Pool Size, Unit of Analysis: Pool

By Brand of Index Test

Eight test kits had higher sensitivity estimates than the overall pooled sensitivity (i.e. 87%). *PowerChek 2019-nCoV* used by Kim et al. (2020) had the highest with sensitivity at 98% (95% CI: 96-99, I^2-statistic: 13%). This was followed by *Sansure* at 96% (95 CI: 79-99), *LightMix E Gene* at 95% (95% CI: 74-99), and *CDC-based Washington State EUA SARS-CoV-2 RT-PCR Assay* at 94% (95% CI: 78-98) used by Pasomsub et al. (2020), Gupta et al. (2020), and Perchetti et al. (2020), respectively.

Further, three test kits had 91% pooled sensitivity with varying confidence intervals or heterogeneity. These are *BGI RT-PCR kit* (95% CI: 77-97, I^2-statistic: 35%) used by three studies (Ben-Ami et al., 2020; Singh et al., 2020; Wacharapluesadee et al., 2020), *Hologic Aptima* (95% CI: 77-97) used by Mitchell & Ventura (2020), and *Roche Cobas* (95% CI: 64-98, I^2-statistic: 0%) used by Schmidt et al. (2020). Meanwhile, *AgPathID RT-PCR* used by Yelin et al. (2020) had a 90% sensitivity estimate (95% CI: 53-99).

It should be noted that the widely popular *2019-nCoV CDC RT-PCR* had 86% pooled sensitivity (95% CI: 76-93; 5 studies) with a substantial heterogeneity within the subgroup. While studies by Paraharaj et al. (2020) which used several brands had a pooled sensitivity of 79% (95% CI: 50-93).

Four test kits had sensitivity estimate below 75%. These include brands used by Yip et al. (2020) such as *STN-COVID-19-N* at 50% (95% CI: 12-88*), STN COVID-19-RdRp/Hel* at 50% (95% CI: 12-88), and the *non-nested STN COVID-19-RdRp/Hel* at 25% (95% CI: 3-76); as well as by Khodare et al. (2020) for the *LightMix RdRP and E Gene* at 63% (95% CI: 50-73).

In addition, the study of Lo et al. (2020) reported a sensitivity estimate of 83%, 72%, and 67% for the brand *Maccura SARS-CoV-2 Fluorescent PCR Kit* using different pool sizes.

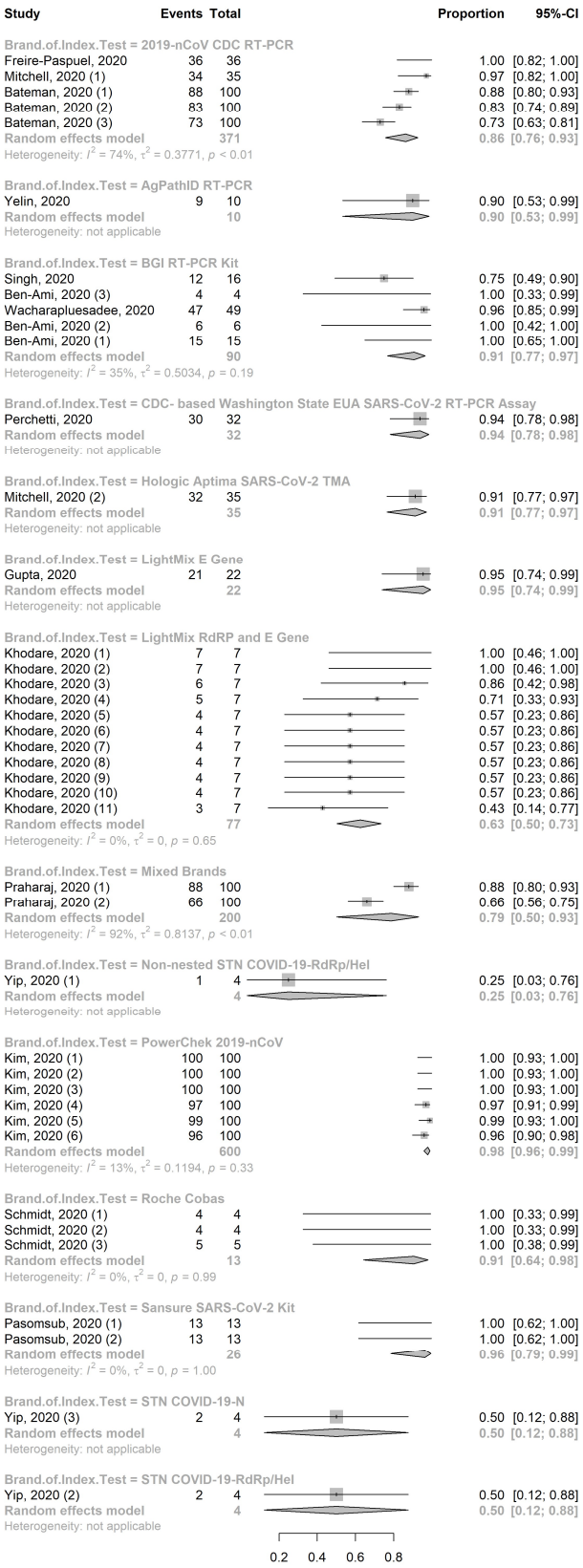

Figure 5. Sensitivity of Pooled Testing by Brand of Index Test, Unit of Analysis: Pool

By Pool Size of 10 and Brand of Index Test

We also explored the effect of pool size and brand of index test on pooled testing sensitivity. For brevity, we categorized the datasets whether they used a pool size of at most 10 or more than 10. From our earlier discussion, we observed that the sensitivity of pooled testing was more stable when the number of samples in the pool is limited to 10.

Among our included studies, seven brands were used in pool sizes <=10 only. Their sensitivity estimates range from 75% [Mixed Brands (k=2)] to 96% [*Sansure SARS-CoV-2* (k=2)]. The remaining brands were *BGI RT-PCR Kit* (k=5), *CDC-based Washington State EUA SARS-CoV2 RT-PCR Assay* (k=1), *Hologic Aptima SARS-CoV-2 TMA* (k=1), *LightMix E Gene* (k=1), and *Roche Cobas* (k=3).

Meanwhile, four brands were used in pool sizes more than 10 only and were notably used on one study only. The range of their sensitivity estimates was from 25% [*Non-nested STN COVID-19-RdRp* (k=1)] to 90% [*AgPath ID RT-PCR* (k=1)]. The other brands were *STN-COVID-19-N* (k=1), and *STN-COVID-19-RdRp/Hel* (k=1).

Finally, three brands were used in both cases. *PowerChek 2019-nCoV* (k=5, 1) had the highest sensitivity estimates at 99% and 96% for pool size of at most and more than 10, respectively. *LightMix RdRP and EGene* (k=5 and 6) had the widest disparity in terms of sensitivity between two pool size categories with 77% for the smaller pool and 55% for the larger pool. All three brands were consistent with this trend including *2019-nCoV CDC RT-PCR* (k=4 and 1).

For comparison with a local validation, *Maccura SARS-CoV-2 Fluorescent PCR Kit* was used in both cases where the sensitivity of the smaller pool sizes was 83% for a pool size of 5 and 72% for a pool size of 10. On the other hand, the sensitivity of the 20-sample pool size was 67% (Lo et al., 2020).

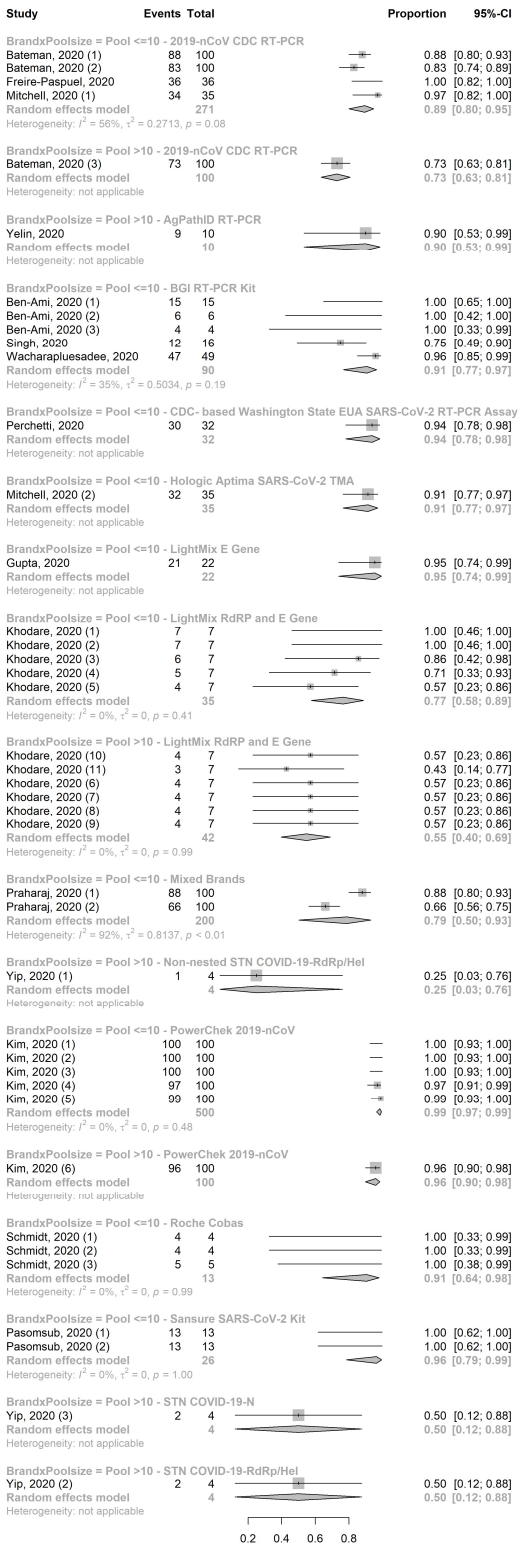

Figure 6. Sensitivity of Pooled Testing by Pool Size of 10 and Brand of Index Test, Unit of Analysis: Pool

By Ct Value

Based on varying Ct values of positive specimens contained in an n pool of samples in the two studies included, there is an observed decrease in sensitivity as the Ct value increases. In the study of Mitchell & Ventura (2020) utilizing two different RT-PCR kits (2019 nCoV CDC RT-PCR and Hologic Aptima SARA-CoV-2 TMA), the pooled sensitivity for Ct values of <34, 34-36, and ≥37 is 98% (95% CI: 87-100, I^2 statistic: 0%), 90% (95% CI: 53-99, I^2 statistic: 0%), and 73% (95% CI: 42-91, I^2-statistic: 19%) respectively. On the other hand, the pooled sensitivity for the study of Praharaj (2020) which utilized different pool sizes had pooled sensitivities of 97% (95% CI: 85-99, I^2 statistic: 0%), 89% (95% CI: 61-98, I^2 statistic: 77%), and 48% (95% CI: 14-85, I^2 statistic: 91%) for Ct values of ≤30, 30-33, and 33-36 respectively. The pooled estimates for each Ct value range came from two data sets from Praharaj et al. (2020) with varying pool sizes of 5 and 10.

Comparing the sensitivity for Ct value of 34-36 in Mitchell, 2020 and 33-36 in Praharaj, 2020 we observe that the pooled estimate of Mitchell (2020) (90%) is higher than that of Praharaj et al. (2020) (48%). This difference may possibly be due to the pool size of 10 in Praharaj (2) (2020) which indicates that sensitivity decreases with the increase in pool size.

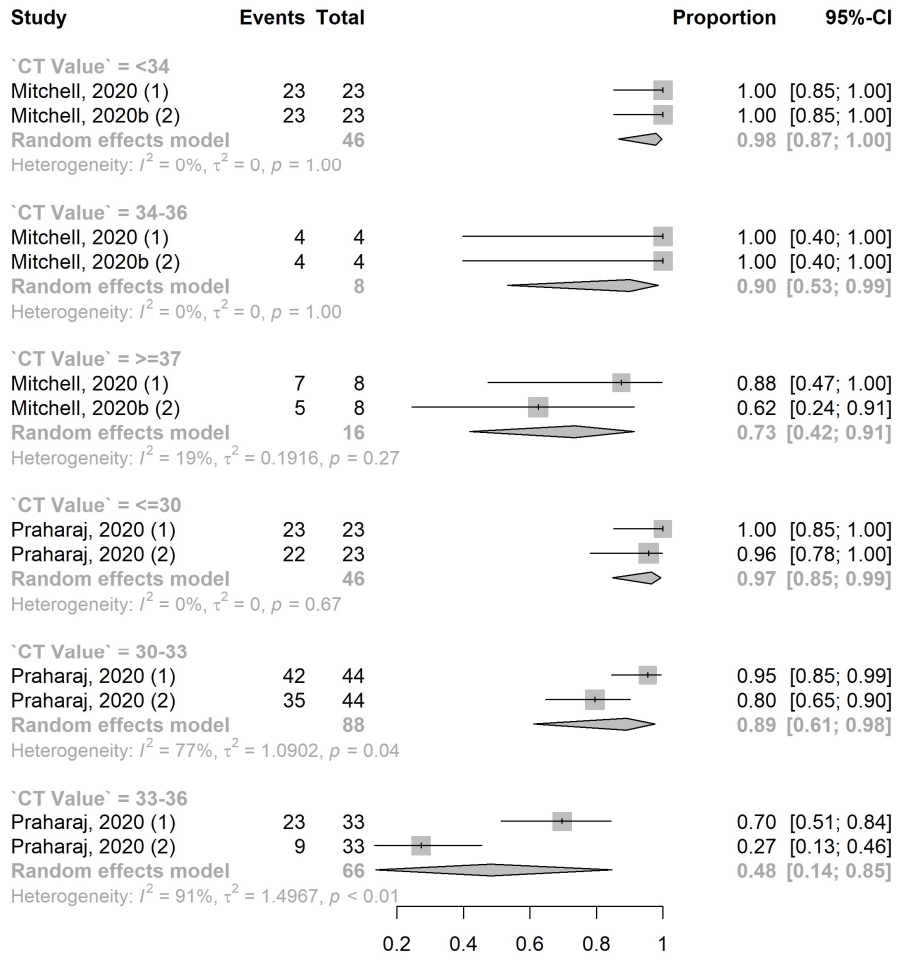

Figure 7. Sensitivity of Pooled Testing by Ct Value, Unit of Analysis: Pool

By Use Case

We also investigated the sensitivity of pooled testing depending on the indicated use case of the study. Pooled testing that was intended for screening had the highest pooled sensitivity estimate at 94% (95% CI: 87-97; I^2 statistic: 70%) from six studies. This was followed closely by those that intended pooled testing for diagnosis with 93% pooled sensitivity estimate (95% CI: 79-98, I^2 statistic: 47%) from four studies. Meanwhile two studies that intended pooled testing for surveillance had a pooled sensitivity of 84% (95% CI: 74-90, I^2 statistic: 77%). In contrast, four studies that did not mention the use case of pooled testing had the lowest pooled sensitivity at 74% (95% CI: 63-82, I^2 statistic: 44%). The study of Lo et al. did not mention the specific use case for the study but reported sensitivity estimates of 67%, 72%, and 83%.

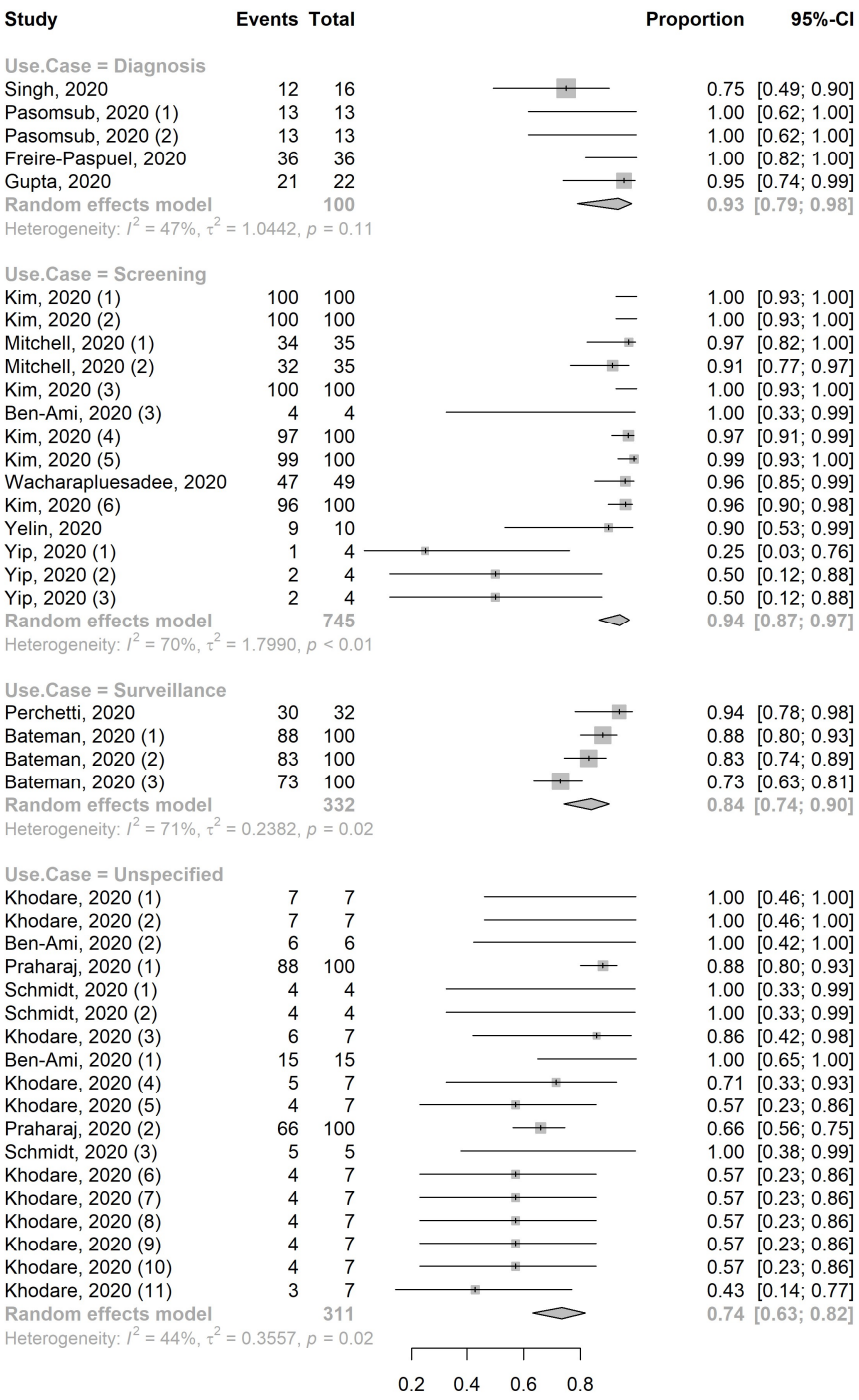

Figure 8. Sensitivity of Pooled Testing by Use Case, Unit of Analysis: Pool

By Presence of Symptoms

The overall pooled sensitivity of pooled testing among asymptomatic patients based on two studies was found to be at 91% (95% CI: 56-99, I^2-statistic: 0%). On the other hand, two studies that conducted five evaluations with varying pool size and pooling strategy that used both symptomatic and asymptomatic

population in pooled testing had a pooled sensitivity of 83% (95% CI: 74-90, I^2 statistic: 60%). Only one data set from one study (Schmidt et al., 2020) reported using symptomatic patients in pooled testing and the reported sensitivity is at 100% (95% CI: 33-99). Given the limited number of studies clearly delineating symptomatic versus asymptomatic population, caution must be considered in generalizing the effect of symptoms on the performance of pooled testing as compared to individual PCR tests. In addition, comparisons with the local study cannot be done since Lo et al (2020) did not specify the characteristics of patient samples used in the first phase of the study in terms of presence of symptoms.

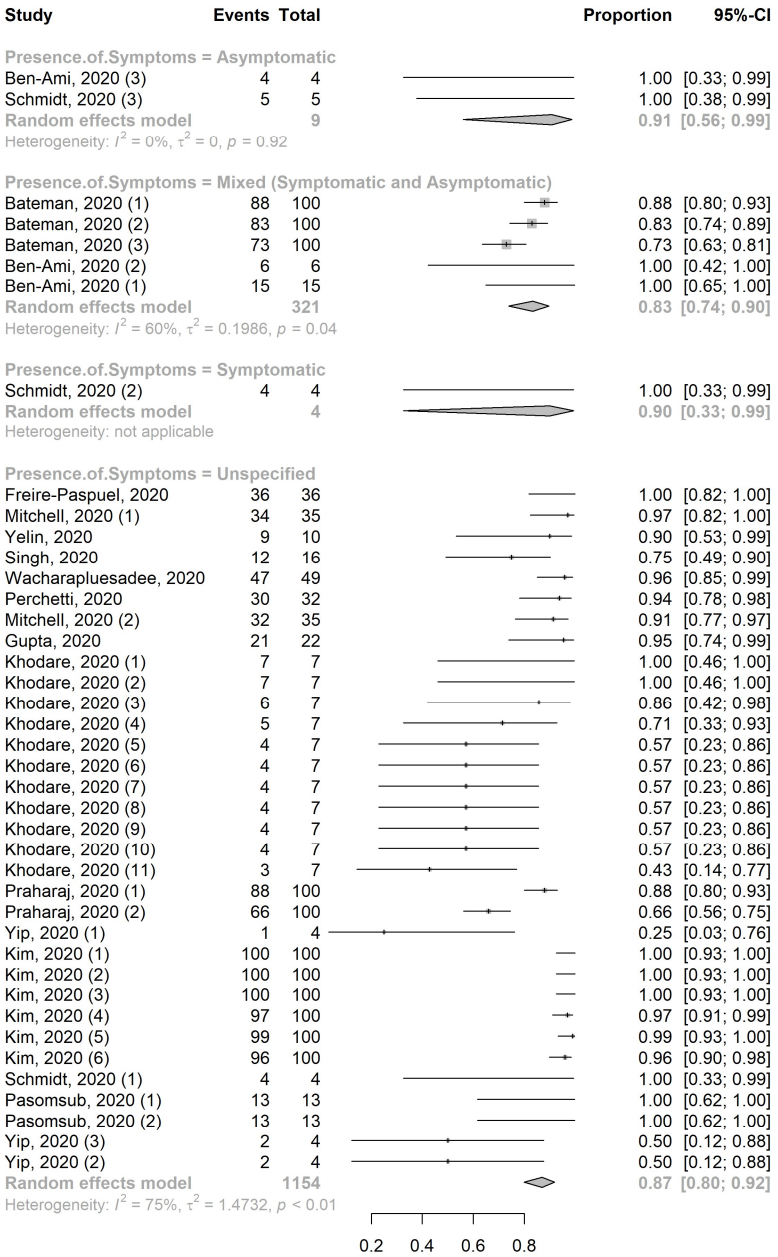

Figure 9. Sensitivity of Pooled Testing by Presence of Symptoms, Unit of Analysis: Pool

By Onset of Symptoms

While we intended to analyze the diagnostic performance of pooled testing as stratified according to the onset of symptoms (for symptomatic) and day since first positive PCR test or day since contact with a confirmed case (for asymptomatic patients), none of the included studies have reported this outcome of interest.

By Prevalence

It is also of our interest to confirm through subgroup analysis whether prevalence of the disease where sample was obtained is a factor in the performance of pooled testing. However, few studies have known or have specified this in their paper. Hence, it is difficult to generalize sensitivity estimates from Singh et al. (2020), and Pasomsub et al. (2020) which indicated a prevalence of 4.8%, and 8%, respectively. Moreover, comparisons with the local context are not feasible since Lo et al (2020) did not specify the prevalence of COVID-19 in the sample obtained in the first part of the study.

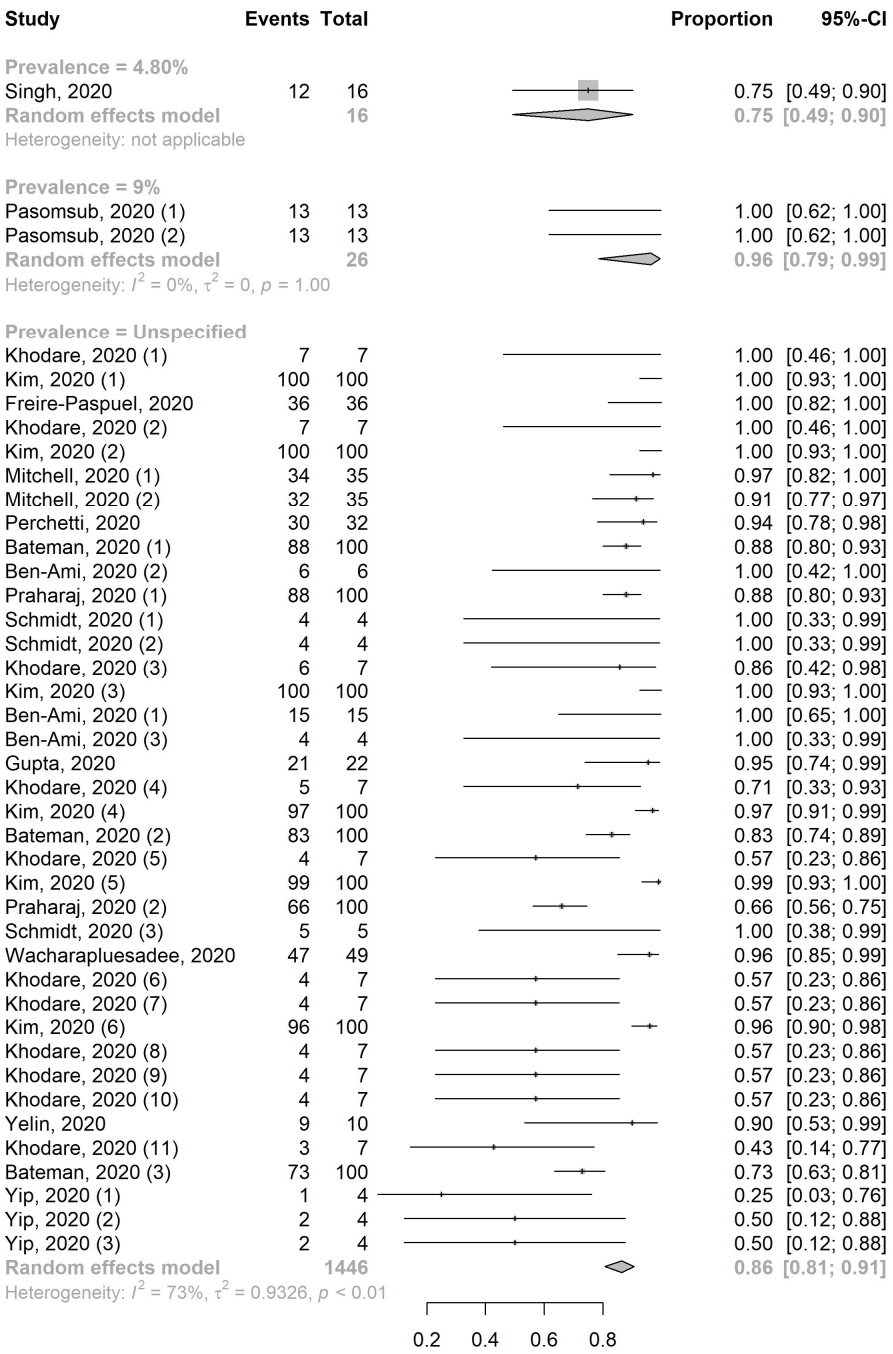

Figure 10. Sensitivity of Pooled Testing by Prevalence of Symptoms, Unit of Analysis: Pool

By Specimen

Pooled testing showed a relatively high sensitivity when using nasopharyngeal and saliva specimens with pooled estimates of 94% (95% CI: 88-97, I^2 statistic: 0%, k=4) and 96% (95% CI: 79-99, I^2 statistic: 0%, k=2) respectively. On the other hand, studies that used combined nasopharyngeal and oropharyngeal specimens had a pooled estimate of 88% (95% CI: 80-93, I^2 statistic: 74%, k=26). In contrast, the local study by

Lo et al (2020) had lower sensitivity estimates of 83%, 72%, and 67% which also used combined nasopharyngeal and oropharyngeal specimens.

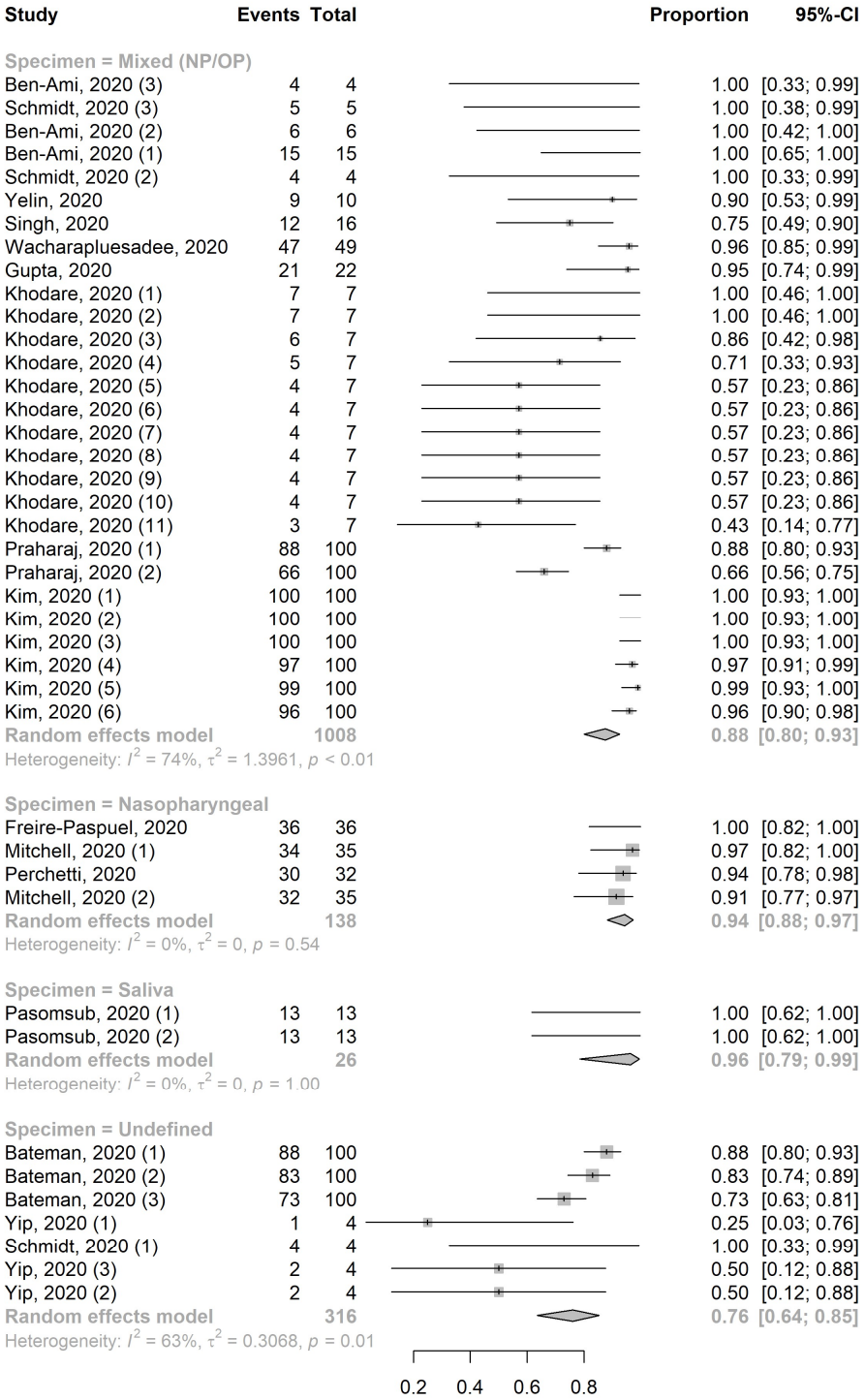

Figure 11. Sensitivity of Pooled Testing Specimen, Unit of Analysis: Pool

5.2.3.2. Unit of analysis: Individuals

For the meta-analysis of pooled testing with individuals as units of analysis, only overall pooled sensitivity and specificity will be presented given that no significant heterogeneity was observed among included studies and there is only a limited number of data sets to conduct subgroup analysis.

Sensitivity

The overall pooled sensitivity of pooled testing with individuals as the unit of analysis is 97% (95% CI: 95-99), with point estimates of individual studies ranging from 92% to 100%. I^2 statistic was found to be 0% showing negligible heterogeneity between included studies.

Meanwhile, in the second phase of the local study of Lo et al. (2020) with individuals as the unit of analysis, the sensitivity of pooled testing in general was lower than the individual international studies and reported sensitivity estimates of 83%, 58%, and 50% for the two-stage (5-1), three-stage (10-5-1), and four-stage (20-10-5-1) Dorfman pooling.

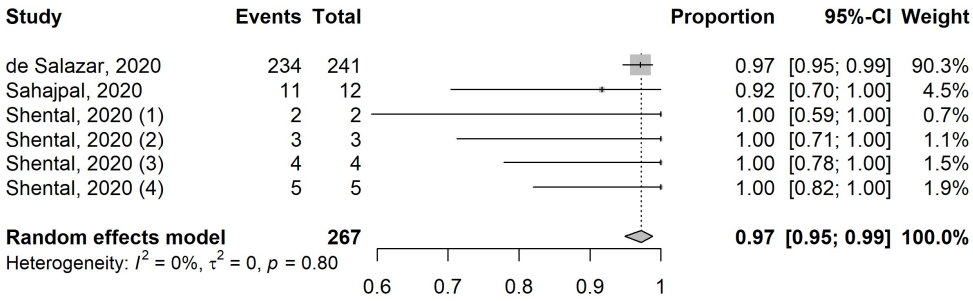

Figure 12. Sensitivity of Pooled Testing, Unit of Analysis: Individuals

Specificity

On the other hand, the overall pooled specificity of pooled testing with individuals as the unit of analysis was found to be 99.99% (95% CI: 99.97-100), with point estimates of individual studies ranging from 99.74% to 100%. Likewise, the I^2 statistic was found to be 0% showing negligible heterogeneity between included studies.

For the study of Lo et al. (2020) a high specificity was generally observed for the three variations in Dorfman pooling and specificity estimates were found to be consistent at 100%. This is similar to the results of the international individual studies which reported high specificity values.

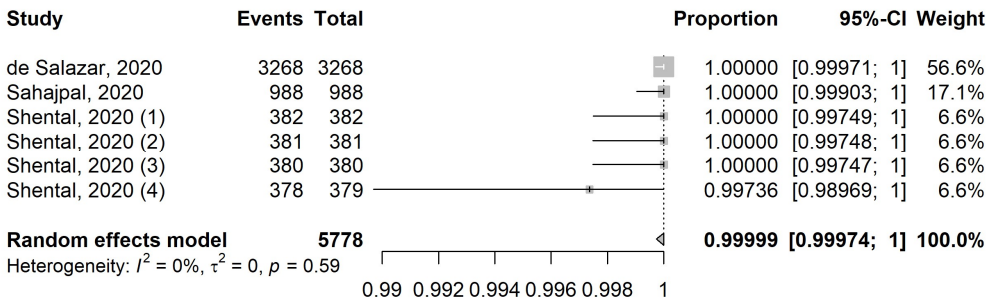

Figure 13. *Specificity of Pooled Testing, Unit of Analysis: Individuals*

PPV, NPV, +LRs, and -LRs are computed for each data set from each study. The point estimates for these measures can be seen in Appendix 2B.

Table 10. Summary of Performance Characteristics of Pooled Testing, Unit of Analysis: Pool

Subgroup title	No. of studies	No. of data sets	No. of pools	Diagnostic Performance Measure	Point estimate (95% CI)	I^2 statistic
SENSITIVITY						
Overall	15	41	1488	Sensitivity (95% C.I.)	87% (81-91)	71%
By pool size						
2	2	2	107	Sensitivity (95% C.I.)	98% (82-100)	38%
3	1	1	36	Sensitivity (95% C.I.)	100% (82-100)	N/A
4	4	5	209	Sensitivity (95% C.I.)	95% (89-98)	3%
5	6	7	243	Sensitivity (95% C.I.)	87% (82-91)	0%
6	2	2	107	Sensitivity (95% C.I.)	97% (50-100)	74%
8	4	5	148	Sensitivity (95% C.I.)	93% (82-98)	41%
10	7	7	374	Sensitivity (95% C.I.)	88% (74-95)	81%
12	1	1	7	Sensitivity (95% C.I.)	57% (23-86)	N/A
16	2	2	107	Sensitivity (95% C.I.)	86% (26-99)	90%
20	1	1	7	Sensitivity (95% C.I.)	57% (23-86)	N/A
24	1	1	7	Sensitivity (95% C.I.)	57% (23-86)	N/A
32	2	2	17	Sensitivity (95% C.I.)	75% (32-95)	54%
48	1	1	7	Sensitivity (95% C.I.)	43% (14-77)	N/A
50	2	4	112	Sensitivity (95% C.I.)	60% (38-79)	36%
By brand of index test						
2019-nCoV CDC RT-PCR	3	5	371	Sensitivity (95% C.I.)	86% (76-93)	74%
AgPathID RT-PCR	1	1	10	Sensitivity (95% C.I.)	90% (53-99)	N/A
BGI RT-PCR Kit	3	5	90	Sensitivity (95% C.I.)	91% (77-97)	35%
CDC-based Washington State EUA SARS-CoV-2 RT-PCR Assay	1	1	32	Sensitivity (95% C.I.)	94% (78-98)	N/A
Hologic Aptima SARS-CoV-2 TMA	1	1	35	Sensitivity (95% C.I.)	91% (77-97)	N/A
LightMix E Gene	1	1	22	Sensitivity (95% C.I.)	95% (74-99)	N/A
LightMix RdRP and E Gene	1	11	77	Sensitivity (95% C.I.)	63% (50-73)	0%
Mixed Brands	1	2	200	Sensitivity (95% C.I.)	79% (50-93)	92%
Non-nested STN COVID-19-RdRp/Hel	1	1	4	Sensitivity (95% C.I.)	25% (3-76)	N/A
PowerCheck 2019-nCoV	1	6	600	Sensitivity (95% C.I.)	98% (96-99)	13%
Roche Cobas	1	3	13	Sensitivity (95% C.I.)	91% (64-98)	0%

Subgroup title	No. of studies	No. of data sets	No. of pools	Diagnostic Performance Measure	Point estimate (95% CI)	I^2 statistic
Sansure SARS-CoV-2 Kit	1	2	26	Sensitivity (95% C.I.)	96% (79-99)	0%
STN-COVID-19-N	1	1	4	Sensitivity (95% C.I.)	50% (12-88)	N/A
STN COVID-19-RdRp/Hel	1	1	4	Sensitivity (95% C.I.)	50% (12-88)	N/A
By brand of test kit and pool size						
2019-nCoV CDC RT-PCR (<=10)	3	4	271	Sensitivity (95% C.I.)	89% (80-95)	56%
2019-nCoV CDC RT-PCR (>10)	1	1	100	Sensitivity (95% C.I.)	73% (63-81)	N/A
AgPathID RT-PCR (>10)	1	1	10	Sensitivity (95% C.I.)	90% (53-99)	N/A
BGI RT-PCR Kit (<=10)	3	4	90	Sensitivity (95% C.I.)	91% (77-97)	35%
CDC-based Washington State EUA SARS-CoV-2 RT-PCR Assay (<=10)	1	1	32	Sensitivity (95% C.I.)	94% (78-98)	N/A
Hologic Aptima SARS-CoV-2 TMA (<=10)	1	1	35	Sensitivity (95% C.I.)	91% (77-97)	N/A
Lightmix E gene (<=10)	1	1	22	Sensitivity (95% C.I.)	95% (74-99)	N/A
Lightmix RdRP and E gene (<=10)	1	4	35	Sensitivity (95% C.I.)	77% (58-89)	0%
Lightmix RdRP and E gene (>10)	1	6	42	Sensitivity (95% C.I.)	55% (40-69)	0%
Mixed Brands (<=10)	1	2	200	Sensitivity (95% C.I.)	79% (50-93)	92%
Non-nested STN COVID-19 RdRp/Hel (>10)	1	1	4	Sensitivity (95% C.I.)	25% (3-76)	N/A
PowerChek 2019-nCoV (<=10)	1	5	500	Sensitivity (95% C.I.)	99% (97-99)	0%
PowerChek 2019-nCoV (>10)	1	1	100	Sensitivity (95% C.I.)	96% (90-98)	N/A
Roche Cobas (<=10)	1	3	13	Sensitivity (95% C.I.)	91% (64-98)	0%
Sansure SARS-CoV-2 Kit (<=10)	1	2	26	Sensitivity (95% C.I.)	96% (79-99)	0%
STN COVID-19-N (>10)	1	1	4	Sensitivity (95% C.I.)	50% (12-88)	N/A
STN COVID-19-RdRp/Hel (>10)	1	1	4	Sensitivity (95% C.I.)	50% (12-88)	N/A
By Ct value						
<34	1	2	46	Sensitivity (95% C.I.)	98% (87-100)	0%
34-36	1	2	8	Sensitivity (95% C.I.)	90% (53-99)	0%
≥37	1	2	16	Sensitivity (95% C.I.)	73% (42-91)	19%
≤30	1	2	46	Sensitivity (95% C.I.)	97% (85-99)	0%
30-33	1	2	88	Sensitivity (95% C.I.)	89% (61-98)	77%
33-36	1	2	66	Sensitivity (95% C.I.)	48% (14-85)	91%
By use case						

Subgroup title	No. of studies	No. of data sets	No. of pools	Diagnostic Performance Measure	Point estimate (95% CI)	I^2 statistic
Diagnosis	4	5	100	Sensitivity (95% C.I.)	93% (79-98)	47%
Screening	6	14	745	Sensitivity (95% C.I.)	94% (87-97)	70%
Surveillance	2	4	332	Sensitivity (95% C.I.)	84% (74-90)	71%
Unspecified	4	18	311	Sensitivity (95% C.I.)	74% (63-82)	44%
By presence of symptoms						
Asymptomatic	2	2	9	Sensitivity (95% C.I.)	91% (56-99)	0%
Mixed (Symptomatic and Asymptomatic)	2	5	321	Sensitivity (95% C.I.)	83% (74-90)	60%
Symptomatic	1	1	4	Sensitivity (95% C.I.)	100% (33-99)	N/A
Unspecified	13	33	1154	Sensitivity (95% C.I.)	87% (80-92)	75%
By prevalence of symptoms						
4.80%	1	1	16	Sensitivity (95% C.I.)	75% (49-90)	N/A
9%	1	2	26	Sensitivity (95% C.I.)	96% (79-99)	0%
Unspecified	13	38	1446	Sensitivity (95% C.I.)	86% (81-91)	73%
By specimen						
Mixed (NP/OP)	9	28	1008	Sensitivity (95% C.I.)	88% (80-93)	74%
Nasopharyngeal	3	4	138	Sensitivity (95% C.I.)	94% (88-97)	0%
Saliva	1	2	26	Sensitivity (95% C.I.)	96% (79-99)	0%
Undefined	3	7	316	Sensitivity (95% C.I.)	76% (64-85)	63%
SPECIFICITY						
Overall	10	26	591	Specificity (95% C.I.)	99.6% (98.9-100)	0%

Table 11. Summary of Performance Characteristics of Pooled Testing, Unit of Analysis: Individuals

Subgroup title	No. of studies	No. of data sets	No. of participants	Diagnostic Performance Measure	Point estimate (95% CI)	I^2 statistic
SENSITIVITY						
Overall	3	6	267	Sensitivity (95% C.I.)	96% (92-98)	0%
SPECIFICITY						
Overall	3	6	5778	Specificity (95% C.I.)	99.99% (99.97-100)	0%

5.2.4. Characteristics of Ongoing Studies

Based on our search in clinicaltrials.gov and the WHO International Clinical Trials Registry Platform, we only found one validation study on pooled testing. The study (CTRI/2020/06/026005) is set to be conducted in India, enrolling samples of nasopharyngeal swabs from patients submitted for COVID testing as per their current testing guidelines. The study will utilize previously collected samples and will subject these to two modes of pooling, namely simple pooling and a combinatorial tapestry pooling based on a computerized algorithm. The results will then becompared to individual testing which will be considered as the gold standard. The primary outcomes are sensitivity and specificity of combinatorial tapestry pooling technique compared to individual testing in varying prevalence rates and degrees of pooling. On the other hand, secondary outcomes will include sensitivity and specificity of the simple pooled technique compared to individual testing across varying prevalence rates and degrees of pooling as well. Sensitivity and specificity of combinatorial tapestry pooling vs individual testing as well as cost for negative and positive tests across the three methods shall also be looked at. The target sample size is 321 patients. Currently, the trial is still not yet recruiting in India.

5.2.5. Critical Appraisal on the Included Studies

Of the nineteen studies that were critically appraised, eight (8) looked into screening as a use case for pooled testing, five (5) looked into the use of pooled testing in the diagnosis of COVID-19, two (2) studies explored the use of pooled testing in COVID-19 surveillance while four (4) studies did not mention a particular use case for pooled testing that was being explored by their study. Three (3) studies looked into the use of pooled testing in the screening of populations that are at risk of COVID-19, such as essential workers while one study that did not specify the specific use case. The local study on the other hand used pooled testing among employees from a supermarket chain.

In terms of the validity of the studies that were critically appraised, their internal validity would range from low to moderate, with eight (8) studies having low internal validity and eleven (11) having moderate internal validity. While most of the studies utilized an acceptable reference standard, common reasons for the low to moderate internal validity include the non-independence of the definition, performance and interpretation of the index test and reference standard. Considering the inherent design of pooled testing as an index test and comparing it with individual RT-PCR testing as the reference standard, the non-independence of the definition is understandable since most studies are likely to use the same criteria for interpretation since they used the same brand of PCR kit.

Most of the studies employed RT-PCR as the reference standard and generally have variations on experimental design (e.g., matrix pooling and ratio pooling) and sample size. The LRs of some of the 19 critically appraised studies cannot be computed because other values were not reported in their manuscripts (e.g., false positive and true negative), hence, other measures such as NPV and PPV cannot be obtained as well. Nevertheless, the majority of the positive and negative LRs calculated from the appraised studies, regardless of the variations due to sampling strategies, show moderate-strong performance in ruling in (confirmatory test) and/or ruling out (screening test) COVID-19 infection. Furthermore, it is important to note that some of the studies that presented very strong statistics on certain parameters (e.g., sensitivity, specificity, PPV and NPV are all 100%) have a small sample size and primarily have wide confidence intervals. Anent this, one study (Freire-Paspuel et al,

2020) had shown bias in the computation for sensitivity: instead of marking the one pooled and one individual sample as inconclusive, they were treated as positive and included in the computation. Overall, most of the appraised studies—from the results of the independent calculations—tend to agree that pooled testing is a useful confirmatory and screening test

The applicability could not be determined since all of the studies presented no relevant details on patient characteristics (some mentioned only symptomatic and asymptomatic), the severity of the disease and other biologic factors such as age and sex, let alone subgroup analysis. Therefore, it is difficult if not impossible to determine the applicability of pooled testing.

The full details of critical appraisal per study are provided in Appendix 4.

5.3. Testing Guideline Recommendations and HTA Evidence Reviews on Pooled Testing

5.3.1 Review of Testing Guideline Recommendations

Data source: Ministry of Health Website

Of the fourteen guidelines reviewed, only four (US CDC, ECDC, Public Health Ontario and Philippine DOH) have existing guidelines on pooled testing in their respective ministry/department of health websites. As for their recommended use cases, the US CDC recommendation has the widest scope of use which is for diagnosis, screening and surveillance. Meanwhile, both the ECDC and the Philippine DOH recommend its use only for screening or surveillance. Lastly, Public Health Ontario (PHO) recommends its use for diagnosis and surveillance. We also note that two guidelines (Philippine DOH and ECDC) explicitly stated that they do not recommend pooled testing for diagnosis.

Diagnosis

The US CDC and PHO currently implement pooled RT-PCR testing for use case diagnosis. The US CDC allows laboratories certified under CLIA to use specimen pooling strategy to expand SARS-CoV-2 diagnostic testing using test kits that are authorized for use in pooled testing by the US FDA. As a diagnostic tool, the US CDC currently recommends pooled testing among patients with symptoms or recent exposure or to determine the resolution of infection. Further, the US CDC recommends the use of a *two-stage specimen pooling strategy* in which samples are pooled together, and if a pooled test result is negative, then all specimens within the pool can be presumed negative with the single test. Whereas if a pooled test result is positive or indeterminate, all the specimens within the pool need to be retested individually. It is also recommended that pooling strategy depend on the prevalence of the virus in the community, and that prevalence be determined using a rolling average of the positivity rate of their own SARS-CoV-2 testing over the previous 7-10 days. Likewise, the US CDC also recommends the use of standardized methodology that factors in sensitivity of the assay used and the cost of testing to determine when the positivity rate is low enough to justify the implementation of a pooling strategy. However, it is important to note that there were no elaborations in the US CDC guideline on how prevalence will guide pooling strategy. We note that the US CDC applies these

recommendations not just for use case diagnosis but for screening and surveillance use cases as well. (CDC, 2020)

Meanwhile, PHO (Public Health Ontario) implemented a validated pooling approach to testing. They recommend the use of pooling on testing a portion of specimens submitted to its laboratory from assessment centers . PHO did not further elaborate on the target population nor characteristics of population being tested for diagnosis. Similar to the US CDC approach, PHO also employs a two-stage specimen pooling. (PHO, 2020)

We note that two guidelines explicitly did not recommend the use of pooled testing for diagnosis. The ECDC does not recommend the use of pooled testing in situations wherein diagnosis is critical due to the possibility of error. For diagnosis, they recommend the samples be retested separately if there is a positive result in the pool. (ECDC, 2020) The Philippine DOH also do not recommend the use of pooled testing in individuals with symptoms (regardless of severity), recovered patients, as well as close contacts of positive individuals.

Screening

Of the four guidelines which have included pooled testing, three (US CDC, ECDC and the Philippine DOH) are currently recommending it for screening purposes.

The US CDC likewise allows laboratories certified CLIA to use specimen pooling strategy to expand SARS-CoV-2 screening testing using test kits that are authorized for use in pooled testing by the US FDA. Further, the US CDC recommends the use of pooled testing as a screening test to identify occurrence at the individual level even if there is no reason to suspect infection. Screening tests intend to identify infected individuals without, or prior to development of symptoms who may be contagious so that measures can be taken to prevent further transmission. As mentioned above, the US CDC recommendations on the pooling strategy and other provisions provided in the previous section on diagnosis apply for this use case as well.

The ECDC, on the other hand, recommends the use of pooled testing to enhance testing of mild and asymptomatic patients.

Lastly, the Philippine DOH recommends the use of pooled testing for border testing at ports of entry for inbound foreign travelers and returning residents, locally-stranded individuals; as well as for departing and returning OFWs. In both screening and surveillance testing, the Philippine DOH recommends a pool sample of five, until an accurate prevalence of case with the presence of SARS-CoV-2 is identified in the populations. Their guidelines, however, emphasized that pooled testing strategies are currently being evaluated and validated; hence, the currently guidelines shall be amended as new developments ensue from the studies and pilot implementation.

Surveillance

Across the three use cases, surveillance is the most applied use case which are recommended by all four guidelines that have included pooled testing as part of their COVID-19 testing strategies.

The US CDC recommends the use of pooled testing as a surveillance test to monitor for a community or population level occurrence, such as infectious disease outbreak or to characterize the occurrence once detected, e.g. looking at incidence and prevalence of the occurrence. Surveillance testing is done to gain information at a population level, rather than individual level and may sample a certain percentage of a specific population to monitor for increasing or decreasing prevalence and to determine the population effect from community interventions. As mentioned in the regulatory section of this report, US FDA does not regulate tests for surveillance, but they recommend that If surveillance testing is performed by a non-CLIA certified laboratory, an individual who tests positive for SARS-CoV-2 should have a confirmatory test performed by a CLIA-certified laboratory. Surveillance with return of results and surveillance with pooled or batched testing should be validated on a test platform and test of high sensitivity and positive tests should have a confirmatory test. US CDC recommends that laboratories conducting surveillance testing with pooling should use an assay and test system that has received an EUA from FDA. Using an FDA-authorized assay and test system helps ensure the quality and reliability of testing.

The ECDC recommends the use of pooled testing in determining prevalence of disease in the community.

Meanwhile, the PHO recommends it in testing asymptomatic patients especially during outbreak investigations.

Lastly, the Philippine DOH recommends the use of pooled testing in surveillance of healthcare workers and all workers in health facilities, essential workers including market vendors, transport workers, frontline government workers and other economy workers. As mentioned above, the current guidelines recommend a pool sample of five until an accurate prevalence of case with the presence of SARS-CoV-2 is identified in the populations. Likewise, the guidelines are anticipated to be updated based on new studies and pilot implementation findings.

Data source: News Articles

To further retrieve information on the use of pooled testing among different countries specifically on pilot implementation efforts or plans which are not usually reported in guidelines, the reviewers conducted a targeted search for news articles from different ministries of health websites as well as independent news agencies. A limitation to this approach is the non-comprehensiveness of information gathered; however, the reviewers still found merit in this type of targeted search seeing as pooled testing is a new strategy, and as such, may not necessarily be included in the testing guidelines, or may still be in the pilot testing stages.

News articles from these countries are classified according to the different use cases being explored in this study.

- Of all the news articles screened, no country has used pooled testing in diagnosing COVID-19 infection.
- Below are countries exploring the use of pooled testing for screening:
  - United Kingdom is introducing the use of COVID-19 screening using pooled testing in universities to help prevent outbreaks and allow campuses to stay open. Independent SAGE (Scientific Advisory Group for Emergencies) recommended the testing of students and staff at the start of the academic year as well as regular

testing after, with University of Cambridge stating in their guidance that students living in college accommodations can be tested every week given their current capacity (Mahase, 2020).
    - Singapore allows the use of pooled testing in migrant worker dormitories and nursing facilities, with their Ministry of Health recommending its use in testing sub-populations with very low prevalence rates of COVID-19, or for mass screening purposes. (Singapore MOH, 2020a) (Singapore MOH, 2020b) (Sin.Y., 2020) (Sun.D., 2020) Singapore recommends using pool size of up to 5 specimens per test.
    - Malaysia recommends the use of pooled testing in mass testing groups with high risk of infection such as members of the Kuching church where the biggest clusters occurred due to mass gatherings. (Choong.J., 2020)
    - Korea recommends using pooled testing in local clusters at a higher risk of acquiring COVID-19 infection using pool sizes of 10. (Sung-sun, 2020)
    - Vietnam uses pooled testing on returnees from Da Nang population, which is considered to be the outbreak epicentre in Vietnam by using pool size of three to five individuals per laboratory test. (Kiet, 2020) (WHO, 2020)
- Lastly, the countries exploring the use of pooled testing for surveillance are as follows:
    - The Indonesian government plans to conduct pooled testing with pool sizes of five each in eight provinces that have been hardest hit by the coronavirus. A thousand samples will be taken from these provinces using a multi-step random sampling. (Sutrisno.B, 2020) They recommended that pool size be not more than 5 at the risk of getting false positive results; they also reiterated that pooled testing only be used in areas where cases are less prevalent, defined as having a positivity rate of less than 2%. (Kumar, 2020) (Sutrisno, B., 2020)
    - Thailand employs pooled testing using saliva samples to accelerate testing 100,000 persons in targeted groups including health and medical professionals, prison inmates, drivers for public buses and migrant workers. (WHO, 2020)
    - China used pooled testing on the entirety of Wuhan population as mass, indiscriminate testing, using pool sizes of five to ten individuals in one laboratory test (BBC, 2020)

We note that of the guidelines reviewed, the WHO and Australia do not have any guidelines regarding pooled testing, nor are there any articles which cites the use of this strategy. The compilation of the different testing guideline recommendation as well as relevant news article excerpts on the use of pooled testing for diagnosis, screening and surveillance from the relevant agencies and countries are presented in Appendix 3.

5.3.2 Review recommendations of HTA agencies

Of the ten (10) HTA agencies reviewed on their assessments and recommendations, none has published a review on available evidence and provided recommendations regarding pooled testing. Additionally, no information was found on existing and ongoing HTA reviews on pooled testing.

5.4. Resource Requirements

Among the nineteen studies that were reviewed, nine studies included information on the resources saved when using their proposed pooling strategy at a certain prevalence rate. Three additional studies were also obtained to give us more information on the reduction in resource requirements

when pooled testing strategies are implemented. The studies expressed resources saved in terms of total money amounts of costs, or in terms of specific resource requirements such as the number of tests, testing hours, and changes in human resources required.

Reduction in costs

Five studies relayed cost savings as an advantage of pooled testing over individual testing. It should be noted that not all declared whether their calculations included costs saved in compensation of personnel, cost saved in utilities, costs saved in test kits, and other costs saved due to the decrease in tests required when pooled testing is employed so their estimates can still vary. The pooled testing strategy done in Hainan, Country saved 49 620 USD in costs (Li et al, 2020). In the study by Singh et al (2020), individual testing of the 545 samples would have cost 3137 USD but their pooled testing strategy with testing of deconvoluted positive pools cost only 1002 USD. They saved 2135 USD or 68% savings in reagent costs. This does not include other costs that can be saved due to the decrease in tests performed. The cost analysis of Sahajpal et al (2020), claims that testing 1 million individuals would normally cost $58 million if tested individually but only 9.1 million USD if using their proposed mass screening by pooled testing. Wacharapluesadee et al (2020) calculated the costs saved when using their 10-sample pooling strategy for different prevalence rates. They report that a test that normally costs the laboratory 35 USD per patient can be discounted to 3.85 USD, 6.85 USD, 17.54 USD, and 26.30 USD for populations with a prevalence rate of 0.10%, 1.00%, 5.00%, and 10.00%, respectively. However, these costs savings are still based on the study's assumption that positive specimens are distributed evenly among pools. Lastly, Campbell et al (2020) calculated the cost savings when conducting 4-sample pooled testing across Canada over a period of 28 days and 42 days for testing in schools. Their calculations claim that testing symptomatic people and individuals with high risk of exposure will save them 15.5 million CAD, 17.3 million CAD for testing the same group and their contacts, 10.4 million CAD for testing employees of acute care hospitals, 13.6 million CAD for testing of community healthcare workers and in long-term care facilities, 35.4 million CAD for testing all essential employees with major public or interpersonal contact, and 83.1 million CAD in savings for testing all children and staff of schools.

Reduction in number of tests

Savings in terms of number of tests saved when compared to individual testing were reported by ten studies. In the study of Khodare et. al. (2020), they reported the efficiency of their sample pooling method as 0.38 (calculated using the online calculator considering the two-stage Dorfman mini-pool strategy of pool testing with the conservative predictions of a sensitivity of 95%, a specificity of 99%, 4% prevalence of SARS-CoV-2 infection and pool size of 6 samples). This means that for one sample testing, only 0.38 test reagents are required (savings of tests required by 62%). Mastrianni et al (2020) were able to report a 64% (340 out of 530) savings on tests when they conducted pooled testing to 530 samples from a population with 0.8% positivity rate. Schmidt et al (2020) reported that by applying their pooled testing protocol of pooling 100 samples into 10 pools, they were able to reduce the number of required tests by up to 80%, without loss of diagnostic sensitivity. Meanwhile, a reduction from 10 extraction and PCR runs (96-well plate format) to 2 runs (80% savings) for testing the status of 940 samples with a positivity rate of 0.6% was observed in the study of Sahajapal et.al (2020), as they performed pooled testing. In the study of de Salazar et. al. (2020), where they tested 3519 samples from a population with 6.86% prevalence, they were

able to save 2167 PCR tests (86%) by combining 9-10 samples per pool. Shental et. al. (2020), on the other hand, observed an eightfold increase in testing efficiency (80% savings) by using 48 pools to simultaneously test 384 subjects from a population with 1.3%. carrier rate. As for the study of Li et. al. (2020), they were able to save 827 (87.6%) tests in testing 944 specimens by using 10:1 pools. Ben-Ami, et. al. (2020) performed pooled testing of samples from a population with 0.23% positivity rate in three batches, and consequently observed 86.4%, 87.5% and 83.1% savings test saved. Lastly, the study by Campbell et. al. (2020) discussed the change in number of laboratory tests run per day when conducting pooled testing (4 samples in 1 pool) of high-risk and at-risk groups as per their calculations, based on a positivity rate of 0.84% at the time when the study was conducted. The study presented up to 71-74% savings on tests run per day.

In general, these studies show that given low positivity rates, using a pooled testing strategy may reduce the number of tests to be conducted individually from 62 – 87%, consequently increasing the capacity of testing for COVID –19.

A local study conducted by Lo et. al (2020) compared different Dorfman pooling methods and reported savings in terms of number of test kits saved from 69% to 83%, for a population with a positivity rate of 3%. The savings were computed by dividing the number of tests saved and the number of samples which is also the number of individual tests that will be used. The Dorfman 5-1 method having the least savings was reported to have used 138 test kits for 440 samples in turn saving 302 test kits (69% savings). The Dorfman 20-10-5-1 method was reported to use only 76 test kits for 440 samples translating to 83% savings, showing an increase in savings with larger pool sizes. The number of tests saved per method can be seen in Table 12. It is to note however that in these calculations, the number of tests included were only the test kits consumed except for repeat tests and those used for quality control. A more complete calculation would have included these values.

Table 12. Test Savings (Lo et.al, 2020)

	Prevalence	**Tests saved**
Dorfman5-1	**3%**	**302 out of 440 (69%)**
Site A	4%	146 out of 220 (66%)
Site B	2%	156 out of 220 (71%)
Dorfman10-5-1	**3%**	**347 out of 440 (79%)**
Site A	4%	177 out of 220 (80%)
Site B	2%	170 out of 220 (77%)
Dorfman20-10-5-1	**3%**	**364 out of 440 (83%)**
Site A	4%	191 out of 220 (87%)
Site B	2%	173 out of 220 (79%)

Notes: Controls not yet included in calculation of test savings

The same study also presented a calculation of average test savings per batches of 100 specimens, calculated using 10, 000 simulations, as seen in Table 13. The study reported that savings is a function of positivity rate and savings will decrease as positivity/prevalence rates increase (Lo et.al, 2020). In the table, we see that even at a prevalence of 20%, we see savings in the 2-stage Dorfman pooling for pool size of 10. Pooled testing using a pool size of 20 when used in settings with prevalence of 14% would have no savings at all and loss in savings were observed with prevalence of 15% and above.

Table 13. Average Test Savings per Batches of 100 Specimens, Calculated Using 10,000 Simulations*

	2- stage Dorfman			3 stage Dorfman		
	Pool size			Pool sizes		
Prevalence	**5**	**10**	**20**	**10&5**	**20&10**	**25&5**
1%	75	80	77	83	84	87
2%	71	72	62	77	73	78
3%	66	64	49	71	64	71
4%	61	56	39	65	56	65
5%	57	50	31	59	48	59
6%	53	44	24	54	42	54
7%	49	38	19	49	36	49
8%	46	34	14	44	31	44
9%	42	29	10	40	26	40
10%	39	25	7	36	21	47
11%	36	21	5	32	17	33
12%	33	18	3	38	14	29
13%	30	15	1	25	10	26
14%	27	12	0	21	8	24
15%	24	9	-1	18	5	21
16%	22	7	-2	15	3	18
17%	20	6	-3	13	1	15
18%	17	4	-3	10	-1	13
19%	15	2	-4	7	-3	11
20%	13	1	-4	5	-4	9

* 10, 000 batches of 100 specimens each were simulated using the R programming language for Statistical Analysis

Reduction in testing hours

Five studies reported a reduction in testing hours when they employed their pooled testing strategy versus individual PCR testing. In the study by Li et. al. (2020), they were able to analyze 944 samples using 117 tests, saving the staff three hours of hands-on time. Campbell et. al. (2020), concluded in their study that for the Canadian population with a positivity rate of 0.84%, pooling samples by 4 can reduce laboratory technician time by 70%. On the other hand, Sahajpal et. al. (2020) determined the status of 940 samples by testing 144 pools of 10, including the second-stage testing for positive pools. Positivity rate was 0.6%. According to the study, testing 940 samples individually would have taken at least 30 hours of processing time but their pooling strategy only required 6 hours, including the time spent on testing the samples of positive pools individually. In the study conducted in India by Singh et. al. (2020), for a population with a point prevalence of 4.8%, 545 samples were grouped into 109 pools with 5 samples each. This pooling strategy reduced laboratory processing time from the expected 67 hours to 23 hours, including the individual testing of samples from deconvoluted positive pools. The four studies show that pooled testing, when done using a strategic pool size among a population with low prevalence, will save the laboratory hours of processing time even when positive pools have to be deconvoluted and tested individually. Lastly, in the study by Ben-Ami et. al. (2020), the testing hours saved was expressed in the form of an increase in throughput. The study claimed that in their population with a positivity rate of 0.12%,

testing pools of 8 samples will result in a 7.3-fold increase in throughput, allowing more tests within a certain period of time.

Compared to the foreign studies that reported reduction in testing hours, the local study by Lo et al (2020) reported delays in turnaround times in pooled testing. The study defined turnaround time as the number of batch runs required to release the results of a sample. The turnaround time measured in the study is based only on the actual testing hours per run and does not include downtimes and waiting times. The study reported delays in turnaround time for both positive and negative samples when compared to the turnaround time of 1 batch run for samples subjected to individual testing.

Positive samples required 2 batch runs for the Dorfman 5-1 pooling strategy, 3 batch runs for Dorfman 10-5-1, and 4 batch runs for Dorfman 20-10-5-1, with the turnaround time being the number of stages in the pooling methodology. However, these delays in turnaround times are expected because pooled testing entails an additional batch run for testing each sample of positive pools. Because the study employed multiple-stage Dorfman pooling strategies, each stage, then, required one batch run comprised of the positive samples from the previous run. There was also a reported delay in turnaround time for negative samples. Negative samples can be released after an average of 1.09 batch runs, 1.21 batch runs, and 1.44 batch runs for Dorfman 5-1, 10-5-1 and 20-10-5-1, respectively. Based on the given data, we found that the turnaround time for negative samples was calculated using the equation,

$$\frac{\sum_{i=1}^{m} iX_i}{x}$$

where m is the number of stages of pooled testing, i is the stage number, X_i is the number of negative samples in the ith stage of pooled testing, and x is the total number of negative samples in the whole run. Lo et al. (2020) also acknowledged that although they found a slight delay in turnaround time, this can still be compensated by the increase in number of patient samples tested per day. The reported increases in turnaround time can be seen in Table 14.

Table 14. Reported Turnaround Times after Pooled Testing (Lo et.al, 2020)

	Prevalence	**Turnaround time (TAT)**	
		Positives (batch runs)	**Negatives (batch runs)**
Dorfman5-1	**3%**	**2**	**1.09**
Site A	4%	2	1.11
Site B	2%	2	1.07
Dorfman10-5-1	**3%**	**3**	**1.21**
Site A	4%	3	1.18
Site B	2%	3	1.24
Dorfman20-10-5-1	**3%**	**4**	**1.44**
Site A	4%	4	1.29
Site B	2%	4	1.59

Reduction in human resources

Among the 12 studies that reported on resource savings, only the study by Campbell et. al. calculated the changes in personnel required when pooled testing is employed. The study deemed that a 4-sample pooling strategy will reduce the number of laboratory staff required but increase the number of clerical staff needed. Calculations from the study claim that in Canada, the pooling strategy will require 473 less laboratory staff and 13 more clerks when testing symptomatic people and individuals with high risk of exposure, 525 less laboratory staff and 19 more clerks when testing the same group and their contacts, 322 less laboratory staff and 3 more clerks when testing all employees of acute care hospitals, 423 less laboratory staff and 3 more clerks when testing community healthcare workers and in long-term care facilities, 1100 less laboratory staff and 3 more clerks when testing all essential employees with major public or interpersonal contact and 1718 less laboratory staff and 2 more clerks are required when testing all children and staff of schools. These calculations are all based on the daily human resource needs for active testing strategies in Canada, with a positivity rate of 0.84% at the time when the study was conducted. Data used for the calculations were the number of SARS-CoV-2 testing sites and the estimated number of tests performed per day as of 17 July 2020. The researchers also referred to data from Statistics Canada to obtain the number of acute care hospitals and long-term care facilities and the number of employees and residents for each. American data was used for the number of community healthcare workers and assumed that the number is proportional to the population of Canada. For calculations for essential employees with major public of interpersonal contact, the researchers used census data from 2016 and adjusted to June 2020 labor force size and classified essential work as those not able to be performed at home. The number of primary and secondary schools and the number of students and employees were obtained from Statistics Canada and provincial reports.

6. LIMITATIONS

This review recognizes the following limitations: First, as this is a rapid review, certain steps of a systematic review were abbreviated such as searching through other search databases. Second, while a bivariate model was ideal in pooling accuracy measures from primary diagnostic accuracy studies, not all reported data on specificity, hence, a univariate approach to pooling was conducted. Lastly, as research on the different facets of COVID-19 is on-going and rapidly evolving, the evidence presented here can rapidly change as well. Hence, updating of evidence would be necessary.

7. CONCLUSION

REGULATORY STANDARDS FOR COVID-19 POOLED TESTING

This review found only one regulatory agency (US FDA) with available information on the regulation of pooled testing for COVID-19. As such, a comparison of performance standards and validation requirements cannot be made. The US FDA being the regulatory body for in vitro diagnostic devices requires the use of EUA authorized COVID-19 assays and test systems for pooled testing whether used as a diagnostic or as a screening test. The US FDA explicitly stated that it does not generally regulate the use of a test for surveillance purposes but requires that if surveillance testing is performed by a non- CLIA-certified laboratory, a confirmatory test on the detected positive individuals should be performed by a CLIA-certified laboratory.

The US FDA currently lists a total of ten molecular diagnostic tests that can be used for pooled testing. Of these, one brand is currently authorized by the Philippine FDA as a PCR based test kit for COVID-19 but for individual testing and not specifically for pooling testing as regulatory standards for the latter does not exist yet in the Philippines. As of now, the US FDA identifies two approaches to specimen pooling: 1) sample/media pooling (n-sample pooling) 2) swab pooling.

What are the validation testing requirements of selected regulatory agencies for COVID-19 pooled testing?

For the validation testing requirements, the US FDA requires that the manufacturer submit an EUA request for pooled testing and conduct clinical validation studies for the proposed pooling strategy in the intended use population. The US FDA believes that a pool size of 5 is a reasonable starting point for validation of the test for pooled testing. However, they still strongly encourage the test developers to choose their pool size in consideration of the positivity rate and percent of weak positive patients (as defined by the US FDA as patients with Ct values close to the assay's limit of detection) in their intended test population and sensitivity of their RT-PCR test. Tests validated for a pool size of n are also authorized and validated for pool sizes lower than n. The US FDA has issued separate guidelines for the validation of pooled testing by sample/media pooling and by swab pooling. For both methods of sampling, the test developer must establish the performance of the test kit when used for pooled testing compared to individual testing. The US FDA also requires two additional studies for the validation of swab pooling to evaluate the performance of the test kit when inhibitions due to (1) high concentrations of swab specimen (e.g. mucin) and (2) high concentrations of viral load are observed.

What are the performance standards used by selected regulatory agencies for the approval of COVID-19 pooled testing for market entry?

For the clinical validation study to establish the performance of pooled testing, the US FDA requires a PPA of at least 85% and that the results must demonstrate that weak positive samples can still be detected when pooled with other samples. For the study that evaluates performance against the interference of swab specimens, the US FDA requires 95% agreement with expected results and an invalid rate of less than 5%. As with the interference of high viral load, all replicates must test as positive or have an invalid rate of ≤5%.

DIAGNOSTIC PERFORMANCE

What is the accuracy of pooled testing in the diagnosis of COVID-19 as compared to individual testing?

- The sensitivity of pooled testing greatly varies when the unit of analysis is by pool (sensitivity point estimates ranging from 25 –100%) where in the number of pools tested is small, compared to when the analysis is by individual (sensitivity point estimates ranging from 92 –100%). The pooled sensitivity of pooled testing analyzed by pool was found to be 87%, (95% CI: 81-91, I^2=71%) while pooled sensitivity analyzed by individual was 97% (95% CI: 95-99, I^2=0%). The pooled sensitivity analyzed by individual suggests that pooled testing has a good rate of correctly identifying COVID-19 positive individuals and consequently, a low rate of false negatives. Meanwhile, the pooled sensitivity analyzed by pools should be interpreted with caution due to high heterogeneity.
- On the other hand, the specificity of pooled testing was consistently high ranging from 97% to 100%. The pooled specificity of pooled testing analyzed by pool [98.9% (95% CI: 89-96, I^2=12%)]

and analyzed by individual [99.99%, CI: 98.9-100, I^2=0%)] implies that the test has a very low rate of both false positive pools and both false positive individuals.

- Given the substantial heterogeneity present in the pooling of sensitivity, we performed subgroup analysis to assess the impact of several factors (e.g., pool size, brand of index test, CT value threshold, presence of symptoms, onset of symptoms, use case, specimen, and prevalence of disease) that may have served as sources of heterogeneity across the studies. The results from the subgroup analysis suggest that:
    - Based on low to moderate quality of evidence from 18 studies, pooled testing showed higher sensitivity estimates when pool sizes used are smaller. This finding was very intuitive because the more specimens in the same pool, the more diluted the shared reagent becomes losing the power of the test.
    - Based on low to moderate quality of evidence from 18 studies the sensitivity varies greatly from one brand to another possibly because of the different processes and protocols each one takes, as well as the different criteria set by their respective manufacturers. In addition, we observed that three brands that were tested in pool sizes of less than or equal to 10 had higher sensitivity as compared to the same brand when tested using greater than 10 samples. However, caution must be taken in the interpretation given that there were only few studies available for each brand.
    - Based on low and moderate quality evidence from 2 studies, an increase in the CT value of positive samples in a pool decreases the sensitivity estimate. However, this needs further investigation since the number of studies analyzed for this variable is small.
    - The sensitivity of pooled testing for screening (6 studies, low to moderate validity) was the highest but was closely followed by those that indicated diagnosis (4 studies, low to moderate validity) as its use case. Though, both use cases were found to have substantial variation within the included studies.
    - Sensitivity estimates for symptomatic individuals, although based only on one moderate quality study, were higher compared to the asymptomatic (2 studies, low and moderate validity) and unspecified population (13 studies, low to moderate validity). Caution must also be taken in the interpretation of this analysis since very few studies had available information regarding the clinical characteristics of the sample.
    - Pooled testing that used saliva had the highest sensitivity but was comparable to those that collected using nasopharyngeal specimen. However, it should be noted that only one moderate quality study used saliva as its specimen. Those that used mixed samples of nasal and oral specimen also had an acceptable sensitivity.
    - Few studies indicated the prevalence of the disease where the sample was obtained so our numerical results for studies with low prevalence still warrant further investigation.
    - The Philippine study by Lo et al (2020) had a lower sensitivity in both units of analysis than the overall pooled sensitivity estimate of our quantitative synthesis. The reported sensitivity estimates of the study for analysis by pools were 83% (95% CI: 67-94), 72% (95% CI: 55-86), and 67% (95% CIL 49-81) while the sensitivity estimate for analysis by individuals were reported to be at 83% (95% CI: 52-98), 58% (95% CI: 28-85), and 50% (95% CI: 21-79). In contrast, the specificity of the local study with individuals as unit of analysis (100%, 95% CI: 99-100) is comparable to the overall specificity estimate of the included studies in meta-analysis. Furthermore, the study of Lo et al showed consistent trend with the observations in the subgroup analysis which tells that lower pool sizes were seen to have higher sensitivity estimates.
- As for the quality of these studies on diagnostic performance, our critical appraisal shows that eight studies had high risk of bias and eleven had moderate risk of bias. Some factors that have affected the validity of these studies include non-independence of the definition, performance, and interpretation of the index and reference test.

- Based on low to moderate quality evidence, the use of pooled testing for COVID-19 shows high specificity but varied sensitivity. Further, the prevalence of the population to which pooled testing can be applied remains unclear.

GUIDELINE RECOMMENDATION AND EVIDENCE SYNTHESIS FROM HTA AGENCIES ON USE OF POOLED TESTING

Which countries have implemented testing strategies using pooled testing for diagnosis, screening and surveillance of COVID-19?

Of the fourteen guidelines reviewed, only four (US CDC, ECDC, Public Health Ontario and Philippine Department of Health) have existing guidelines on pooled testing for COVID-19.

Guidelines Recommending RT-PCR pooled testing	Recommended Use Case		
	Diagnosis	Screening	Surveillance
US CDC			
ECDC			
PHO			
Philippine DOH			

To further retrieve information on the use of pooled testing among different countries specifically on pilot implementation efforts or plans which are not usually reported in guidelines, the reviewers conducted a targeted search for news articles from different ministries of health websites as well as independent news agencies.

News Articles Describing Pooled Testing Pilot Implementation	Recommended Use Case		
	Diagnosis	Screening	Surveillance
China			
Indonesia			
South Korea			
Malaysia			
Singapore			
Thailand			
United Kingdom			
Vietnam			

Diagnosis

Two guidelines (US CDC and PHO) currently recommend the use of pooled testing for diagnosis of COVID-19 infection. As a diagnostic tool, the US CDC currently recommends pooled testing among patients with symptoms or recent exposure or to determine the resolution of infection, whereas PHO uses pooled testing a portion of specimens submitted to its laboratory from assessment centers . PHO did not further elaborate on the target population nor characteristics of population being tested for diagnosis.

In both cases, they recommend using a two-stage specimen pooling strategy in which samples are pooled together, and if a pooled test result is negative, then all specimens within the pool can be presumed negative with the single test; whereas if a pooled test result is positive or indeterminate, all the specimens within the pool need to be retested individually. Meanwhile,

ECDC and the Philippine DOH explicitly stated in their guidelines that they do not recommend the use of pooled testing for diagnosis. The Philippine DOH do not recommend it for certain populations which include individuals with symptoms (regardless of severity), recovered patients and close contacts of positive individuals, while ECDC does not recommend it due to the possibility of error.

Screening

Three guidelines (US CDC, ECDC, and Philippine DOH) currently recommend the use of pooled testing in screening infected individuals without, or prior to development of symptoms who may be contagious so that measures can be taken to prevent further transmission. The US CDC does not specify the target population of this use case, while the ECDC specifically recommends it for mild and asymptomatic patients. The Philippine DOH recommends the use of pooled testing in screening the following populations: inbound travellers, overseas filipino workers (deployment and returning), and locally stranded individuals.

In addition to testing guidelines, several news articles from countries like United Kingdom, Singapore, Malaysia, Korea and Vietnam have mentioned the use of pooled testing as screening tests in different settings and populations. United Kingdom is introducing the use of COVID-19 screening using pooled testing in universities to help prevent outbreaks and allow campuses to stay open (Mahase, 2020). Singapore on the other hand, allows the use of pooled testing in migrant worker dormitories and nursing facilities, with their ministry of health recommending its use in testing sub-populations with very low prevalence rates of COVID-19, or for mass screening purposes. (Singapore MOH, 2020a) (Singapore MOH, 2020b) (Sin.Y., 2020) (Sun.D., 2020) Malaysia recommends the use of pooled testing in mass testing groups with high risk of infection such as members of the Kuching church where the biggest clusters occurred due to mass gatherings. (Choong.J., 2020) Like Malaysia, Korea also recommends using pooled testing in local clusters at a higher risk of acquiring COVID-19 infection using pool sizes of 10 (Sung-sun, 2020) Vietnam uses pooled testing on returnees from Da Nang population, which is considered to be the outbreak epicentre in Vietnam by using pool size of three to five individuals per laboratory test. (Kiet, 2020) (WHO, 2020)

Surveillance

Across the three use cases, pooled testing is most commonly used for surveillance purposes, based on the guidelines reviewed.

All four guidelines (US CDC, ECDC, PHO, and Philippine DOH) recommend the use of pooled testing for surveillance. The US CDC recommends the use of pooled testing to monitor a community of population level occurrence, such as an infectious disease outbreak or to characterize the occurrence once detected, as well as to look at incidence and prevalence of occurrence. The ECDC recommends the use of pooled testing in determining prevalence of disease in the community or to enhance testing of mild and asymptomatic patients; while PHO recommends it in testing asymptomatic patients especially during outbreak investigations. The Philippine DOH recommends the use of pooled testing in the surveillance of the following populations: healthcare workers and all workers in health facilities, essential workers including market vendors, transport workers, frontline government workers and other economy workers.

In addition to testing guidelines, several news articles from countries like Indonesia, Thailand and China have mentioned the use of pooled testing for surveillance under different circumstances. The Indonesian government plans to conduct pooled testing with pool sizes of five each in eight provinces that have been hardest hit by the coronavirus. A thousand samples

will be taken from these provinces using a multi-step random sampling. (Sutrisno.B, 2020) Thailand also employs pooled testing using saliva samples to accelerate testing 100,000 persons in targeted groups including health and medical professionals, prison inmates, drivers for public buses and migrant workers. (WHO, 2020) China used pooled testing on the entirety of Wuhan population as mass, indiscriminate testing, using pool sizes of five to ten individuals in one laboratory test (BBC, 2020)

We note that of the guidelines reviewed, the WHO and Australia do not have any guidelines regarding pooled testing, nor are there any articles which cites the use of this strategy.

What is the current position of HTA agencies regarding the use of pooled testing for COVID-19?

Currently, no HTA agencies have published a review on the use of pooled testing in COVID-19 nor are there any ongoing studies by HTA agencies on the use of pooled. Hence, the current position of HTA agencies on the use of pooled testing in COVID-19 remains unclear.

RESOURCE REQUIREMENTS

What are the resource requirements needed to use pooled testing for COVID-19?

In terms of resource requirements, we had originally planned to look at the resources needed to implement pooled testing however, the included studies reported the resources saved instead. In summary, studies reported savings in terms of money amounts of cost or in specific reductions in number of test kits, testing hours, and personnel required when pooled testing was used. Values were reported in consideration of the pool size and the positivity rate of the study. The included studies have reported positivity rates ranging from 0.12-3% or prevalence of 4-6.86%. In addition, the studies also used a variety of pool sizes ranging from 3-20 samples.

8. DECLARATION OF CONFLICT OF INTERESTS

The reviewers declare that they have no competing interests.

9. REFERENCES

1. Bateman, A. C., Mueller, S., Guenther, K., & Shult, P. (2020). Assessing the dilution effect of specimen pooling on the sensitivity of SARS-CoV-2 PCR tests. J Med Virol. doi:10.1002/jmv.26519.

2. Ben-Ami, R., Klochendler, A., Seidel, M., Sido, T., Gurel-Gurevich, O., Yassour, M., . . . Drier, Y. (2020). Large-scale implementation of pooled RNA extraction and RT-PCR for SARS-CoV-2 detection. Clin Microbiol Infect, 26(9), 1248-1253. doi:10.1016/j.cmi.2020.06.009.

3. Campbell, J. R., Uppal, A., Oxlade, O., Fregonese, F., Bastos, M. L., Lan, Z., . . . Menzies, D. (2020). Active testing of groups at increased risk of acquiring SARS-CoV-2 in Canada: costs and human resource needs. CMAJ 2020. doi:10.1503/cmaj.201128.

4. Center for Disease Controls (2020). Interim Guidance for Use of Pooling Procedures in SARS-CoV-2 Diagnostic, Screening and Surveillance testing Retrieved Nov. 3, 2020 from: https://www.cdc.gov/coronavirus/2019-ncov/lab/pooling-procedures.html#:~:text=What%20is%20pooling%3F,virus%20that%20causes%20COVID%2D19.

5. de Salazar, A., Aguilera, A., Trastoy, R., Fuentes, A., Alados, J. C., Causse, M., . . . García, F. (2020). Sample pooling for SARS-COV-2 RT-PCR screening. Clin Microbiol Infect. doi:10.1016/j.cmi.2020.09.008.

6. Department of Health (DOH). (2020). COVID-19 Tracker. Retrieved October 16, 2020 from https://www.doh.gov.ph/covid19tracker

7. Dong, E., Du, H., Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. Retrieved last October 16, 2020. DOI:https://doi.org/10.1016/S1473-3099(20)30120-1

8. Freire-Paspuel, B., Vega-Mariño, P., Velez, A., Cruz, M., & Garcia-Bereguiain, M. A. (2020). "Sample pooling of RNA extracts to speed up SARS-CoV-2 diagnosis using CDC FDA EUA RT-qPCR kit". Virus Res, 290, 198173. doi:10.1016/j.virusres.2020.198173.

9. Gupta, E., Padhi, A., Khodare, A., Agarwal, R., Ramachandran, K., Mehta, V., . . . Sarin, S. K. (2020). Pooled RNA sample reverse transcriptase real time PCR assay for SARS CoV-2 infection: A reliable, faster and economical method. PLoS One, 15(7), e0236859. doi:10.1371/journal.pone.0236859.

10. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (Eds). (2017). Cochrane Handbook for Systematic Reviews of Interventions version 5.2 (updated June 2017). Retrieved June 9, 2020, from www.training.cochrane.org/handbook

11. Khodare, A., Padhi, A., Gupta, E., Agarwal, R., Dubey, S., & Sarin, S. K. (2020). Optimal size of sample pooling for RNA pool testing: An avant-garde for scaling up severe acute respiratory syndrome coronavirus-2 testing. Indian J Med Microbiol, 38(1), 18-23. doi:10.4103/ijmm.IJMM_20_260.

12. Kim, S. Y., Lee, J., Sung, H., Lee, H., Han, M. G., Yoo, C. K., . . . Hong, K. H. (2020). Pooling Upper Respiratory Specimens for Rapid Mass Screening of COVID-19 by Real-Time RT-PCR. Emerg Infect Dis, 26(10), 2469-2472. doi:10.3201/eid2610.201955.
13. Li, H., Sun, K., Persing, D. H., Tang, Y. W., & Shen, D. (2020). Real-time Screening of Specimen Pools for Coronavirus Disease 2019 (COVID-19) Infection at Sanya Airport, Hainan Island, China. Clin Infect Dis. doi:10.1093/cid/ciaa1074.

14. Lo, R., Barrientos, A., Espiritu, B., Santiago, F., Tandoc III, A., Veloso, J., & Yañez, S. (2020). An Evaluation of Pooling Strategies for qRT-PCR Testing for SARS-CoV-2 Infection. Manuscript sumbitted for publication.

15. Mastrianni, D., Falivena, R., Brooks, T., McDermott, B., Tan, J., Vandell, R., & Holland, M. (2020). Pooled Testing for SARS-CoV-2 in Hospitalized Patients. J Hosp Med, 15, 538-539. doi:10.12788/jhm.3501.

16. Mitchell, S. L., & Ventura, S. E. (2020). Evaluation and Comparison of the Hologic Aptima SARS-CoV-2 and the CDC 2019 nCoV real-time RT-PCR Diagnostic Panel using a Four-Sample Pooling Approach. J Clin Microbiol. doi:10.1128/JCM.02241-20.

17. Pasomsub, E., Watcharananan, S. P., Watthanachockchai, T., Rakmanee, K., Tassaneetrithep, B., Kiertiburanakul, S., & Phuphuakrat, A. (2020). Saliva sample pooling for the detection of SARS-CoV-2. J Med Virol. doi:10.1002/jmv.26460.

18. Perchetti, G. A., Sullivan, K. W., Pepper, G., Huang, M. L., Breit, N., Mathias, P., . . . Greninger, A. L. (2020). Pooling of SARS-CoV-2 samples to increase molecular testing throughput. J Clin Virol, 131, 104570. doi:10.1016/j.jcv.2020.104570. Epub 2020 Aug 2.

19. Philippine Society for Microbiology and Infectious Diseases. (2020). Interim Guidelines on the Clinical Management of Adult Patients with Suspected or Confirmed COVID-19 Infection. Retrieved April 16, 2020 from https://www.dropbox.com/s/hja8h49q8k9b77r/PSMID%20COVID%20tx%20guidelines%20v.3.31.20a.pdf?dl=0

20. Praharaj, I., Jain, A., Singh, M., Balakrishnan, A., Dhodapkar, R., Borkakoty, B., . . . Gupta, N. (2020). Pooled testing for COVID-19 diagnosis by real-time RT-PCR: A multi-site comparative evaluation of 5- & 10-sample pooling. Indian J Med Res, 152(1), 88-94. doi:10.4103/ijmr.IJMR_2304_20.

21. Sahajpal, N. S., Mondal, A. K., Njau, A., Ananth, S., Jones, K., Ahluwalia, P. K., . . . Kolhe, R. (2020). Proposal of RT-PCR-Based Mass Population Screening for Severe Acute Respiratory Syndrome Coronavirus 2 (Coronavirus Disease 2019). J Mol Diagn, 22(10), 1294-1299. doi:10.1016/j.jmoldx.2020.07.001. Epub 2020 Jul 30.

22. Schmidt, M., Hoehl, S., Berger, A., Zeichhardt, H., Hourfar, K., Ciesek, S., & Seifried, E. (2020). Novel multiple swab method enables high efficiency in SARS-CoV-2 screenings without loss of sensitivity for screening of a complete population. Transfusion. doi:10.1111/trf.15973.

23. Schwarzer, G., Chemaitelly, H., Abu-Raddad, L. J., & Rücker, G. (2019). Seriously misleading results using inverse of Freeman-Tukey double arcsine transformation in meta-analysis of single proportions. Research synthesis methods, 10(3), 476–483. https://doi.org/10.1002/jrsm.1348

24. Shental, N., Levy, S., Wuvshet, V., Skorniakov, S., Shalem, B., Ottolenghi, A., . . . Hertz, T. (2020). Efficient high-throughput SARS-CoV-2 testing to detect asymptomatic carriers. Sci Adv, 6(37). doi:10.1126/sciadv.abc5961. Print 2020 Sep.

25. Singh, A. K., Nema, R. K., Joshi, A., Shankar, P., Nema, S., Raghuwanshi, A., . . . Singh, S. (2020). Evaluation of pooled sample analysis strategy in expediting case detection in areas with emerging outbreaks of COVID-19: A pilot study. PLoS One, 15(9), e0239492. doi:10.1371/journal.pone.0239492. eCollection 2020.

26. Trikalinos TA, Trow P, & Schmid CH. (2013). Simulation-Based Comparison of Methods for Meta-Analysis of Proportions and Rates [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US). Available from: https://www.ncbi.nlm.nih.gov/sites/books/NBK179166/

27. United States Food and Drug Administration. (2020, August 24). Pooled Sample Testing and Screening Testing for COVID-19. Retrieved October 12, 2020, from https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/pooled-sample-testing-and-screening-testing-covid-19

28. United States Food and Drug Administration. (2020, October 20). FAQs on Testing for SARS-CoV-2. Retrieved October 12, 2020, from https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/faqs-testing-sars-cov-2

29. United States Food and Drug Administration. (2020, October 20). In Vitro Diagnostics EUAs. Retrieved October 12, 2020, from https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/vitro-diagnostics-euas

30. Wacharapluesadee, S., Kaewpom, T., Ampoot, W., Ghai, S., Khamhang, W., Worachotsueptrakun, K., . . . Hemachudha, T. (2020). Evaluating the efficiency of specimen pooling for PCR-based detection of COVID-19. Journal of medical virology, 92(10), 2193-2199. Retrieved from http://www.epistemonikos.org/documents/af0be4e35bbbe5315e0914ddfe146bfadf899a85

31. World Health Organization (WHO). (2020). Laboratory testing for coronavirus disease (COVID-19) in suspected human cases. Retrieved April 16, 2020 from https://apps.who.int/iris/rest/bitstreams/1272454/retrieve

32. Yelin, I., Aharony, N., Shaer Tamar, E., Argoetti, A., Messer, E., Berenbaum, D., . . . Kishony, R. (2020). Evaluation of COVID-19 RT-qPCR test in multi-sample pools. Clin Infect Dis. doi:10.1093/cid/ciaa531.

33. Yip, C. C., Sridhar, S., Leung, K. H., Ng, A. C., Chan, K. H., Chan, J. F., . . . To, K. K. (2020). Development and Evaluation of Novel and Highly Sensitive Single-Tube Nested Real-Time RT-PCR Assays for SARS-CoV-2 Detection. Int J Mol Sci, 21(16). doi:10.3390/ijms21165674.

10. APPENDICES

Appendix 1. Characteristics of Included Studies

Study ID	Country	Population	Sample Size	Index Test	Brand of PCR used for Index Test	Pool Size	Reference Standard	Brand of PCR used for Reference Standard	Specimen	Prevalence	Outcomes
Bateman, 2020	USA	Highest risk symptomatic patients for testing (hospitalized, healthcare workers, etc.) and symptomatic patients from many populations, as well as individuals tested as part of public health investigations (asymptomatic or symptomatic).	100 specimens that attain the distribution of CT values of the first 838	Individual RT-PCR testing of diluted samples to mimic pooled testing	CDC RT-PCR Assay	5, 10, 50	Individual RT-PCR testing of undiluted samples to mimic individual testing	U.S. Centers for Disease Control and Prevention (CDC) 2019-nCoV Real-Time RT-PCR Diagnostic Panel	Specimens before freezing and dilution	Unspecified	Sensitivity Undiluted (N1 or N2): 99% (95% CI: 94.6%–99.9%) Undiluted (N1 and N2): 95% (95% CI: 88.7%–98.4%) Undiluted (N1 only): 99% (95% CI: 94.6%–99.9%) Undiluted (N2 only): 95% (95% CI: 88.7%–98.4%) 1:5 dilution (N1 or N2): 93% (95% CI: 86.1%–97.1%) 1:5 dilution (N1 and N2): 88% (95% CI: 80.0%–93.6%) 1:5 dilution (N1 only): 89% (95% CI: 81.2%–94.4%) 1:5 dilution (N2 only): 92% (95% CI:84.8%–96.5%) 1:10 dilution (N1 or N2): 91% (95% CI: 83.6%–95.8%) 1:10 dilution (N1 and N2): 83% (95% CI: 74.2%–89.8%) 1:10 dilution (N1 only): 89% (95% CI:81.2%–94.4%) 1:10 dilution (N2 only):

											85% (95% CI: 76.5%–91.4%) 1:50 dilution (N1 or N2): 81% (95% CI:71.9%–88.2%) 1:50 dilution (N1 and N2): 73% (95% CI:63.2%–81.4%) 1:50 dilution (N1 only): 78% (95% CI: 68.6%–85.7%) 1:50 dilution (N2 only): 76% (95% CI: 66.4%–84.0%)
Ben-Ami, 2020	Israel	Symptomatic patients and asymptomatic populations such as hospital employees and workers in essential industries	Positive = unknown Negative = unknown Indeterminate = 5 Total = 184	RT-PCR (Dorfman pooling)	Real-time fluorescent RT-PCR kit (BGI)	8	RT-PCR (individual)	Real-time fluorescent RT-PCR kit (BGI)	Combined deep nasal and oropharyngeal swabs	Unspecified	This approach yielded highly accurate results, with no loss of diagnostic assay sensitivity: each of the pools that contained one or more positive sampleswas found to be positive, and all the pools that contained only negative samples were found to be negative.
			Positive = 3 Negative = 72 Total = 75	RT-PCR (matrix pooling)		5 x 5 matrix					As expected, only six pools (one row and one column per matrix) were positive for SARS-COV-2, while 24 pools had threshold cycle (Ct) > 40 (Undetected).
			Positive = 5 Negative = 2163 Total = 2168	RT-PCR (Dorfman pooling)		8					TP = 4, FP = 1, FN = 0, TN = 266
De Salazar, 2020	Spain	Nasopharyngeal swabs from patients or health professionals	3519 nasopharyngeal swabs	Pooled testing	Viasure SARS-CoV-2 Real Time PCR (CerTest) TaqMan	9, 10	Individual RT-PCR testing	Viasure SARS-CoV-2 Real Time PCR (CerTest) TaqMan 2019-nCoV Assay Kit	Nasopharyngeal swabs	Unspecified	Sensitivity Major discordance: 97.01% (94.11% to 98.82%) All discordance: 85.48% (80.39%-89.67%)

					2019-nCoV Assay Kit v1 (Thermo Fisher Scientific) Allplex 2019-nCoV Assay (Seegene) Light Mix E gene (Roche)			v1 (Thermo Fisher Scientific) Allplex 2019-nCoV Assay (Seegene) Light Mix E gene (Roche)			Specificity Major discordance: 100% (99.89% to 100%) All discordance: 85.48% (99.89 to 100%) PPV Major discordance:100% All discordance: 100% NPV Major discordance: 99.79% (CI 99.56% to 99.90%) All discordance: 98.94% (CI 98.57% to 99.22%) Accuracy Major discordance: 99.80 (99.59% to 99.92%) All discordance: 99.00% (98.62% to 99.30%)
Freire-Paspuel, 2020	Ecuador	Nasopharyngeal swabs from individuals selected during SARS-CoV-2 surveillance	Positive = 38 Negative = 76 Total = 114	RT-PCR (pooled)	2019-nCoV CDC EUA kit (IDT, USA)	3	RT-PCR (individual)	2019-nCoV CDC EUA kit (IDT, USA)	Nasopharyngeal	Unspecified	Sensitivity: 100%
Gupta, 2020	India	Unspecified	Positive = 40 Negative = 240 Total = 280	RT-PCR (pooled)	LightMix® SarbecoV E-gene (TIB MOLBIOL)	8	RT-PCR (individual)	LightMix® SarbecoV E-gene (TIB MOLBIOL) and LightMix® Modular SARS-CoV-2 RdRP (TIB MOLBIOL)	Combined nasopharyngeal and oropharyngeal swabs	Unspecified	Sensitivity: 95.4% Specificity: 100% PPV: 100% NPV: 92.86%
Khodare, 2020	India	Unspecified	Positive and negative nasopharyngeal and oropharyngeal sample elutes	RT-PCR (pooled)	LightMix® SarbecoV E-gene (TIB MOLBIOL) and LightMix®	2, 4, 6, 8, 10, 12, 16, 20, 24, 32, 48	RT-PCR	For the positive samples: LightMix® SarbecoV E-gene (TIB MOLBIOL) and	Nasopharyngeal and oropharyngeal sample elutes	Unspecified	Out of the 77 pools, only 53 (68.8%) were found positive. The sensitivity of pools of 2–48 samples was decreased from 100%

			Positive samples= 7 Negative samples = 48		Modular SARS-CoV-2 RdRP (TIB MOLBIOL)			LightMix® Modular SARS-CoV-2 RdRP (TIB MOLBIOL)			(95% confidence interval [CL]; 98.4–100) to 41.41% (95% CL; 34.9–48.1). The maximum size of the pool with acceptable sensitivity (>95%) was found to be of six samples. For the pool size of six samples, the sensitivity was 97.8% and the efficiency of pooling was 0.38.
Kim, 2020	South Korea	Using clinical specimens from 3 hospitals in South Korea: Seoul Medical Center and National Medical Center, both in Seoul, and Jeonbuk National University Hospital in Jeonju.	50 positives, 300 negatives	RT-PCR	PowerCheck 2019-nCoV Real-Time Detection	2, 4, 6, 8, 10, 16	RT-PCR	E gene and RdRP gene: STANDARD M nCoV Real-time Detection or PowerCheck 2019-nCoV Real-Time Detection	Pooled upper respiratory specimens	Unspecified	Sensitivity: 1:2 – 100 (96-100) 1:4 – 100 (96-100) 1:6 – 100 (96-100) 1:8 – 97 (92-99) 1:10 – 99 (95 – 100) 1:16 – 96 (90- 98)
Mitchell, 2020	USA	Patients presenting to Pittsburgh-based UPMC medical facilities	35 positive pools, 20 negative pools	RT-PCR (pooled sampling)	CDC 2019 nCoV Real-Time RT-PCR Diagnostic Panel (CDC); Hologic Aptima SARS-CoV-2 TMA assay (TMA)	4	RT-PCR (individual sampling)	Cepheid SARS-CoV-2 EUA	Frozen residual nasopharyngeal swabs	Unspecified	All 20 negative pools were negative by both CDC and TMA assay, resulting in 100% specificity. For positive pools, the CDC assay correctly detected 34/35, with a sensitivity of 97.4%. For the TMA assay, it was able to correctly detect 32/35 positive pools, with a sensitivity of 91.4%.
Pasomsub, 2020	Thailand	Patients under investigation for COVID-19	200 RNA specimens	SARS-CoV-2 Nucleic Acid	Sansure, Changsha, China	5, 10	SARS-CoV-2 Nucleic Acid Diagnostic Kit	Sansure, Changsha, China	Saliva samples	9.00%	Of the 40 pools of five samples, there were 27 negative pools. Eleven

		during the outbreak in Bangkok, Thailand		Diagnostic Kit (pooled sampling) targeting ORFlab and N gene			(individual sampling)				pools detected both ORFlab and N genes, and two pools detected only the N gene. Of the 20 pools of ten samples, there were seven negative pools. Twelve pools detected both ORFlab and N genes, and one pool detected only the ORFlab gene.
Perchetti, 2020	USA	Unspecified	32 distinct positive samples pooled into negative specimens, 32 four-way pools of negative specimens	Four-way pooling using RT-PCR kit targetting N1 and N2	CDC-based Washington state EUA SARS-CoV-2 RT-PCR assay	4	Individual RT-PCR testing	CDC-based Washington state EUA SARS-CoV-2 RT-PCR assay	Nasopharyngeal swabs	Unspecified	Thirty out of 32 (94%) positive sample pooled into negative specimens were detected by pooling. The only two missed samples by pooling had Cts ≥ 35. Two low positives (Ct > 33.5) were inconclusive pooling (one target positive) and according to EUA protocol would be considered positive with individual specimens repeated. These inconclusive results were confirmed as low positives when repeated from neat sample.
Praharaj, 2020	India	Unspecified	1000 positive samples, number of negative samples unspecified	RT-PCR (pooled sampling)	TIB Molbiol 2019 nCoV Kit (TIB Molbiol, Germany). Standard M nCoV Real-Time Detection Kit (SD Biosensor	5, 10	RT-PCR	TIB Molbiol 2019 nCoV Kit (TIB Molbiol, Germany). Standard M nCoV Real-Time Detection Kit (SD Biosensor Inc., Republic of Korea) and	Nasopharyngeal and oropharyngeal samples	Unspecified	A total of 110 each of 5- and 10-sample pools were evaluated. Concordance between the 5-sample pool and individual sample testing was 100 per cent in the Ct value ≤30 cycles and 95.5 per cent for Ct values ≤33 cycles. Overall concordance

					Inc., Republic of Korea) and PathoDetect COVID-19 Detection Kit (Mylab Discovery Solutions, Maharashtra).			PathoDetect COVID-19 Detection Kit (Mylab Discovery Solutions, Maharashtra).			between the 5-sample pooled and individual sample testing was 88 per cent while that between 10-sample pool and individual sample testing was 66 per cent. Although the concordance rates for both the 5- and 10-sample pooled testing varied across laboratories, yet for samples with Ct values ≤33 cycles, the concordance was ≥90 per cent across all laboratories for the 5-sample pools.
Sahajpal, 2020	USA	Unspecified	940 samples, 934 negative, 6 positive	RT-PCR (pooled sampling)	Perkin Elmer Assay	10	Cannot be confirmed	Cannot be confirmed	Nasopharyngeal samples	Unspecified	Of the 94 pools/wells, four were positive [Ct values: N (22.7 to 28.3), ORF1ab (23.3 to 27.2), and internal control (34.4 to 35.4)]. The 40 samples comprising the four pools were identified and reanalyzed individually; six samples were positive, with Ct values of N gene, ORF1ab, and internal control comparable to their respective wells. Additional experiments were performed on samples with high Ct values, and overall results showed 91.6% positive and 100% negative agreement

											compared with individual testing approach. Thus, 940 samples were tested in 148 reactions compared with 940 reactions in routine screening.
Schmidt, 2020	Germany	Samples from a proficiency panel test provider (INSTAND) with predetermined concentrations of SARS-CoV-2	Positive = 4 Negative = unknown	RT-PCR (multiple swab pooling)	Roche Cobas SARS-CoV-2	5	RT-PCR (individual)	Roche Cobas SARS-CoV-2	Swabs from the pharyngeal region and nasal region	Unspecified	We determined that all multiple-swab tubes containing a SARS-CoV-2–positive sample were correctly identified in the multiple-swab protocol, independent of the virus concentration in the original sample. All multiple-swab tubes containing no SARS-CoV-2–positive sample were also true negative.
		Patients with clinical symptoms and moderate likelihood of SARS-CoV-2 infection	Positive = 4 Negative = 46 Total = 50			5					Each of the four pools containing a positive sample was correctly identified withthe multiple-swab method. Multiple-swab tubes containing no positive sample were also correctly identified to be negative in multiple-swab tubes of five swabs.
		Asymptomatic residents of a nursing home	Positive = 8 Negative = 92 Total = 100			10					All 5 multiple-swab tubes containing a total of 8 positive swabs were correctly identified. All 5 multiple-swab tubes containing no positive swab sample were also true negative.
Shental, 2020	Israel	For testing of P-BEST as a method (initial validation): 384	For testing of P-BEST as a method (initial validation): 384	Pooled RT-PCR using P-BEST method	Seegene, CA, USA	48	Individual RT-PCR testing (only for initial validation of	Seegene, CA, USA	Naso- and oropharynx swabs	Unspecified but P-BEST pooling scheme was	For initial validation of P-BEST as a method:

		samples were used, pooling design was constructed for a carrier rate of up to ~1% (4 sets of 384 samples) For using P-BEST to screen asymptomatic health care employees: 1115 participants were recruited. -690 female, 425 male -Soroka University Medical Center Staff: physicians (n= 165), nurses (n=157), nurse assistants (n=43), other clinical staff (n=119), administrative staff (n=631) -296 subjects worked in direct contact with COVID-19 patients -Presence of symptoms: 926 totally asymptomatic, 71 with mild	samples For using P-BEST to screen asymptomatic health care employees: samples from 1115 participants				P-BEST as a method); for screening of asymptomatic health care employees, no reference standard was used			designed for carrier rate of < 1.3%	2 positive carriers (2+)/384 samples - True positives: 2 - False positives: 0 - True negatives: 382 - False negatives: 0 3 positive carriers (3+)/384 samples - True positives: 3 - False positives: 0 - True negatives: 381 - False negatives: 0 4 positive carriers (4+)/384 samples - True positives: 4 - False positive: 0 - True negatives: 380 - False negatives: 0 5 positive carriers - True positives: 5 - False positives: 1 - True negatives: 378 - False negatives: 0 For using P-BEST in the screening of asymptomatic healthcare workers, all pools tested were negative. The third batch was blindly spiked with a sample from a patient with COVID-19 and this was correctly identified using P-BEST.

		cough, 70 with rhinorrhea									
Singh, 2020	India	Suspected COVID-19 patients	545	Pooled testing	Real-Time Fluorescent RT-PCR Kit (BGI, Hong Kong)	5	Individual RT-PCR testing	Real-Time Fluorescent RT-PCR Kit (BGI, Hong Kong)	Nasopharyngeal and oropharyngeal swabs	4.8% point prevalence	Sensitivity: 75% (47.6% to 92.7%) Specificity: 98.9% (94.2% to 100%) PPV 41.3% (1% Prevalence) 58.7% (2% Prevalence) 68.3% (3% Prevalence) 74.4% (4% Prevalence) 78.6% (5% Prevalence) NPV 99.8% (1% Prevalence) 99.5% (2% Prevalence) 99.2% (3% Prevalence) 99% (4% Prevalence) 98.7% (5% Prevalence)
Wacharapluesadee, 2020	Thailand	Patients under investigation (PUI)	Pooled previously 50 negative SARS-CoV-2 specimens and Forty-nine PCR positive NT specimens	Ratio pooling	qPCR (BGI, Shenzhen, China)	10	Individually tested samples using the standard realtime quantitative PCR (qPCR)	qPCR (BGI, Shenzhen, China)	Nasopharyngeal and throat swabs	Unspecified	This study demonstrates that specimen pooling (either 1X or 2X pooling ratios) does not compromise the sensitivity of detecting SARS-CoV-2 provided the Ct value of the individually tested sample is lower than 35. There were no significant difference between the Ct values of individual testing vs ratio pools 2X L+L, 2X H+L, or 2X H+H (P=.063, .507, and .6766, respectively). Thus, sensitivity was not affected by pooling specimens, while accuracy was maintained.

Yelin, 2020	Israel	Unspecified	Positive = 5 Negative = 67 Total = 72	RT-PCR (pooled)	AgPath-ID One-Step RT-PCR Reagents (Thermo Fisher Scientific)	32	RT-PCR (individual)	AgPath-ID One-Step RT-PCR Reagents (Thermo Fisher Scientific)	Swabs from both nostril and throat	Unspecified	Of the 10 tested replicates, only duplicate B of sample 2 did not cross the threshold in pools of 32.
Yip, 2020	Hong Kong	Suspected COVID-19 patients	Positive samples with low viral load-unspecified number 49 negative samples	Nested RT-PCR assay	STN real-time RT-PCR assays (STN COVID-19-RdRp/Hel Assay and STN COVID-19-N Assay)	4	Non-nested assay	Non-nested COVID-19-RdRp/Hel assay	Respiratory specimens	Unspecified	To evaluate the performance of our novel STN assays in pooled specimens, we created four sample pools, with each pool consisting of one low positive specimen and 49 negative specimens. While the non-nested COVID-19-RdRp/Hel assay was positive in only one of four sample pools (25%), both of the STN assays were positive in two of four samples pools (50%).

Appendix 2A. Diagnostic accuracy of included studies (Unit of analysis: Pools)

Study ID	N	Sensitivity		Specificity		Positive Predictive Value (PPV)		Negative Predictive Value (NPV)		Positive Likelihood Ratio		Negative Likelihood Ratio	
		%	95% CI	%	95% CI	%	95% CI	%	95% CI	LR+	95% CI	LR-	95% CI
Bateman, 2020 (1)	100	88	79.98, 93.64	-	-	-	-	-	-	-	-	-	-
Bateman, 2020 (2)	100	83	74.18, 89.77	-	-	-	-	-	-	-	-	-	-
Bateman, 2020 (3)	100	73	63.20, 81.39	-	-	-	-	-	-	-	-	-	-
Ben-Ami, 2020 (1)	18	100	78.20, 100.00	100	29.24, 100	100	-	100	-	3001	0, $2.46x10^{31}$	<0.01	0, $5.50x10^{22}$
Ben-Ami, 2020 (2)	30	100	54.07, 100.00	100	85.75, 100	100	-	100	-	23997	0, $1.98x10^{19}$	<0.01	0, $1.37x10^{23}$
Ben-Ami, 2020 (3)	271	100	39.76, 100.00	99.56	97.57, 99.99	80	36.14, 96.58	100	-	267	37.74, 1888.15	<0.01	0, $2.06x10^{23}$
Freire-Paspuel, 2020	36	100	90.26, 100.00	-	-	-	-	-	-	-	-	-	-
Gupta, 2020	35	95.45	77.16, 99.88	100	75.29, 100	100	-	92.86	65.70, 98.88	12410	0, $1.02x10^{31}$	0.05	0, 0.31
Khodare, 2020 (1)	8	100	59.04, 100	100	2.5, 100	100	-	100	-	1002	0, $8.03x10^{29}$	<0.01	0, $1.18x10^{23}$
Khodare, 2020 (2)	8	100	59.04, 100	100	2.5, 100	100	-	100	-	1002	0, $8.03x10^{29}$	<0.01	0, $1.18x10^{23}$
Khodare, 2020 (3)	8	85.71	42.13, 99.64	100	2.5, 100	100	-	50	14.01, 85.99	859	0, $6.89x10^{29}$	0.14	0.02, 0.88
Khodare, 2020 (4)	8	71.43	29.04, 96.33	100	2.5, 100	100	-	33.33	13.42, 61.73	716	0, $5.75x10^{29}$	0.29	0.09, 0.92
Khodare, 2020 (5)	8	57.14	18.41, 90.1	100	2.5, 100	100	-	25	12.41, 43.95	573	0, $4.60x10^{29}$	0.43	0.18, 1.01
Khodare, 2020 (6)	8	57.14	18.41, 90.1	100	2.5, 100	100	-	25	12.41, 43.95	573	0, $4.60x10^{29}$	0.43	0.18, 1.01
Khodare, 2020 (7)	8	57.14	18.41, 90.1	100	2.5, 100	100	-	25	12.41, 43.95	573	0, $4.60x10^{29}$	0.43	0.18, 1.01
Khodare, 2020 (8)	8	57.14	18.41, 90.1	100	2.5, 100	100	-	25	12.41, 43.95	573	0, $4.60x10^{29}$	0.43	0.18, 1.01
Khodare, 2020 (9)	8	57.14	18.41, 90.1	100	2.5, 100	100	-	25	12.41, 43.95	573	0, $4.60x10^{29}$	0.43	0.18, 1.01
Khodare, 2020 (10)	8	57.14	18.41, 90.1	100	2.5, 100	100	-	25	12.41, 43.95	573	0, $4.60x10^{29}$	0.43	0.18, 1.01
Khodare, 2020 (11)	8	42.86	9.90, 81.59	100	2.5, 100	100	-	20	11.63, 32.20	429	0, $3.45x10^{29}$	0.57	0.3, 1.09
Kim, 2020 (1)	100	100	96, 100	-	-	-	-	-	-	-	-	-	-
Kim, 2020 (2)	100	100	96, 100	-	-	-	-	-	-	-	-	-	-
Kim, 2020 (3)	100	100	96, 100	-	-	-	-	-	-	-	-	-	-
Kim, 2020 (4)	100	97	92, 99	-	-	-	-	-	-	-	-	-	-
Kim, 2020 (5)	100	99	95, 100	-	-	-	-	-	-	-	-	-	-
Kim, 2020 (6)	160	96	90, 98	97	87, 99	97.90	92.4, 99.5	93.60	84.7, 97.4	29	7.4, 112.5	0.04	0.02, 0.11
Mitchell, 2020 (1)	55	97.14	85.08, 99.93	100	83.16, 100	100	-	95.24	74.34, 99.28	19430	0, $1.60x10^{31}$	0.03	0, 0.20
Mitchell, 2020 (2)	55	91.43	76.94, 98.2	100	83.16, 100	100	-	86.96	69.32, 95.16	18287	0, $1.51x10^{31}$	0.09	0, 0.25
Mitchell, 2020 (3)	23	100	85.18, 100	-	-	-	-	-	-	-	-	-	-
Mitchell, 2020 (4)	4	100	39.76, 100	-	-	-	-	-	-	-	-	-	-
Mitchell, 2020 (5)	8	87.50	47.35, 99.68	-	-	-	-	-	-	-	-	-	-
Mitchell, 2020 (6)	23	100	85.18, 100	-	-	-	-	-	-	-	-	-	-
Mitchell, 2020 (7)	4	100	39.76, 100	-	-	-	-	-	-	-	-	-	-
Mitchell, 2020 (8)	8	62.50	24.49, 91.48	-	-	-	-	-	-	-	-	-	-
Pasomsub, 2020 (1)	40	100	75.29, 100	100	87.23, 100	100	-	100	-	26999	0, $2.23x10^{31}$	<0.01	0, $6.34x10^{22}$
Pasomsub, 2020 (2)	20	100	75.29, 100	100	59.04, 100	100	-	100	-	7000	0, $5.6x10^{30}$	<0.01	0, $6.34x10^{22}$
Perchetti, 2020	64	93.75	79.19, 99.23	100	89.11, 100	100	-	94.12	80.70, 98.39	30000	0, $2.48x10^{31}$	0.06	0, 0.24

Study ID	N	Sensitivity		Specificity		Positive Predictive Value (PPV)		Negative Predictive Value (NPV)		Positive Likelihood Ratio		Negative Likelihood Ratio	
		%	95% CI	%	95% CI	%	95% CI	%	95% CI	LR+	95% CI	LR-	95% CI
Praharaj, 2020 (1)	100	88	79.98, 93.64	-	-	-	-	-	-	-	-	-	-
Praharaj, 2020 (2)	100	66	55.85, 75.18	-	-	-	-	-	-	-	-	-	-
Praharaj, 2020 (3)	23	100	85.18, 100	-	-	-	-	-	-	-	-	-	-
Praharaj, 2020 (4)	44	95.45	84.53, 99.44	-	-	-	-	-	-	-	-	-	-
Praharaj, 2020 (5)	33	69.70	51.29, 84.41	-	-	-	-	-	-	-	-	-	-
Praharaj, 2020 (6)	23	95.65	78.05, 99.89	-	-	-	-	-	-	-	-	-	-
Praharaj, 2020 (7)	44	79.55	64.70, 90.20	-	-	-	-	-	-	-	-	-	-
Praharaj, 2020 (8)	33	27.27	13.30, 45.52	-	-	-	-	-	-	-	-	-	-
Schmidt, 2020 (1)	5	100	39.76, 100	100	2.50, 100	100	-	100	-	1002	0, $8.02x10^{29}$	<0.01	0, $2.05x10^{23}$
Schmidt, 2020 (2)	10	100	39.76, 100	100	54.07, 100	100	-	100	-	6001	0, $4.94x10^{30}$	<0.01	0, $2.05x10^{24}$
Schmidt, 2020 (3)	10	100	47.82, 100	100	47.82, 100	100	-	100	-	5001	0, $4.11x10^{30}$	<0.01	0, $1.64x10^{23}$
Singh, 2020	109	75	47.62, 92.73	98.92	94.15, 99.97	92.31	62.60, 98.85	95.83	90.78, 98.17	70	9.73, 500.08	0.25	0.11, 0.59
Wacharapluesadee, 2020	49	95.92	86.02, 99.50	-	-	-	-	-	-	-	-	-	-
Yelin, 2020	12	90	55.5, 99.75	100.00	15.81, 100	-	-	-	-	1801	0, $1.47x10^{30}$	0.1	0.02, 0.64
Yip, 2020 (1)	4	25	0.63, 80.59	-	-	-	-	-	-	0.5	0, $5.58x10^{18}$	1.5	0, $1.63x10^{19}$
Yip, 2020 (2)	4	50	6.76, 93.24	-	-	-	-	-	-	1	0, $1.09x10^{19}$	1	0, $1.09x10^{19}$
Yip, 2020 (3)	4	50	6.76, 93.24	-	-	-	-	-	-	1	0, $1.09x10^{19}$	1	0, $1.09x10^{19}$

Appendix 2B. Diagnostic accuracy of included studies (Unit of analysis: Individuals)

Study ID	N	Sensitivity		Specificity		Positive Predictive Value		Negative Predictive Value		Positive Likelihood Ratio		Negative Likelihood Ratio	
		%	95% CI	%	95% CI	%	95% CI	%	95% CI	LR+	95% CI	LR-	95% CI
de Salazar, 2020	3509	97.10	94.11, 98.82	100	99.89, 100	100	-	99.79	99.56, 99.90	$3.17x10^{6}$	0, $2.62x10^{33}$	0.03	0.01, 0.05
Sahajpal, 2020	1000	91.67	61.52, 99.79	100	99.63, 100	100	-	99.90	99.34, 99.98	$9.1x10^{5}$	0, $7.49x10^{32}$	0.08	0.01, 0.54
Shental, 2020 (1)	384	100	15.81, 100	100	99.04, 100	100	-	100	-	$3.81x10^{5}$	0, $3.16x10^{32}$	<0.01	0, $4.07x10^{23}$
Shental, 2020 (2)	384	100	29.24, 100	100	99.04, 100	100	-	100	-	$3.81x10^{5}$	0, $3.15x10^{32}$	<0.01	0, $2.73x10^{23}$
Shental, 2020 (3)	384	100	39.76, 100	100	99.03, 100	100	-	100	-	$3.80x10^{5}$	0, $3.14x10^{32}$	<0.01	0, $2.05x10^{23}$
Shental, 2020 (4)	384	100	47.82, 100	99.74	98.54, 99.99	83.33	41.39, 97.25	100	-	378.92	2683.16	<0.01	0, $1.65x10^{23}$

Appendix 3. Country Guidelines on the Use of Pooled Testing

Country	Use Case	Use of Pooled Testing	References
United Kingdom	**screening**	**There are no available guidelines on pooled testing released by the National Health Service; however, according to news articles, UK recommends the use of pooled testing on screening students returning to universities. Number of samples in a pool was not elaborated** Sept 29 Some UK universities are introducing covid-19 screening programmes using pooled testing to help prevent outbreaks and allow campuses to stay open. The University of Cambridge and the University of Nottingham are both using pooled testing, which involves mixing several samples together and then testing the pooled sample. If the result comes back positive the people in the group then need to be tested individually. This approach increases the number of people who can be tested using the same amount of resources—saving time, supplies, and money. However, some experts have raised concerns over whether the costs, benefits, and harms of such programmes have been evaluated, and they have called for advice from the UK National Screening Committee. Guidance from the University of Cambridge said that all students living in college accommodation would be eligible to take part in its scheme, which has a capacity of 2000 tests a week and can test around 16 000 students using the pooling method. This means that all students living in college accommodation can be tested every week. The guidance said, "Compared with other members of the population, young adults have a higher chance of asymptomatic infection. Public health experts have therefore called for asymptomatic screening in high risk settings, such as universities. As well as protecting students	Mahase, E. (2020). COVID-19 Universities roll out pooled testings of students in bid to keep campuses open. Retrieved Oct. 16, 2020 from: https://www.bmj.com/content/370/bmj.m3789

		directly, controlling transmission between students will help prevent onward transmission to staff and the wider Cambridge community." However, there are downsides to pooling. The US Food and Drug Administration says that, because the samples are diluted, pooling can "result in less viral genetic material available to detect" and a greater likelihood of false negative results. Allyson Pollock, professor of public health and co-director of the Newcastle University Centre of Research Excellence in Regulatory Science, told *The BMJ*, "The UK National Screening Committee should be consulted and involved with this, because screening is a very, very complex public health intervention. It should be carefully evaluated so that we know what the costs, harms, and benefits are, because screening like this will be extremely costly." Pollock has raised concerns over the potential for many false positives, as well as the inability of current testing technology to determine infectiousness. "You could end up quarantining students and staff and their contacts and their whole households unnecessarily and causing a lot of hardship," she said, adding that the financial harm this could cause to students—many of whom will be working—must also be considered. Despite these concerns Independent SAGE has recommended regular testing of students and staff. In a report on universities published on 28 September the group called for all staff and students to be tested at the start of the academic year, before initiating any in-person contact. Institutions should also "initiate structured surveillance programmes for high risk settings such as residential halls of residence—this could include innovative approaches such as sewage testing, pooled sample testing, as well as random sampling," the document said.	
United States	**Diagnosis** **Screening** **Surveillance**	**The US CDC allows use of pooled sample strategy for diagnostic, screening and surveillance testing. Pooling strategy for diagnostic and screening tests depends on the community prevalence of virus, and pool size will need to be adjusted accordingly, pooling strategy should only be used when the prevalence of COVID-19 is low. CDC recommends that laboratories should determine prevalence based on a rolling average of**	Center for Disease Controls (2020). Interim Guidance for Use of Pooling Procedures in SARS-CoV-2 Diagnostic, Screening and Surveillance testing Retrieved Nov. 3, 2020 from: https://www.cdc.gov/coronavirus/2019-ncov/lab/pooling-

		the positivity rate of their own SARS-CoV-2 testing over the previous 7–10 days *Diagnostic or Screening Testing Using a Pooling Strategy* General Guidance Laboratories certified under the Clinical Laboratory Improvement Amendments (CLIA) can use a specimen pooling strategy to expand SARS-CoV-2 nucleic acid diagnostic or screening testing capacity when using a test authorized for such use by FDA. If a pooled test result is negative, then all specimens can be presumed negative with the single test. If the test result is positive or indeterminate, then all the specimens in the pool need to be retested individually. The advantages of this two-stage specimen pooling strategy include preserving testing reagents and resources, reducing the amount of time required to test large numbers of specimens, and lowering the overall cost of testing. A pooling strategy depends on the community prevalence of virus, and pool size will need to be adjusted accordingly. CDC recommends that laboratories should determine prevalence based on a rolling average of the positivity rate of their own SARS-CoV-2 testing over the previous 7–10 days. Laboratories should use a standardized methodology or calculator that factors in the sensitivity of the assay they are using and their costs of testing to determine when the positivity rate is low enough to justify the implementation of a pooling strategy. Laboratories should also understand and, where appropriate, communicate the limitations associated with pooled testing, which are described in greater detail below. *Limitations of Pooled Diagnostic or Screening Testing* Based on limited data, using a pooling testing procedure for SARS-CoV-2 has some limitations. In a pooling procedure, the laboratory cannot ensure the diagnostic integrity of an individual specimen because it is combined with other specimens before testing. Specimen integrity can be affected by the quality of swab specimen collection, which could result in some swabs having limited amounts of viral genetic material	procedures.html#:~:text=What%20is%20pooling%3F,virus%20that%20causes%20COVID%2D19.

		for detection. Inadequate individual specimens, including those with limited amounts of viral genetic material, might not be eliminated from the pooled specimen before testing. Even if each individual specimen in a pool is adequate, the specimens in a pooled procedure are diluted, which could result in a low concentration of viral genetic material below the limit of detection of a given test. These limitations mean that monitoring the prevalence of disease and properly validating the assay and the instrumentation are important to limit the potential for false-negative results. In general, the larger the pool of specimens, the higher the likelihood of generating false-negative results. The prevalence of COVID-19 in a population also affects the efficiency of pooled testing strategies. In general, lower disease prevalence may enable a laboratory to use a larger optimal pool size. A recent study by the Nebraska Public Health Laboratory found that nucleic acid tests for SARS-CoV-2 reliably returned a positive result when one positive sample was mixed with four negatives, and could reduce the number of tests needed by >50% in certain scenarios (such as a COVID-19 prevalence of 5%). However, as the prevalence of COVID-19 increases, the cost savings of a pooling strategy decreases because more pooled tests will return positive results and those specimens will need to be retested individually. CDC continues to pursue research studies on pooling strategies for testing for SARS-CoV-2 and will update this guidance as needed. *Surveillance Testing Using a Pooling Strategy* General Guidance Surveillance testing can be conducted in a laboratory that has a CLIA certificate, or in a laboratory that does not have a CLIA certificate. CMS indicates that ***during the COVID-19 public health emergency and associated authorizations***, "facilities performing SARS-CoV-2 surveillance testing using a pooled sampling procedure to report ***non patient-specific*** SARS-CoV-2 cohort results will not require CLIA certification." FDA's FAQs on Testing for SARS-CoV-2external icon state that FDA generally does not regulate surveillance testing.	

		A specimen pooling strategy can expand a laboratory's capacity to conduct SARS-CoV-2 nucleic acid surveillance testing. If a pooled test result is negative, then all specimens can be presumed negative with the single test. If the test result is positive or indeterminate, then all the specimens in the pool need to be retested individually. The advantages of this two-stage specimen pooling strategy include preserving testing reagents and resources, reducing the amount of time required to test large numbers of specimens, and lowering the overall cost of testing. A pooling strategy should only be used when the prevalence of COVID-19 is low. CDC recommends that laboratories should determine prevalence based on a rolling average of the positivity rate of their own SARS-CoV-2 testing over the previous 7–10 days. Laboratories should use a standardized methodology or calculator that factors in the sensitivity of the assay they are using and their costs of testing to determine when the positivity rate is low enough to justify the implementation of a pooling strategy. Laboratories should also understand the limitations associated with pooled testing, which are described in greater detail below. *Assays and Test Systems for Pooling of Surveillance Testing* Although FDA generally does not regulate surveillance testing, CDC recommends that laboratories conducting surveillance testing with pooling should use an assay and test system that has received an EUA from FDA. Using an FDA-authorized assay and test system helps ensure the quality and reliability of testing. FDA authorized SARS-CoV-2 in vitro diagnostic devices are included on FDA's list of In Vitro Diagnostics EUAs . Laboratories should use an existing authorized nucleic acid assay, and, if not authorized for use with pooling, evaluate and validate the performance of that assay for a pooling strategy according to FDA's guidance in its Molecular Diagnostic Template for Laboratories. If laboratories modify the authorized assay by incorporating alternative components, including extraction methods, PCR instruments, and software versions, for the purposes of a pooling strategy, the laboratories should also evaluate and validate the performance of the altered test system. *Reporting Pooled Surveillance Testing Results*	

		Results of surveillance testing can be returned in aggregate to the requesting institution, such as a university or public health agency. Negative pooled surveillance test results should be reported as "presumptive negative" to the requesting institution. Positive and indeterminate pooled surveillance test results should not be reported to the requesting institution; they should be retested individually before being reported in aggregate. Facilities, regardless of their CLIA status, should not officially report the results of surveillance testing, including surveillance testing that uses a pooling procedure, to the local, state, tribal, or territory health department as diagnostic or screening test results. If a local, state, tribal, or territory health department requests access to the results of surveillance testing for SARS-CoV-2 that uses a pooling procedure, the laboratory should state in the report to the health department that the data are pooled surveillance testing results that do not represent COVID-19 diagnostic or screening test results. Only a facility with a CLIA certificate may officially report a patient-specific diagnostic or screening COVID-19 test result to the local, state, tribal, or territory health department. Facilities that conduct surveillance testing, including surveillance testing that uses a pooling procedure, should not report test results to individuals whose specimens have been tested, or to the individual's health care provider, employer, etc. If at any time a facility intends to report a patient-specific test result, it must first obtain a CLIA certificate and meet all requirements to perform testing. *Limitations of Pooled Surveillance Testing* Based on limited data, using a pooling testing procedure for SARS-CoV-2 has some limitations. In a pooling procedure, the laboratory cannot ensure the diagnostic integrity of an individual specimen because it is combined with other specimens before testing. Specimen integrity can be affected by the quality of swab specimen collection, which could result in some swabs having limited amounts of viral genetic material for detection. Inadequate individual specimens, including those with limited amounts of viral genetic material, might not be eliminated from	

		the pooled specimen before testing. Even if each individual specimen in a pool is adequate, the specimens in a pooled procedure are diluted, which could result in a low concentration of viral genetic material below the limit of detection. These limitations mean that monitoring the prevalence of disease and properly validating the assay and the instrumentation are important to limit the potential for false-negative results. In general, the larger the pool of specimens, the higher the likelihood of generating false-negative results. The prevalence of COVID-19 in a population also affects the efficiency of pooled testing strategies. In general, lower prevalence may enable a laboratory to use a larger optimal pool size. As the prevalence of COVID-19 increases, the cost savings of a pooling strategy decreases because more pooled tests will return positive results and those specimens will need to be retested individually. CDC continues to pursue research studies on pooling strategies for testing for SARS-CoV-2 and will update this guidance as needed.	
Australia		**There are no available guidelines on pooled testing released by the Communicable Diseases Network Australia or Australian Government Department of Health; There are also no news reports regarding the use of pooled testing in Australia.**	
Canada	**Diagnosis** **Screening**	**Public Health Ontario recommends the use of pooled testing on diagnosis, outbreak investigation and asymptomatic screening, using a pool size of 3.** Sept 10 As of September 10, 2020, PHO Laboratory implemented a validated pooling approach to testing. Specimen pooling will primarily be used for a proportion of specimens submitted to PHOL from assessment centres. It may also be used for testing asymptomatic patients in a variety of other patient settings (e.g. investigations, outbreaks). Pooled testing will allow PHO Laboratory to increase testing throughput significantly, without greatly compromising clinical test sensitivity. On	Public Health Ontario (2020). COVID-19 Updates. Retrieved October 16, 2020 from: https://www.publichealthontario.ca/en/laboratory-services/test-information-index/covid-19

		average, the pooling of 3 specimens, where 1 specimen tests positive for COVID-19, results in an increase in cycle threshold (Ct) of the pool of 1.5 cycles when compared to the Ct of the positive specimen if not pooled (e.g. if a single specimen will give a positive pool result with Ct of 32, it will generate a Ct of 30.5 when retested individually). To conduct pooled testing, a portion of three individual specimens are combined into a single pool and run on the SARS-CoV-2 PCR assay as a single test. Reflex testing is done based on the intermediate result of the pool, as follows: i) If the intermediate pool result is NOT DETECTED, all three specimens are individually reported as NOT DETECTED. ii) If the intermediate pool result is DETECTED, INDETERMINATE or INVALID, each individual specimen is then tested individually and reported according to the result obtained for the individual specimen. **Note:** the intermediate pool result does NOT appear on the individual specimen PHO Laboratory report; only the final specimen level result appears on the formal report. The following specimen note will appear on the report for each individual specimen tested using a pooling approach: "This assay was tested using a pooling approach whereby 3 specimens are combined into a single pool. This method has been validated for clinical testing at PHO Laboratory. For further information see the PHO Laboratory Coronavirus Disease 2019 PCR Test Information Sheet.	

Singapore	**Screening**	**There are no available guidelines on pooled testing released by the ministry of health; however, according to news articles, Singapore uses pooled testing on migrant worker dormitories and other sub-populations with very low prevalence rates of COVID-19 or for mass screening purposes, using pool size of up to five individuals in one laboratory test.** May 12: Singapore uses pooled testing in migrant worker dormitories. Such pooled tests involve combining swabs of up to five individuals into one laboratory test , which does not affect sensitivity of the tests. Where a pooled test is positive, the original five individuals could be re-tested individually to identify the affected person. This is an effective strategy where the infection prevalence rates are likely to be low. June 15 The Inter-agency Taskforce (ITF) is continuing its efforts to systematically test and clear workers and dormitories. Up to 10,000 tests, involving individual swabs, pooled swabs and serological testing, are conducted for migrant workers daily. The ITF strives to clear all the dormitories in the most expedient and efficient way possible, while safeguarding the good health of residents. Dormitories are prioritized based on their readiness for testing, which depends on factors including the prevalence of infection, length of time since the onset of infection, and adherence to safe distancing July 25 Pooled testing can reduce the use of resources, and this strategy can be used in settings where the Covid-19 prevalence is low as well as higher-risk areas like dormitories, said the Health Ministry's director of medical services Kenneth Mak. May 12	Ministry of Health Singapore (2020) Controlling the outbreak, preparing for next phase. Retrieved October 5, 2020 from: https://www.moh.gov.sg/news-highlights/details/controlling-the-outbreak-preparing-for-the-next-phase Ministry of Health Singapore (2020) Steady progress in dormitory clearance aggressive testing and tracing in phase 2. Retrieved October 5, 2020 from: https://www.moh.gov.sg/news-highlights/details/steady-progress-in-dormitory-clearance-aggressive-testing-and-tracing-in-phase-2 Sin. Y (2020). Pooled testing can be used in dorms, community. Retrieved October 5, 2020 from: https://www.straitstimes.com/singapore/pooled-tests-can-be-used-in-dorms-community

		An MOH spokesman said it has piloted the use of pooled testing with the help of the National Public Health Laboratory and Agency for Integrated Care. The pilot was on nursing home staff, who number about 9,000 across 80 nursing homes here. It is not known exactly how many staff members pooled testing was used on. The MOH spokesman noted that pooled testing may be used in sub-populations with very low prevalence rates for Covid-19, or for mass screening purposes. She added that the pilot here used a maximum of five patient samples per pool.	Sun. D (2020). Singapore pilots pooled testing on nursing home staff. Retrieved October 5, 2020 from: https://www.tnp.sg/news/singapore/singapore-pilots-pooled-testing-nursing-home-staff
Malaysia	**screening**	**There are no available guidelines on pooled testing released by the ministry of health; however, according to news articles, Malaysia recommends the use of pooled testing on mass testing high-risk groups, using pool size of up to five individuals or more in one laboratory test.** **April 6** PUTRAJAYA, April 6 — The Health Ministry's latest strategy in the fight against the Covid-19 pandemic is to look at mass testing of groups at high risk of infection. Health director-general Datuk Dr Noor Hisham Abdullah said focusing on these groups would be a sound decision to maximize resources. "With the limited resources the ministry has, it would be best to target such high-risk groups like the Sri Petaling tabligh members or the	Choong J. (2020) Health Ministry targets at risk groups for COVID-19 mass testing. Retreived Oct. 7, from: https://www.malaymail.com/news/malaysia/2020/04/06/health-ministry-targets-at-risk-groups-for-covid-19-mass-testing/1854081

		Kuching church, nationwide," he said during the daily Health Ministry press conference. Similarly Dr Noor Hisham said the number of people who can be tested on a daily basis, currently at a capacity of 11,500, can be further increased with rapid testing kits "We also want to increase laboratory capacity in terms of the number of samples, so instead of doing a single sample we can then have a pool of five or so samples. Equally important and crucial is to enhance our labs' services. International Medical University (IMU), a private medical university in Kuala Lumpur, responded to the call for national service by volunteering to perform the SARS-CoV-2 pooled testing to help ease the congestion at the National Public Health Laboratory (NPHL), MoH.	
Indonesia	**surveillance**	**There are no available guidelines on pooled testing released by the ministry of health; however, according to news articles, Indonesia used/uses pooled testing on mass testing populations hardest provinces, however, they reiterated that it is only recommended for areas where the cases are less prevalent and using pool size of up to five individuals or more in one laboratory test.** **May 12** The guidelines from the ministry on the pooled testing for Covid 19 have been developed post a feasibility study by the /ICMR Virus Research & Diagnostic Laboratory (VRDL) at King George's Medical University (KGMU), Lucknow. The study by KGMU has suggested that the real-time PCR for the Covid-19 by pooling up to 5 samples from an area with a low infectivity rate is feasible. The study has suggested the interpretation that all samples in a pool of 5 will be labeled negative in case the pooled sample tests negative while in case of a pooled sample testing positive, deconvoluted testing should be followed which means testing samples in a pool separately. The ministry has recommended to not pool more than 5 samples for testing to avoid the samples testing false negative. Detailing the guidelines for the pooled testing, the Ministry of Health has recommended the test only for the areas where the cases are less	Sutrisno B. (2020) Government to conduct PCR pool tests in 8 provinces. Retrieved October 8, 2020 from: https://www.thejakartapost.com/news/2020/05/13/covid-19-government-to-conduct-pcr-pool-tests-in-8-provinces.html

		prevalent– less than 2 per cent positivity rate. In areas with a positivity rate between 2 to 5 per cent, the ministry has asked for sample pooling for only in community survey or surveillance among asymptomatic individuals. The ministry's guidelines have asked to strictly avoid pooling samples of individuals with known contact with confirmed cases, Health Care Workers, and rather their samples should be tested directly by rt-PCR and not by pooling. As per the guidelines issued by the health ministry, a pooled test for Covid-19 should be done by pooling 5 samples but more than 2 samples can also be pooled for testing. In order to avoid the possibility of missing samples which are positive but with low viral load, the ministry has strongly recommended against pooling more than 5 samples.	Kumar (2020) Coronavirus test: Health Ministry issues guidelines for pooled testing in all districts; check details. Retrieved Oct. 16, 2020 from: https://www.financialexpress.com/lifestyle/health/coronavirus-test-health-ministry-issues-guidelines-for-pooled-testing-in-all-districts-check-details/1956437/
Thailand	**Surveillance**	**There are no available guidelines on pooled testing released by the ministry of health; however, according to news articles, XX used/uses pooled testing on XX population, using XX pool size.** The Department of Medical Sciences is adopting the saliva pooled sampling method to accelerate COVID-19 testing in 100,000 persons from targeted groups including health and medical professionals, prison inmates, drivers of public buses and migrant workers. The method is effective and can substantially reduce the cost of COVID-19 testing. Currently, COVID-19 positive cases represent only 0.1% of all tests and more than 40,000 pooled saliva tests have been completed to date.	World Health Organization (2020). COVID-10 WHO Thailand Situation Report. Retrieved October 5, 2020 from: https://www.who.int/docs/default-source/searo/thailand/2020-06-05-tha-sitrep-90-covid19.pdf?sfvrsn=852eca78_2
Vietnam	**screening**	**There are no available guidelines on pooled testing released by the ministry of health; however, according to news articles, Vietnam uses pooled testing on returnees from Da Nang population, using pool size of three to five individuals in one laboratory test.** **August 5:** According to Dr. Ton That Thanh, director of the Danang Center for Disease Control (CDC), pool testing consists of combining samples from three to five subjects into one sample and running the test on it. If the sample is positive, then the samples will be tested separately to determine which one is positive. This approach helps screen a large amount of people in a short period of time. **August 13:**	Kiet (2020). Danang, Vietnam's Covid-19 epicenter, carries out coronavirus pool testing. Retrieved October 6, 2020 from: http://hanoitimes.vn/danang-vietnams-covid-19-epicenter-carries-out-coronavirus-pool-testing-313705.html

		A strategy to test those that returned from Da Nang since the 1st of July is being implemented by individual provinces, such as Ho Chi Minh City (HCMC) and Ha Noi. Ha Noi screened over 50 000 returnees by serological RDT and is now conducting RT-PCR testing of roughly 100 000 returnees. HCMC plans to test over 50 000 returnees. To match this high demand for testing, a scheme for pooled testing of returnees is being designed. Pooled testing is additionally being designed by Pasteur Institute Nha Trang to keep up with demand for laboratory testing in Da Nang. Official guidelines for pooled testing are currently being developed by the Ministry of Health. However, shortages in laboratory testing reagents are being reported by laboratories. Ministry of Health is additionally revising the national laboratory testing strategy in light of the current situation to ensure guidance is up to date. In Ha Noi, 4 major laboratories have been mobilized to support rt RT-PCR testing (Bach Mai hospital, National Paediatric hospital, NIHE and National Hospital of Tropical Diseases). As of 13 Aug, 17 699 samples have been collected of which 14 466 (81.7%) samples have been tested and all turned negative. According to Dr. Ton That Thanh, director of the Danang Center for Disease Control (CDC), pool testing consists of combining samples from three to five subjects into one sample and running the test on it. If the sample is positive, then the samples will be tested separately to determine which one is positive. This approach helps screen a large amount of people in a short period of time.	World Health Organization (2020) WHO Vietnam Situation Report. Retrieved October 6 from: https://www.who.int/docs/default-source/wpro---documents/countries/viet-nam/covid-19/vnm-moh-who-covid-19-sitrep4.pdf?sfvrsn=38270192_4
China	**surveillance**	**There are no available guidelines on pooled testing released by China's Center for Disease Control; however, according to news articles, China used pooled testing on the entirety of Wuhan population as mass, indiscriminate testing, using pool size of five to ten individuals in one laboratory test .** June 8 The other way they sped up the process was to use a batch testing method, which groups individual test samples together. Reports suggest they used batches of between five and 10 samples in Wuhan, only carrying out individual tests if a batch proved positive. And as many as 25% of all tests were done using this method. This is an efficient way to	BBC (2020). Coronavirus: China's plan to test everyone in Wuhan. Retrieved October 5, 2020 from: *https://www.bbc.com/news/world-asia-china-52651651*

		test large numbers of people where infection levels are low, as most batches would produce negative results. And it appears to have worked in Wuhan because 97% of local communities across the city reported no positive tests, according to the official data. The authorities said they had found just 300 positive cases (all without symptoms) out of all the tests done, and traced a further 1,174 close contacts of these people.	
South Korea	**screening**	**There are no available guidelines on pooled testing released by the ministry of health and welfare; however, according to news articles, Korea uses pooled testing on an *unspecified population*, using pool size of 10 individuals in one laboratory test.** **April 10:** The government is expected to test combined samples from multiple people of a local cluster at a higher risk of the new coronavirus (COVID-19) simultaneously. The method is called "pool testing" or "sample pooling," which can be useful to prevent infection in high-risk groups, including those in nursing homes with no COVID-19 symptoms. The Korea Centers for Disease Control and Prevention (KCDC) and the Korean Society for Laboratory Medicine (KSLM) said Thursday they completed a pool testing protocol, mixing several samples into one and re-testing individual samples if the combined sample comes out positive. The KCDC and three medical institutions under the KSLM collaborated to design the protocol to be used in the Korean-specific testing environment. The protocol underwent 650 times of testing, the authorities said. Under the pool testing protocol, mixing 10 specimens could still maintain over 96 percent of the sensitivity of testing, compared to individual testing, they added.	Sung-sun. K (2020) Korea considers testing pooled samples for vulnerable groups. Retrieved Oct 7, 2020 from:http://www.koreabiomed.com/news/articleView.html?idxno=7966
WHO		**WHO does not have any existing guidelines on the use of pooled testing.**	

Philippines	**Surveillance** **Screening**	**The Department of Health recommends the use of pooled testing for surveillance testing using pool size of 5, until the prevalence of SARS-CoV-2 infection in the specific community can be determined** The recently released Department Memo 2020-0439 *Omnibus Testing Guidelines on Prevention, Detection, Isolation, Treatment and Reintegration Strategies for COVID-19* includes some recommendations on pooled testing; However, DOH reiterated that such methodology may only be used once results of ongoing pilot testing are positive and favorable, based on the recommendations of experts. Here are the key recommendations of the said guideline. 1. DOH recommends the use of pooled testing among asymptomatic persons belonging to the following targeted population: a) Communities with prevalence rate of 10% or less; b) Surveillance of healthcare workers and all workers in the health facility; c) Workplace testing including market vendors, transport workers, and those considered as economy workers; d) Border testing at ports of entry for inbound foreign travelers and returning residents; e) Overseas deployment of Overseas Filipino Workers (OFWs), and returning OFWs; f) Frontline government workers; g) Locally stranded individuals. 2. DOH does not recommend the use of pooled testing on the following subgroups: a) Individuals with severe/ critical symptoms and relevant history of travel and/or contact; b) Individuals with mild symptoms and relevant history of travel and/or contact, and considered vulnerable. Vulnerable populations include those elderly and with pre-existing medical conditions that predispose them to severe presentation and complications of COVID-19.	Department of Health (2020). Department Memo 2020-0439 Omnibus Testing Guidelines on Prevention, Detection, Isolation, Treatment and Reintegration Strategies for COVID-19

		c) Individuals with mild symptoms, and relevant history of travel and/or contact. d) Individuals with no symptoms but with relevant history of travel and/or contact or high-risk of exposure. These include: i. Contact-traced individuals ii. Healthcare workers, who shall be prioritized for testing to ensure the stability of our healthcare system. iii. Returning Overseas Filipino Workers, who shall immediately be tested at the point of entry; iv. Filipino citizens in a specific locality within the Philippines who have expressed intention to return to their place of residence/home origin (Locally Stranded Individuals) may be tested subject to the existing protocols of the IATF 3. DOH also does not recommend the use of pooled testing, on any population that fall under the following cohorts: e) Symptomatic individuals; f) Recovered patients, regardless if symptomatic or asymptomatic; and g) Close contacts of positive individuals. 4. DOH recommends a pool sample of five, until an accurate prevalence of case with the presence of SARS-CoV-2 is identified in the populations. 5. DOH recommends that laboratories that will conduct pooled testing shall develop guidelines and procedures for the purpose of pooled testing in accordance with the recommendations of local studies on pooled testing. DM 2020-0439 reinforces that these pooled-testing strategies are currently being evaluated and validated. These guidelines shall be further amended as new developments ensue from the studies and pilot implementation.	
EU CDC	**Surveillance**	**EU CDC recommends the use of pooled testing as an alternative approach in the event of reagent shortage. EU CDC recommends that**	European Center for Disease Controls. (2020) Laboratory support for COVID-19 in the EU/EEA.

	Screening	**this strategy be used in prevalence studies or to enhance mild/asymptomatic patients testing, and not when diagnosis is considered critical, due to the possibility of error.** The current test reagent and equipment shortages affect laboratories in all EU/EEA countries and their diagnostic capacity. In the event of severe shortages of reagents, the following alternative approaches have been proposed in the latest risk assessment and may be considered after thorough validation in the individual laboratory: • RT-PCR screening of only a single discriminatory target, using one set of primers, instead of two. Confirmatory testing (and additional sampling if necessary) should be performed only for specimens where the first result is technically not interpretable. • performing a sample preheating step, instead of RNA extraction. This should be followed by the use of an internal control (e.g human gene target) to ensure that sufficient RNA has been included in the RT-PCR reaction. • pooling of low-risk samples from different individuals in one testing run (group testing); this can be used in prevalence studies or to enhance testing of mild/asymptomatic patients. This should not be used in cases where diagnosis is critical, due to the possibility of error. For diagnosis, the samples will need to be retested separately if there is a positive result in the pooled sample. • oropharyngeal and nasopharyngeal swabbing from one patient can be performed with one swab and combined into one diagnostic test. sterile saline can be used instead of viral transport media.	Retrieved October 5, 2020 from: https://www.ecdc.europa.eu/en/novel-coronavirus/laboratory-support

Appendix 4. Critical Appraisal of Included Studies

Assessing the dilution effect of specimen pooling on the sensitivity of SARS-CoV-2 PCR tests

Bateman et al, 2020

1. DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of patients (P), examination (E) used and disease or outcome (O) being diagnosed?

Research Question	
P	Positive specimens for COVID-19 (N=838) *"... the criteria have been repeatedly widened to include symptomatic patients from many populations, as well as individuals tested as part of public health investigations (asymptomatic or symptomatic)" (page 2, par. 2)* *"We included the first 838 SARS-CoV-2 positive specimens tested at WSLH which were tested from March 7 through April 24, 2020." (page 2, par. 3)*
I	Individual Testing of diluted samples with a validated KingFisher Flex extraction platform with the Maxwell® HT Viral TNA Kit, Custom (Cat. #AX2340; Promega) and tested using the CDC RT-PCR assay (N1 and N2 targets)
C	Individual Testing U.S. Centers for Disease Control and Prevention (CDC) 2019-nCoV Real-Time RT-PCR Diagnostic Panel
O	Sensitivity and specificity

2. APPRAISAL OF VALIDITY

2.1. Was the reference standard an acceptable one?

Yes. The reference standard used is the standard test for diagnosis of COVID-19.

2.2. Was "definition" of the index test and the reference standard independent?

No. There is an overlap in terms of the criteria used to determine positive cases. Both tests used the CT value of genes (N1 and N2) to detect positive cases.

"To evaluate the effect of dilution on the sensitivity, we tested specimens undiluted (as the control), diluted 1:5, diluted 1:10, and diluted 1:50." (page 2, par. 7)

"According to the Instructions For Use for the 2019-nCoV Real-Time RT-PCR Diagnostic Panel, if both N1 and N2 are positive, the interpretation is positive, and if either N1 or N2 is positive, the interpretation is inconclusive." (page 2, par. 9)

2.3. Was "performance" of the index test and the reference standard independent?

No. Only selected data based on the frequency distribution of CT values were used for the analysis. The index test was only compared to 100 specimens from the frequency distribution of the CT values of the N1 target of the 838 specimens.

"After plotting the Ct distribution of the N1 target of these 838 specimens, we then selected 100 specimens from this frequency distribution to attain specimens with the same distribution as the first 838." (page 2, par. 3)

2.4. *Was "interpretation of the index test and the reference standard independent?*
Likely no. Since the sample for the index test was selected after knowing the distribution of CT values of the reference standard, it is likely that the interpretation of the index test and reference standard was not independent.

3. APPRAISING THE RESULTS

The likelihood ratios cannot be computed and reconstruction of the 2x2 table cannot be performed since only sensitivity was provided and there is no available data for true positive, false positive, false negative, and true negative.

Dilution	Positive	Sensitivity	95% CI
Undiluted	N1 or N2	99%	(94.6%–99.9%)
	N1 and N2	95%	(88.7%–98.4%)
	N1 only	99%	(94.6%–99.9%)
	N2 only	95%	(88.7%–98.4%)
1:5 dilution	N1 or N2	93%	(86.1%–97.1%)
	N1 and N2	88%	(80.0%–93.6%)
	N1 only	89%	(81.2%–94.4%)
	N2 only	92%	(84.8%–96.5%)
1:10 dilution	N1 or N2	91%	(83.6%–95.8%)
	N1 and N2	83%	(74.2%–89.8%)
	N1 only	89%	(81.2%–94.4%)
	N2 only	85%	(76.5%–91.4%)
1:50 dilution	N1 or N2	81%	(71.9%–88.2%)
	N1 and N2	73%	(63.2%–81.4%)

	N1 only	78%	(68.6%–85.7%)
	N2 only	76%	(66.4%–84.0%)

4. APPLICABILITY

No information. The paper noted that the highest risk symptomatic patients were prioritized during the data collection period. However, asymptomatic patients that are part of public health investigation were also included. Hence, we cannot ascertain the applicability of the paper due to limited information.

5. CONCLUSION

We deem that the paper has **low internal validity** because of the non-independence of the performance, and the interpretation between the index test and the reference standard, as well as the lack of information regarding the definition of the tests.

Large-scale implementation of pooled RNA extraction and RT-PCR for SARS-CoV-2 detection

Ben-Ami et al, 2020

1. DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of patients (P), examination (E) used and disease or outcome (O) being diagnosed?

Yes. The study provides a direct enough answer to the clinical question.

Research Question	
P	Deep nasal and oropharyngeal samples from symptomatic patients and asymptomatic populations for COVID-19 testing *(page 2, column 2, paragraph 1)*
I	Pooled testing by RT-PCR using real-time fluorescent RT-PCR kit (BGI) *(page 2, column 2, paragraph 5)* Three procedures were performed: - 23 pools were tested; each pool had 8 samples with unknown positivity and negativity status *(page 3, column 2, paragraph 1)* - 30 pools were tested using a matrix pooling procedure; each of the 3 matrices had 5 columns and 5 rows with 1 positive sample and 24 negative samples in each matrix *(page 4, figure 2)* - 271 pools were tested; each pool had 8 samples each with unknown positivity and negativity status *(page 3, column 2, paragraph 4)*
C	Individual testing by RT-PCR using real-time fluorescent RT-PCR kit (BGI)

	(page 2, column 2, paragraph 5; page 2, column 2, paragraph 1)
o	Number of true positives, true negatives, false positives, and false negatives *(page 3, table 2; page 4, figure 1; page 4, figure 2)*

2. APPRAISAL OF VALIDITY

2.1. *Was the reference standard an acceptable one?*

Yes. The reference standard used in the study is individual testing by RT-PCR, which remains to be the standard test for confirming COVID-19.

2.2. *Was "definition" of the index test and the reference standard independent?*

No. Both the index test and the reference standard had the same definition for samples that were positive and negative for SARS-CoV-2.

> *"According to the clinical guidelines of the Israeli Ministry of Health at the time experiments were conducted, a sample was defined as positive if the viral genome was detected at threshold cycle (Ct) values ≤ 35, as indeterminate at Ct values >35 and ≤ 38, and as negative at Ct values >38." (page 2, column 2, paragraph 2)*
>
> *"Therefore, all pools detected with Ct ≤ 39 were retested (see Table 2, batch 3), while maintaining standard criteria for the individual tests when retesting." (page 3, column 1, paragraph 2)*

2.3. *Was "performance" of the index test and the reference standard independent?*

Three methods of the index test were presented in the study – Dorfman pooled testing on 184 samples, matrix pooled testing on 75 samples, and Dorfman pooled testing on 2168 samples.

Yes. The index test and the reference standard were performed in all 184 individual samples and in all 23 pools of 8 samples.

> *"We tested the pooling of 184 samples into 23 pools of eight samples each, and also tested in parallel each sample individually." (page 3, column 2, paragraph 1)*

Yes. The index test and the reference standard were performed in all 75 samples and in all 30 pools of 5 samples.

> *"Twenty-five samples sorted in a 5 x 5 matrix and each row and each column is pooled into a total of ten pools, on which RNA extraction, reverse transcription and qPCR are performed." (page 4, figure 2)*
>
> *"… 25 lysates that were previously tested individually" (page 4, figure 2)*

No. The index test and the reference standard were not performed in all 2168 samples and in the 271 pools of 8 samples. Individual testing was only performed on each sample of pools that were defined as positive in pooled testing.

> *"In the first three batches run at the HMC (Table 2) we tested 2168 samples by pooling, using 311 RNA extraction and RT-PCR reactions (a mere 14% of kits that would have been used in the full individual testing, an increase of sevenfold in throughput)." (page 3, column 2, paragraph 4)*

2.4. Was "interpretation" of the index test and the reference standard independent?

Three methods of the index test were presented in the study – Dorfman pooled testing on 184 samples, matrix pooled testing on 75 samples, and Dorfman pooled testing on 2168 samples.

No information. Although both the individual tests and pooled tests were done in parallel, there was no mention that the researchers were blinded when interpreting the results of the index test and reference standard.

No. The researchers formed the pools and matrices based on the results of the individual testing. They already had an idea of the expected results of the matrix pooled testing.

> *"Three 5 x 5 pool matrices were generated (30 pools from 75 lysates). Each matrix (25 lysates that were previously tested individually) included a single lysate positive for SARS-COV-2. As expected, only six pools (one row and one column per matrix) were positive for SARS-COV-2, while 24 pools had threshold cycle (Ct) > 40 (Undetected)." (page 4, figure 2)*

No. The researchers were informed of the results of the pooled testing before they proceeded to individual testing of each sample of positive pools.

> *"A negative result implies that all samples in the pool are negative, while a positive result implies that at least one sample in the pool is positive. In the second stage, the samples of each pool that tested positive are individually tested." (page 2, column 2, paragraph 3)*

3. APPRAISING THE RESULTS

What were the likelihood ratios of the various test results?

Three methods of the index test were presented in the study – Dorfman pooled testing on 184 samples, matrix pooled testing on 75 samples, and Dorfman pooled testing on 2168 samples.

The figure below was the data presented by the researchers from the comparison of pooled testing and individual testing on 184 samples and 23 pools of 8 samples.

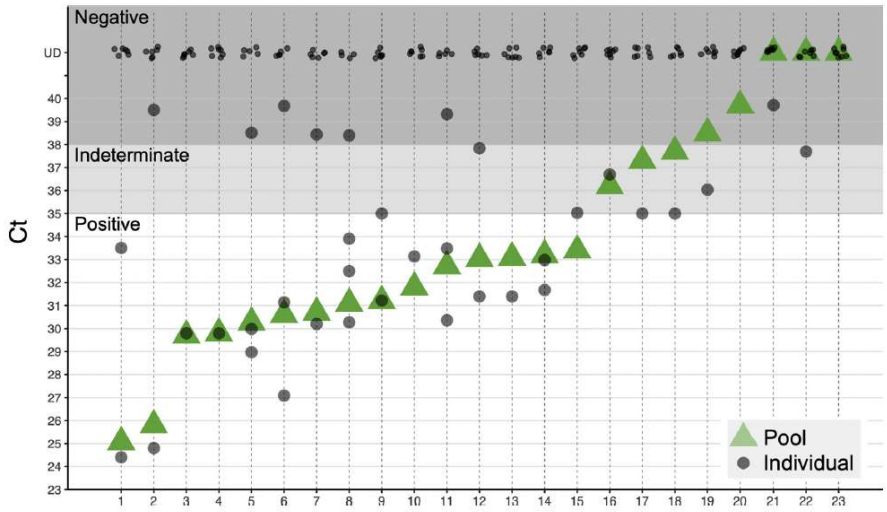

Fig. 1. Pooling eight lysates retains clinical sensitivity. Shown are results of 23 pooling experiments, with eight lysates in each pool; 15 pools with positive samples indeed come up positive (pools 1–15), three pools without positive samples come up negative (pools 20, 21, 23) and four out of five pools containing a single indeterminate sample detected as indeterminate (pools 16, 17, 18, 19, 22); Pools containing one or two samples with low amount of SARS-CoV-2 are detected at a similar Ct (pools 9–18), showing clinical sensitivity is retained and the risk of false negatives is minimal. Ct, threshold cycle.

From the presented data, we constructed a 3 x 3 table to aid in the independent calculation of values for sensitivity, specificity, positive predictive values, negative predictive values, and likelihood ratios.

Pooled Testing	Individual Testing			Total
	(+)	Indeterminate	(-)	
(+)	15	0	0	15
Indeterminate	0	4	0	4
(-)	0	1	3	4
Total	15	5	3	23

The researchers arrived at indeterminate results for both individual testing and pooled testing because of the established definitions for positive, indeterminate, and negative samples based on the Ct value. They mentioned that indeterminate samples should be retested; however, they did not include the results from retesting. Therefore, we have come up with different scenarios to integrate these indeterminate results into 2 x 2 or 2 x n tables.

The 2 x 2 table below shows the results of the method if the indeterminate results were not regarded as positive or negative samples.

Pooled Testing	Individual Testing		Total
	(+)	(-)	
(+)	15	0	15
(-)	0	3	3
Total	15	3	18

Sensitivity of the Dorfman pooled testing is 100.00% (95%CI: 78.20%, 100.00%); specificity is 100.00% (95%CI: 29.24%, 100.00%); positive predictive value is 100.00%; negative predictive value is 100.00%; positive likelihood ratio is 4.00 (95%CI: 0.73, 21.84); and negative likelihood ratio is

0.06 (95%CI: 0.01, 0.42). The values of the likelihood ratios indicate that the test had moderately positive and strongly negative LRs, making it useful for ruling out COVID-19.

The 2 x 3 table below shows the results of the method if the indeterminate results were regarded as positive in the reference standard.

Pooled Testing	Individual Testing		Total
	(+)	(-)	
(+)	15	0	15
Indeterminate	4	0	4
(-)	1	3	4
Total	20	3	23

The positive likelihood ratio is 2.25, intermediate likelihood ratio is 0.60, and the negative likelihood ratio is 0.05. In this case, the values of the likelihood ratios indicate that the test has weakly positive, weakly intermediate, and weakly negative LRs, which means the test will not change the likelihood of COVID-19 by a considerable level.

Lastly, the 2 x 3 table below shows the results of the method if the indeterminate results were regarded as negative in the reference standard.

Pooled Testing	Individual Testing		Total
	(+)	(-)	
(+)	15	0	15
Indeterminate	0	4	4
(-)	0	4	4
Total	15	8	23

The positive likelihood ratio is 8.00, intermediate likelihood ratio is 0.13, and the negative likelihood ratio is 0.13. In this case, the values of the likelihood ratios indicate that the test has moderately positive, moderately intermediate, and moderately negative LRs, which means the test will not change the likelihood of COVID-19 by a considerable level.

The figure below describes the method and results for the matrix pooled testing.

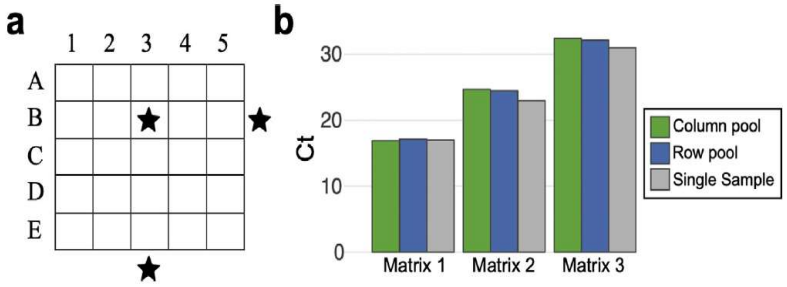

Fig. 2. Matrix pooling. (a) Scheme for 5 × 5 matrix pooling. Twenty-five samples sorted in a 5 × 5 matrix and each row and each column is pooled into a total of ten pools, on which RNA extraction, reverse transcription and qPCR are performed. In this illustration row B and column 3 are positive (black stars), hence sample B3 is the only positive sample. If more than one row and one column are positive then all samples in intersections need to be retested, as some may be negative. (b) Three 5 × 5 pool matrices were generated (30 pools from 75 lysates). Each matrix (25 lysates that were previously tested individually) included a single lysate positive for SARS-COV-2. As expected, only six pools (one row and one column per matrix) were positive for SARS-COV-2, while 24 pools had threshold cycle (Ct) > 40 (Undetected). Reverse transcription-PCR Ct values of positive pools were nearly identical in the column pool (green) and the row pool (blue), and similar to the values of the individual test of the positive sample (grey).

From these results, the 2 x 2 table was constructed to aid in the independent calculation of values for sensitivity, specificity, positive predictive values, negative predictive values, and likelihood ratios.

Pooled Testing	Individual Testing		Total
	(+)	(-)	
(+)	6	0	6
(-)	0	24	24
Total	6	24	30

Sensitivity of the matrix pooled testing is 100.00% (95%CI: 54.07%, 100.00%); specificity is 100.00% (95%CI: 85.75%, 100.00%); positive predictive value is 100.00%; negative predictive value is 100.00%; positive likelihood ratio is 25.00 (95%CI: 3.66, 170.59); and negative likelihood ratio is 0.14 (95%CI: 0.02, 0.88). The values of the likelihood ratios indicate that the test had strongly positive and moderately negative LRs, making it useful for ruling in COVID-19.

The following are the results given by the researchers from their testing on 2168 samples.

Table 2
Three pooled tests run at the Hadassah Medical Centre

Batch	No. of samples	No. of pools	No. of positive pools	Ct of positive pools	No. of positive individual samples in the positive pools	Ct of positive individual samples	No. of total tests	No. of tests saved, compared to single sample tests (%)
1	720	90	1	21.8	1	19.4	98	622 (86.4%)
2	720	90	0	—	—	—	90	630 (87.5%)
3	728	91	3 positives, 1 indeterminate	25.39 37.44 35.26 38.67 indeterminate	2 1 1 0	22.05, 28.72 37.67[a] 34.66 38.24 (determined as negative)	123	605 (83.1%)

Ct, cycle threshold.
[a] According to Hadassah Medical Centre's protocol for indeterminate values, reverse transcription-PCR was repeated with a different kit, and eventually was determined as positive.

From these data, the 2 x 2 table was constructed to aid in the independent calculation of values for sensitivity, specificity, positive predictive values, negative predictive values, and likelihood ratios.

Pooled Testing	Individual Testing		Total
	(+)	(-)	
(+)	4	1	5
(-)	0	266	266
Total	4	267	271

After independent calculations, we arrived at a sensitivity of 100.00% (95%CI: 39.76%, 100.00%); a specificity of 99.63% (95%CI: 97.93% to 99.99%); positive predictive value of 80.00% (95%CI: 36.12%, 96.59%); negative predictive value of 100.00%; positive likelihood ratio of 267.00 (95%CI: 37.75 to 1888.57); and negative likelihood ratio of 0.20 (95%CI: 0.03, 1.16). The values of the likelihood ratios indicate that the test had strongly positive and moderately negative LRs, making it useful for ruling in and COVID-19.

4. APPLICABILITY

Are there biologic issues that may affect accuracy of the test? (Consider the influence of sex, co-morbidity, race, age, and pathology). Are there socio-economic issues that may affect accuracy of the test?

No information. Although they mentioned that samples came from symptomatic patients and asymptomatic individuals, the number of samples with symptoms was not given and presence or absence of symptoms was not included in their analysis. There were no other patient characteristics mentioned in the study; hence, applicability cannot be determined.

5. CONCLUSION

The study has low internal validity due to the non-independence of the definition of the index test and reference standard. Bias may also come from the non-independence of the interpretation of the results, given that in 2 of their 3 methods, the researchers were already informed of the results before proceeding to the next stage of their tests; and in their validation of the Dorfman pooling, there was no mention of blinding the interpreters of the outcomes. In all three methods, independent calculations showed that pooled testing had varied likelihood ratios for each method that the researchers had done.

Sample pooling for SARS-COV-2 RT-PCR screening

De Salazar et al, 2020

1. DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of patients (P), examination (E) used and disease or outcome (O) being diagnosed?

Research Question	
P	nasopharyngeal swabs (n =3519) collected from patients or health professionals

	"*Nasopharyngeal and pharyngeal swabs were collected at the same time, and both swabs were placed in the same tube with transport media*." (page 2, par. 6)
I	Pooled testing of nine or ten samples using Mixed Brands such as Maxwell RSC Viral Total Nucleic Acid (Promega), m2000sp (Abbot), eMAG (bioM_erieux), STARMag (Seegene), MagMAX Viral RNA Isolation Kit (Thermo Fisher Scientific), cobas SARS-CoV-2 test (Roche), Viasure SARS-CoV-2 Real Time PCR (CerTest) , TaqMan 2019-nCoV Assay Kit v1 (Thermo Fisher Scientific), Allplex 2019-nCoV Assay (Seegene), Light Mix E gene (Roche) *"Nine or ten individual samples were pooled."* (page 2, par. 7)
C	Individual Testing using RT-PCR using Mixed Brands such as Maxwell RSC Viral Total Nucleic Acid (Promega), m2000sp (Abbot), eMAG (bioM_erieux), STARMag (Seegene), MagMAX Viral RNA Isolation Kit (Thermo Fisher Scientific), cobas SARS-CoV-2 test (Roche), Viasure SARS-CoV-2 Real Time PCR (CerTest) , TaqMan 2019-nCoV Assay Kit v1 (Thermo Fisher Scientific), Allplex 2019-nCoV Assay (Seegene), Light Mix E gene (Roche) "...screening was performed using RT-PCR targeting the same target as for individual samples." (page 2, par. 7)
O	Sensitivity and specificity

2. APPRAISAL OF VALIDITY

2.1. Was the reference standard an acceptable one?

Yes. The reference standard used is the standard practice for diagnosis of COVID-19.

2.2. Was "definition" of the index test and the reference standard independent?

No. The criteria for declaring an individual sample positive overlaps with the criteria of positive result using pooled testing since the individual testing will be performed if a pool is positive.

"Nine or ten individual samples were pooled, and screening was performed using RT-PCR targeting the same target as for individual samples." (page 2, par. 7)

2.3. Was "performance" of the index test and the reference standard independent?

Yes. All collected specimens were tested for both index and reference tests. However, the paper noted that a pool of 10 samples were invalid for unmentioned reasons.

"For both individual testing and pooled analysis, samples were inactivated 1:1 in lysis buffer and processed according to the existing methodology in each laboratory" (page 2, par. 8)

"We found that 253 pools, made up of 2519 samples, were negative (242 pools of ten samples and 11 pools of nine samples); and 99 pools, made up of 990 samples, were positive (99 pools of

ten samples). One pool comprising ten samples was invalid." (page 2, par. 13)

2.4. *Was "interpretation of the index test and the reference standard independent?*
Yes. It is mentioned that the samples for pooling were selected randomly according to availability of each site. This ascertains the independence of pooled testing with the results of the reference standard in terms of interpretation.

"Pooled testing was performed at each of the participating sites in the study. Pooling was performed by hand, after inactivation of each sample that made up part of the pool. Samples for pooling were selected randomly, according to availability at each site during the study period." (page 2, par. 7)

3. APPRAISING THE RESULTS

The reported results are divided by the inclusion of discordances classified as follows:

"Major discordance was defined as a negative pool result when at least one of the individual samples showed cycle threshold (Ct) values of <35 for one or more SARS-CoV-2 genes. Minor discordance occurred when at least one individual sample had Ct > 35 in one or two of the SARS-CoV-2 genes assayed and the pool scored negative." (page 2, par. 9)

Test Result (Major Discordances)	Reference Standard		
	Positive	Negative	Row Total
Positive	234	0	234
Negative	7	3268	3275
Column Total	241	3268	3509*

**One pool comprising ten samples was considered invalid.*

Statistic	Value	95% CI
Sensitivity	97.10%	94.11% to 98.82%
Specificity	100.00%	99.89% to 100.00%
PPV	100.00%	
NPV	99.79%	99.56% to 99.90%
Accuracy	99.80%	99.59% to 99.92%
Kappa	0.984	

Test Result (All Discordances)	Reference Standard		
	Positive	Negative	Row Total
Positive	206	0	206
Negative	35	3268	3303
Column Total	241	3268	3509*

**One pool comprising ten samples was considered invalid.*

Statistic	Value	95% CI
Sensitivity	85.48%	80.39% to 89.67%
Specificity	100.00%	99.89% to 100.00%
PPV	100.00%	
NPV	98.94%	98.57% to 99.22%

Accuracy	99.00%	98.62% to 99.30%
Kappa	0.916	

To validate the results reported by the authors, independent calculations were done and presented below:

Major Discordance

Statistic	Value	95% CI
Sensitivity	97.10%	94.11% to 98.82%
Specificity	100.00%	99.89% to 100.00%
Positive Likelihood Ratio*	3174.05	447.17 to 22529.60
Negative Likelihood Ratio	0.03	0.01 to 0.06
Disease prevalence	6.87%	6.05% to 7.76%
PPV	100.00%	
NPV	99.79%	99.56% to 99.90%
Accuracy	99.80%	99.59% to 99.92%

All Discordances

Statistic	Value	95% CI
Sensitivity	85.48%	80.39% to 89.67%
Specificity	100.00%	99.89% to 100.00%
Positive Likelihood Ratio*	2794.25	393.44 to 19845.05
Negative Likelihood Ratio	0.15	0.11 to 0.20
Disease prevalence	6.87%	6.05% to 7.76%
PPV	100.00%	
NPV	98.94%	98.57% to 99.22%
Accuracy	99.00%	98.62% to 99.30%

*Using imputed 2x2 table

Based on our independent calculations, the computed sensitivity, specificity, PPV, and NPV matched the reported values in the study for both major discordances and all discordances.

4. APPLICABILITY

No information. The paper did not report any details regarding the patients. Thus, applicability cannot be assessed.

5. CONCLUSION

We conclude that the paper has **moderate internal validity**. The non-independence of the definition between the index test and the reference standard is due to the actual design of the index and does not result in biased estimates. The results presented in the study were consistent with our independent calculation. These results showed a strongly positive likelihood ratio which gives us confidence that it can be used to rule in the disease.

"Sample pooling of RNA extracts to speed up SARS-CoV-2 diagnosis using CDC FDA EUA RT-qPCR kit"

Freire-Paspuel et al, 2020

1. DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of patients (P), examination (E) used and disease or outcome (O) being diagnosed?

Yes. The study provides a direct enough answer to the clinical question.

Research Question	
P	114 nasopharyngeal swab specimens from individuals selected during SARS-CoV-2 surveillance in Galapagos Islands *(page 1, column 2, paragraph 3)*
I	Pooled testing by RT-PCR using 2019-nCoV CDC EUA kit (IDT, USA) *(page 1, column 2, paragraph 1)* 38 pools were tested; each pool had 1 positive and 2 negative samples *(page 2, column 1, paragraph 2)*
C	Individual testing by RT-PCR using 2019-nCoV CDC EUA kit (IDT, USA) *(page 2, column 1, paragraph 1)*
O	Sensitivity *(page 2, column 1, paragraph 5)*

2. APPRAISAL OF VALIDITY

2.1. Was the reference standard an acceptable one?

Yes. The reference standard used in the study is the testing of individual samples using RT-PCR. Individual testing using RT-PCR remains to be the standard test for confirming COVID-19.

2.2. Was "definition" of the index test and the reference standard independent?

No. There is an overlap in the criteria for considering a sample positive in the index test and reference standard. The researchers considered both an individual sample and a pool positive when the Ct values for the N1 and N2 probes are 40 or smaller.

"According to CDC protocol, a sample is considered positive when Ct values for N1 and N2 are 40 or smaller (Interim Guidelines for Collecting, Handling and Testing Clinical Specimens from Persons for Coronavirus Disease, 2019)." (page 2, paragraph 2)

2.3. Was "performance" of the index test and the reference standard independent?

No. Not all 114 clinical specimens were subjected to both the index test and reference standard. The samples were first tested individually, then were pooled based on the number of individually positive and individually negative tests. They formed pools of three, with one positive sample and 2 negative samples, with the exception of one pool that had three positive samples. There were 38 positive specimens but only 36 pools, so 6 negative samples were not used for pool testing.

"The 114 samples included on this study were tested for SARS-CoV2 following the standard protocol described on the methods. 38 of this samples tested positive for N1 and N2 viral probes, and Ct values are detailed on Table 1. These 114 samples were also pool on group of 3 samples after RNA extraction and prior to RT-PCR following the pool protocol detailed on the methods. The Ct values for N1 and N2 for the 38 positives samples on the RT-PCR pool reaction are detailed on Table 1. All positive pools included only a positive sample and two negative samples, with exception of a pool that included 3 positives samples (samples 10, 11 and 12 on Table 1)." (page 2, paragraph 5)

2.4. Was "interpretation of the index test and the reference standard independent?

No. The researchers were already informed of the results from the individual testing, which became the basis of their pool size and composition. The researchers already had an idea of the expected results of each pool.

"We performed this protocol for 38 SARS-CoV-2 positive and 76 negative samples individually, but also pooling one positive sample with two negative samples at the RT-PCR reaction mix." (page 2, paragraph 2)

"All positive pools included only a positive sample and two negative samples, with exception of a pool that included 3 positives samples (samples 10, 11 and 12 on Table 1)." (page 2, paragraph 5)

3. APPRAISING THE RESULTS

What were the likelihood ratios of the various test results?

The likelihood ratio for pooled testing cannot be computed because the pooling scheme used in the study did not include negative pools so the number of false positive tests and true negative tests cannot be determined. Hence, the values for specificity, positive predictive value, and negative predictive values cannot be computed.

Below are the results presented by the researchers from the pooled and individual RT-PCR tests performed. The results given are in the form of Ct values, from which a sample is considered positive when the Ct values for N1 and N2 are less than or equal to 40. From this, they claimed

that the sensitivity obtained for the 3 sample pooling protocol for 36 pools was 100%. However, the researchers did not report the inconclusive results for n=3 and n=7 where the Ct value for N2 was greater than 40; instead, they considered these values as positive. In the CDC protocol for the kits that were used in the study, this statement was included under interpretation of results:

> *"When all controls exhibit the expected performance and the cycle threshold growth curve for any one marker (N1 or N2, but not both markers) crosses the threshold line within 40.00 cycles (< 40.00 Ct) the result is inconclusive."*

Table 1
Ct values for N1 and N2 for single sample and 3 samples pool RT-qPCR protocol for the 38 SARS-CoV-2 positive samples included on the study.

n	N1 Pool PCR Ct value	N1 Single PCR Ct value	N2 Pool PCR Ct value	N2 Single PCR Ct value
1	30.09	30.60	33.86	32.28
2	32.34	31.02	36.47	32.46
3	36.84	35.87	40.56	38.72
4	36.13	34.43	38.64	35.84
5	33.19	32.42	36.50	35.08
6	34.74	33.57	38.37	36.03
7	34.74	35.91	38.37	41.72
8	30.27	29.71	33.30	31.39
9	28.26	27.15	32.21	29.33
10	27.09	25.15	30.37	26.59
11	27.09	36.26	30.37	39.87
12	27.09	35.45	30.37	38.08
13	34.61	34.04	37.12	35.83
14	35.32	34.05	35.51	34.43
15	36.69	35.49	36.29	36.24
16	37.21	36.35	36.43	36.45
17	28.01	27.01	30.90	28.19
18	23.84	22.61	25.50	24.02
19	34.65	33.90	36.85	37.34
20	35.73	34.76	36.18	36.80
21	36.52	35.81	37.07	37.05
22	30.10	28.80	30.73	29.98
23	29.76	28.27	31.19	29.82
24	29.48	28.11	31.68	29.54
25	31.86	30.64	33.48	32.08
26	30.66	32.41	32.89	36.46
27	27.77	27.45	28.61	32.80
28	24.18	26.70	25.67	32.92
29	33.36	32.76	35.22	37.49
30	29.89	31.34	31.87	36.15
31	26.63	27.30	29.07	31.96
32	31.38	30.40	34.03	35.31
33	25.57	26.06	27.12	30.08
34	31.60	33.11	33.88	38.40
35	27.97	29.86	29.39	34.04
36	24.33	26.86	25.67	32.46
37	30.70	31.90	33.00	35.00
38	38.30	35.81	38.72	37.05

Table 2
Average Ct values (mean ± SD) for N1 and N2 for samples tested on single sample and 3 samples pool RT-qPCR protocol.

	N1		N2	
	Pool PCR Ct value	Single PCR Ct value	Pool PCR Ct value	Single PCR Ct value
Mean ± SD	31.16 ± 4.04	31.30 ± 3.69	33.25 ± 3.96	34.09 ± 3.83

4. APPLICABILITY

Are there biologic issues that may affect accuracy of the test? (Consider the influence of sex, co-morbidity, race, age, and pathology). Are there socio-economic issues that may affect accuracy of the test?

No information. The researchers did not mention any patient characteristics considered in the study; hence, applicability cannot be assessed.

5. CONCLUSION

The study has low internal validity given that some negative samples were not subjected to pooled testing without explanation and that the researchers have already been informed of the results of individual testing before conducting pooled testing. The definitions of the index test and the reference standard were also non-independent. There is also a bias in reporting the results because one pooled sample and one individual sample should have been marked as inconclusive. Instead, they were considered as positive and were included in the computation for sensitivity.

Pooled RNA sample reverse transcriptase real time PCR assay for SARS CoV-2 infection: A reliable, faster and economical method

Gupta et al, 2020

1. DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of patients (P), examination (E) used and disease or outcome (O) being diagnosed?

Yes. The study provides a direct enough answer to the clinical question.

Research Question	
P	280 combined nasopharyngeal and oropharyngeal swabs for COVID-19 testing *(page 2, paragraph 2; page 4, paragraph 3)*
I	Pooled testing by RT-PCR using LightMix SarbecoV E-gene (TIB MOLBIOL) *(page 2, paragraph 3; page 3, paragraph 1)* 35 pools were tested; each pool had 8 samples with unknown positivity or negativity status *(page 4, paragraph 3)*
C	Individual testing by RT-PCR using LightMix SarbecoV E-gene (TIB MOLBIOL) followed by LightMix Modular SARS-CoV-2 RdRP (TIB MOLBIOL) for confirmatory testing *(page 2, paragraph 3)*
O	Sensitivity, specificity, positive predictive value, negative predictive value

(page 4, paragraph 3)

2. APPRAISAL OF VALIDITY

2.1. Was the reference standard an acceptable one?

Yes. The reference standard used in the study is individual testing by RT-PCR, which remains to be the standard test for confirming COVID-19.

2.2. Was "definition" of the index test and the reference standard independent?

No. There is an overlap in the criteria for giving a sample a positive result. For the index test, a pooled sample is positive if the Ct value for the detection of the E gene is ≤ 40. For the reference standard, an individual sample is considered to be positive if both the E gene and the RdRP gene were detected with Ct value of ≤ 40.

> *"Results were seen on the ABI software and each reaction was read for E gene after confirmation of the performance of EAC as well as positive control and negative control results. Ct value for each positive test was recorded and as per the WHO criteria, sample with Ct value ≤ 40 were considered as positive. All initial E gene positive were confirmed as positive if RdRP gene was also detected with Ct value ≤ 40." (page 3, paragraph 1)*

2.3. *Was "performance" of the index test and the reference standard independent?*

No. Pooled and individual screening of the E gene using RT-PCR were performed on all 280 samples; however, only samples that were positive for the E gene were individually tested for the presence of the RdRP gene.

> *"All samples that were screened positive for E gene were confirmed by performance of RT-qPCR for the detection of specific RdRP gene of SARS-CoV-2 using LightMix Modular SARS-CoV-2 RdRP (TIB MOLBIOL) using similar PCR conditions as described above." (page 2, paragraph 3)*

2.4. Was "interpretation of the index test and the reference standard independent?

Yes. The researchers performed both the index test and the reference standard as soon as the RNA elutes arrived in the laboratory. Both the individual testing and pooled testing were done in the same plate.

> *"Subsequently, prospectively as RNA elutes were received in the laboratory, they were randomly pooled into pools of 8 RNA elutes on a 96 well plate as well as ID test as shown in Fig 1. Both ID and pooled RNA RT-qPCR for the screening E gene was done in the same plate (Fig 1)." (page 3, paragraph 1)*

> *"Our study was different from this published study as we did the pooling prospectively, as and when the samples were received without knowing the results of the test and both pool test and ID test were done in the same PCR run. Therefore, in our study pooling in a real-life situation where number of positive samples in a pool cannot be predicted was studied." (page 6, paragraph 2)*

3. APPRAISING THE RESULTS

What were the likelihood ratios of the various test results?

The data below were presented by the researchers to show the ability to detect the E gene in pooled testing and in individual testing.

Table 2. Comparison between the mean Ct values of individual test and pool test in different pool combinations.

Pool of 8 RNA samples		E gene detection			
Pool combinations	Total Number of Pools	Not detected in number of pools	Mean of ID test Ct value (± SD)	Mean of Pool test Ct value (± SD)	P value
1 positive + 7 negative	11	1	33.1 (±3.3)	35.6 (±3.3)	0.04
2 positive + 6 negative	5	0	30.4 (±3.7)	33(±3.1)	0.13
3 positive + 5 negative	5	0	34.2 (±3.7)	34.9(±2.6)	0.58
4 positive + 4 negative	1	0	31.7 (± 4.6)	31.7	-
8 negative	13	13	NA		-
Total	**35**	**14**			

Ct: cycle threshold, RNA: ribonucleic acid, ID: individual

From the given data, we constructed the 2 x 2 table to aid in independent calculations for sensitivity, specificity, positive predictive value, negative predictive value, and likelihood ratios.

Pooled Testing	Individual Testing		Total
	(+)	(-)	
(+)	21	0	21
(-)	1	13	14
Total	22	13	35

Sensitivity of the pooled test for detection of E gene is 95.45% (95% CI: 77.16%, 99.88%); specificity is 100.00% (95% CI: 75.29%, 100.00%); positive predictive value is 100.00%; and negative predictive value is 92.86% (95% CI: 65.70%, 98.88%). These values are similar to the performance values presented by the researchers.

Pooled testing has a strongly positive likelihood ratio of 13.36 (95% CI: 2.02 to 88.54) and a strongly negative likelihood ratio of 0.05 (95% CI: 0.01, 0.33).

4. APPLICABILITY

Are there biologic issues that may affect accuracy of the test? (Consider the influence of sex, co-morbidity, race, age, and pathology). Are there socio-economic issues that may affect accuracy of the test?

No information. There were no patient characteristics presented in the study; hence, applicability cannot be assessed.

5. CONCLUSION

The study has moderate internal validity because of non-performance of the confirmatory testing in pooled samples. Comparison of the performance of pooled testing against individual testing was only based on the detection of the E gene. There is also an overlap in the criteria for considering a sample positive. Independent calculations of performance values were similar to

those reported. Likelihood ratios are both strongly positive and strongly negative which indicates that pooled testing has a good performance based on the reported data.

Optimal size of sample pooling for RNA pool testing: An avant-garde for scaling up severe acute respiratory syndrome coronavirus-2 testing

Khodare, 2020

1. DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of patients (P), examination (E) used and disease or outcome (O) being diagnosed?

Yes. The study provides a direct enough answer to the clinical question of the review.

Research Question	
P	**Positive and negative nasopharyngeal and oropharyngeal sample elutes** **Positive samples= 7** **Negative samples = 48** *"We used repeatedly tested positive clinical sample elutes having different levels of SARS CoV 2 RNA and negative sample elutes to prepare seven series of 11 pools each, having pool sizes ranging from 2 to 48 samples to estimate the optimal pool size." (page 1, materials and methods)* *"Combined nasopharyngeal and oropharyngeal swabs were collected and transported in viral transport media (VTM) maintaining the proper cold chain and sent to the virology laboratory of the Institute of Liver and biliary sciences, New Delhi, India." (page 2, materials and methods, paragraph 5)*
I	**2, 4, 6, 8, 10, 12, 16, 20, 24, 32, 48 sample pools (1 positive, n-1 negative samples done in double dilution series); LightMix® SarbecoV E-gene (TIB MOLBIOL) and LightMix® Modular SARS-CoV-2 RdRP (TIB MOLBIOL)** *The 5 µl of the extracted RNA elute/sample was subjected to RT-qPCR for the qualitative detection of SARSCoV-2 RNA utilising with AgPath-IDTM One-Step RT-PCR Reagents (Thermo Fisher Scientific) using an Applied biosystem (ABI) 7500 Real-Time PCR system*

	(Thermo Fisher Scientific) and LightMix® SarbecoV E-gene (TIB MOLBIOL). (page 3, paragraph 1) *All samples that were screened positive for the E gene were confirmed by the performance of RT-qPCR for the detection of specific RdRp gene of SARS-CoV-2 using LightMix® Modular SARS-CoV-2 RdRP (TIB MOLBIOL) using similar PCR conditions as described above. (page 3, paragraph 2)*
C	**For the positive samples: LightMix® SarbecoV E-gene (TIB MOLBIOL) and LightMix® Modular SARS-CoV-2 RdRP (TIB MOLBIOL)** **For the negative samples: Unsure how negative samples were confirmed to be negative** *In this study, clinical samples previously established as positive for the SARS CoV-2 virus were chosen to determine whether they can detectable when their elutes mixed with negative samples elutes in different dilutions. Positive sample elutes for SARS CoV-2 with a different range of Ct values of E gene target were chosen, serially diluted with negative sample elutes, and RT-qPCR was performed. (page 4, paragraph 1)* *The frequency of Ct values (E gene only) distribution of individually tested samples was derived from 227 SARS CoV 2 E and RdRp gene positive samples detected from the 1st of March to 30th April 2020 using the same PCR reagents used for pool testing. (page 4, paragraph 4)*
O	**Sensitivity, Specificity, Negative Predictive Value, Positive Predictive Value, Efficiency of sample pooling, Optimal pool size** *(page 4-5)*

2. APPRAISAL OF VALIDITY

2.1. Was the reference standard an acceptable one?

Yes for positive samples, No for negative samples. The reference standard used for the positive samples were LightMix® SarbecoV E-gene (TIB MOLBIOL) and LightMix® Modular SARS-CoV-2 RdRP (TIB MOLBIOL) which employs RT-PCR to determine the presence of the virus. RT-PCR is still considered the gold standard in diagnosis of COVID-19, which makes it an acceptable reference standard. There are some issues with the validity of the negative samples used. It was not stated whether these samples were negative samples based on a negative result from an RT-PCR test or if these samples are taken from a specimen bank pre-pandemic, and assumed to be negative ipso facto.

2.2. Was "definition" of the index test and the reference standard independent?

No information. The criteria for the index test and reference standard were not defined by the study authors.

2.3. Was "performance" of the index test and the reference standard independent?

No. It was unclear whether the reference standard tests were performed on the negative samples, or if these samples were just assumed to be negative.

The frequency of Ct values (E gene only) distribution of individually tested samples was derived from 227 SARS CoV 2 E and RdRp gene positive samples detected from the 1st of March to 30th April 2020 using the same PCR reagents used for pool testing. (page 4, paragraph 4)

2.4. *Was "interpretation of the index test and the reference standard independent?*
No. There was no blinding done in the interpretation of the results because positive samples were deliberately included in negative pools to make different dilutions. In addition, the study employed retrospective collection of samples. Diagnostic status of the samples were already known before it was subjected to the index test.

"Arbitrarily 7 twice tested SARS CoV 2 E and RdRp gene positive RNA elutes with varying cycle threshold values and 48 twice tested negative elutes for SARS CoV 2 E and RdRp gene tested by utilising AgPath- IDTM One-Step RT-PCR Reagents were selected. A total of 48 negative sample elutes were used to make eight series of 11 pools each (total of 77 pools) in an equal volume (3 µl each) of 2, 4, 6, 8, 10, 12, 16, 20, 24, 32 and 48 sample elutes. Each of the 11 pools was mixed with 1 positive elute to make its seven dilutions
series of 1:2, 1:4, 1:6, 1:8, 1:10, 1:12, 1:16, 1:20, 1:24, 1:32 and 1:48 dilutions." (page 3, paragraph 3)

3. APPRAISING THE RESULTS

The study computed sensitivity by taking into consideration the distribution of particular Ct values in the results obtained during routine diagnostic testing of individual samples received. Seven ranges of Ct values for pooling were defined. For instance, in the 6 sample pools, only the first six Ct values were positive in the pooling strategy, hence the study authors obtained how much of individual positive samples during the pandemic were distributed within the ct value range of the positive pools, hence the sensitivity of 97.8%.

Dilution	Sensitivity (95% CI)	Specificity	PPV	NPV
1:2	100% (98.4-100)	100%	100%	100 %
1:4	100% (98.4-100)	100%	100%	100 %
1:6	97.80% (94.9-99.3)	100%	100%	97.2 %
1:8	84.58% (79.2-89)	100%	100%	96.08 %
1:10	64.76% (58.2-70.9)	100%	100%	95.45 %
1:12	64.76% (58.2-70.9)	100%	100%	96.25 %
1:16	64.76% (58.2-70.9)	100%	100%	97.22 %
1:20	64.76% (58.2-70.9)	100%	100%	97.79 %
1:24	64.76% (58.2-70.9)	100%	100%	98.17 %
1:32	64.76% (58.2-70.9)	100%	100%	98.64 %
1:48	41.41% (34.9-48.1)	100%	100%	99.10 %

However, based on our definition of sensitivity (specificity) for pools which involves the number of pools which were identified as positive (negative) that was supposed to be positive (negative), we recalculated the values, along with PPV, NPV, and LRs yielding the following results:

Dilution	TP	TN	FP	FN	Sensitivity	Specificity	PPV	NPV	+LR	- LR

1:2	7	1	0	0	100% (59.04-100)	100% (2.5-100)	100% (-)	100%	1002 (0-8.03x10^{23})	<0.01 (0-1.18x10^{23})
1:4	7	1	0	0	100% (59.04-100)	100% (2.5-100)	100% (-)	100%	1002 (0-8.03x10^{23})	<0.01 (0-1.18x10^{23})
1:6	6	1	0	1	85.71% (42.13 - 99.64)	100% (2.5-100)	100% (-)	50% (14.01-85.99)	859 (0-6.89x10^{29})	0.14 (0.02, 0.88)
1:8	5	1	0	2	71.43% (29.04-96.33)	100% (2.5-100)	100% (-)	33.33% (13.42-61.73)	716 (0-5.75x10^{29})	0.29 (0.09, 0.92)
1:10	4	1	0	3	57.14% (18.41-90.10)	100% (2.5-100)	100% (-)	25% (12.41, 43.95)	573 (0-4.6x10^{29})	0.43 (0.18, 1.01)
1:12	4	1	0	3	57.14% (18.41- 90.10)	100% (2.5-100)	100% (-)	25% (12.41, 43.95)	573 (0-4.6x10^{29})	0.43 (0.18, 1.01)
1:16	4	1	0	3	57.14% (18.41- 90.10)	100% (2.5-100)	100% (-)	25% (12.41, 43.95)	573 (0-4.6x10^{29})	0.43 (0.18, 1.01)
1:20	4	1	0	3	57.14% (18.41-90.10)	100% (2.5-100)	100% (-)	25% (12.41, 43.95)	573 (0-4.6x10^{29})	0.43 (0.18, 1.01)
1:24	4	1	0	3	57.14% (18.41- 90.10)	100% (2.5-100)	100% (-)	25% (12.41, 43.95)	573 (0-4.6x10^{29})	0.43 (0.18, 1.01)
1:32	4	1	0	3	57.14% (18.41 -90.10)	100% (2.5-100)	100% (-)	25% (12.41, 43.95)	573 (0-4.6x10^{29})	0.43 (0.18, 1.01)
1:48	3	1	0	4	42.86% (9.90-81.59)	100% (2.5-100)	100% (-)	20% (11.63, 32.20)	429 (0-3.45x10^{29})	0.57 (0.3, 1.09)

What were the likelihood ratios of the various test results?
In all dilutions, positive LRs are all greater than 10 (strongly positive) which indicates the utility of the test for ruling in disease or as a confirmatory test. In pool sizes of 2 and 4, the negative likelihood ratios were found to be less than 0.01 (strongly negative) which indicates the utility of the test for ruling out disease. The negative LRs for pool sizes of 6 and 8 were moderately negative while for pool sizes of 10-48, the negative LRs were found to be weakly negative.

4. APPLICABILITY

No information. Patient characteristics were not elaborated in the study, and therefore, the possible effect of sex, age, co-morbidities, pathology of disease and other socio-economic factors on the accuracy cannot be ascertained.

5. CONCLUSION

We deem that the study has low internal validity due to non-independence of the definition, performance and interpretation of reference and index test. In addition, there was no information on patient characteristics to identify other factors that may affect the accuracy of the test. Furthermore, the sensitivity presented by the study authors were not the same as that of our calculations.

Pooling Upper Respiratory Specimens for Rapid Mass Screening of COVID-19 by Real-Time RT-PCR

Kim, 2020

1. DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of patients (P), examination (E) used and disease or outcome (O) being diagnosed?

Yes. The study provides a direct enough answer to the clinical question of the review.

Research Question	
P	**Positive samples= 50** **Negative samples= 300** *Pooled upper respiratory specimens were prepared from 50 individual SARS-CoV-2–positive specimens and 300 individual SARS-CoV-2–negative specimens. (Page 1, Paragraph 4)*
I	**2, 4, 6, 8, 10, 16 sample pools; PowerCheck 2019-nCoV Real-Time Detection** *We performed rRT-PCR using PowerCheck 2019-nCoV for all pooled specimens. (page 2, paragraph 4)*
C	**STANDARD M nCoV Real-time Detection** or **PowerCheck 2019-nCoV Real-Time Detection** *Laboratory diagnosis of SARS-CoV-2 infection was performed with all specimens using the following rRT-PCR kits targeting the E and RdRp genes: STANDARD M nCoV Real-time Detection or PowerCheck 2019-nCoV Real-Time Detection (Page 1, Paragraph 4)*
O	**Sensitivity and Specificity** *We evaluated the clinical sensitivity and specificity of SARS-CoV-2 rRT-PCR using pooled upper respiratory specimens from confirmed cases. (page 3, paragraph 2)*

2. APPRAISAL OF VALIDITY

2.1. Was the reference standard an acceptable one?

Yes. The reference standards used were STANDARD M nCoV Real-time Detection (SD Biosensor, https://sdbiosensor.com) or PowerCheck 2019-nCoV Real-Time Detection (Kogene Biotech, https://kogene.co.kr) which employs RT-PCR. Since rT-PCR is the currently considered standard for diagnosis of COVID-19, then the reference standard is considered acceptable.

2.2. Was "definition" of the index test and the reference standard independent?

No. The study used STANDARD M nCoV Real-time Detection or PowerCheck 2019-nCoV Real-Time Detection as the reference test and only used PowerCheck 2019-nCoV Real-Time Detection as the index test. SD biosensor detects RdRP as the target gene while Powercheck detects E gene as the target gene. E gene is not specific for SARS-CoV-2 and may be present in other strains of coronaviruses while RdRP gene is specific for SARS-CoV-2. Using Powercheck as the index test may unnecessarily and possibly cause false positive results by detecting other coronaviruses that may or may not be of the SARS-CoV-2 strain. This may result in non-concordance of the index test to the reference test.

2.3. Was "performance" of the index test and the reference standard independent?

Yes. The positive specimens and negative specimens that were used in the pooling strategy were tested individually with RT-PCR.

Laboratory diagnosis of SARS-CoV-2 infection was performed with all specimens using the following rRT-PCR kits targeting the E and RdRp genes: STANDARD M nCoV Real-time Detection or PowerCheck 2019-nCoV Real-Time Detection (Page 1, Paragraph 4)

We performed rRT-PCR using PowerCheck 2019-nCoV for all pooled specimens. (page 2, paragraph 4)

2.4. Was "interpretation of the index test and the reference standard independent?

No. The study employed a retrospective collection of samples where diagnostic status of the samples were already known before it was subjected to the index test. There was no blinding done in the interpretation of the results because positive samples were deliberately included in negative pools to make different dilutions.

We pooled the selected individual SARS-CoV-2–positive specimens with different numbers of SARS-CoV-2–negative specimens to generate 50 sets of pooled specimens in duplicate; the pool sizes of each set were 2, 4, 6, 8, 10, and 16. (page 2, par. 1)

3. APPRAISING THE RESULTS

The study authors reported the following results *(page 2)*:

Table. Test performance of pooled specimens compared with individual specimens for severe acute respiratory syndrome coronavirus 2

No. specimens in pool	Amplification in *E* or *RdRp* gene, %	No amplifications	Sensitivity of pools, % (95% CI)	Cumulative sensitivity, %*
2	100	0	100 (96–100)	100
4	100	0	100 (96–100)	100
6	100	0	100 (96–100)	100
8	97	3	97 (92–99)	97
10	99	1	99 (95–100)	96
16	96	4	96 (90–98)	92

*Calculated sensitivity based on the accumulated discrepancy numbers under the dilution fold

In addition, the study authors also mentioned:
The clinical specificity of pool size 16 was 97% (58/60, 95% CI 87%-99%). (page 3, par. 1)

Based on the given data, we independently calculated the sensitivity, and specificity where we obtained similar results. In addition, we also calculated for PPV, NPV, and LRs when data is available. The independent calculations done can be seen in the table below.

Pool size	TP	TN	FP	FN	Sensitivity (95% CI)	Specificity (95%CI)	PPV (95% CI)	NPV (95% CI)	LR+ (95% CI)	LR- (95% CI)
2	100	-	-	0	100 (96-100)	-	-	-	-	-
4	100	-	-	0	100 (96-100)	-	-	-	-	-
6	100	-	-	0	100 (96-100)	-	-	-	-	-
8	97	-	-	3	97 (92-99)	-	-	-	-	-
10	99	-	-	1	99 (95-100)	-	-	-	-	-
16	96	58	2	4	96 (90-98)	97% (87-99)	97.9% (92.4-99.5)	93.6% (84.7-97.4)	28.8 (7.4-112.5)	0.04 (0.02-0.11)

What were the likelihood ratios of the various test results?
The likelihood ratio was only calculated for the 16 pool size. Based on our calculations, the positive likelihood ratio is 28.8 while the negative likelihood ratio is 0.04. This implies that the test has strongly positive and negative LRs, making them useful for ruling in and ruling out disease.

4. APPLICABILITY

No information. Patient characteristics were not elaborated in the study, and therefore, the possible effect of sex, age, co-morbidities, pathology of disease and other socio-economic factors on the accuracy cannot be ascertained.

5. CONCLUSION

Overall, we deem the study to have moderate internal validity with some concerns on the independence of the criteria and interpretation of the index and reference test. In addition, we obtained similar calculations to that of the study authors and based on our computations, we found that the index test has a strongly positive and negative likelihood ratio for a pool size of 16.

An Evaluation of Pooling Strategies for qRT-PCR Testing for SARS-CoV-2 Infection

Lo et al., 2020

1. DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of patients (P), examination (E) used and disease or outcome (O) being diagnosed?

Yes. The study provides a direct enough answer to the clinical question of the review.

Research Question	
P	Phase 1: nasopharyngeal and oropharyngeal specimens that have been previously collected and tested in RITM, with Ct values ranging between 30 to 38 *(page 13, par. 4)* Phase 2: Volunteer employees from a local supermarket chain *(page 14, par. 4)*
I	Phase 1: Pooled testing with pool sizes of 5, 10, and 20, using 1 positive specimen and a certain amount of diluent/buffer; RT-PCR using Maccura SARS-CoV-2 Fluorescent PCR Kit *(page 14, par. 2-3)* Phase 2: Dorfman pooling (5-1, 10-5-1, 20-10-5-1), using Sansure Novel Coronavirus Nucleic Acid Diagnostic Kit *(page 17, par. 2)*
C	Phase 1: Previously characterized positive by RT-PCR *(page 13-14, par. 5)* Phase 2: Individual testing using Sansure Novel Coronavirus Nucleic Acid Diagnostic Kit *(page 17, par. 2)*
O	Sensitivity, specificity, test savings, turnaround time

2. APPRAISAL OF VALIDITY

2.1. Was the reference standard an acceptable one?

Yes. The reference standard used was RT-PCR which is the currently acceptable standard for confirming presence of SARS-CoV-2 in an individual.

2.2. Was "definition" of the index test and the reference standard independent?

No. For phase 2 of the study, the study mentioned that the same interpretation as per manufacturer's specifications will be followed for both pooled and individual samples. On the other hand, no information was provided in phase 1 as to what the criteria for interpreting results are for the reference test.

For purposes of pooled testing, a negative pool is one that shows no target gene amplification. Any target gene amplification (ORF1 and N genes) regardless of Ct value, degree of amplification or curve properties (sigmoid or non-sigmoid) will be considered positive. Individual samples will undergo the same interpretation as per manufacturer's specifications. (page 17, par. 3)

2.3. Was "performance" of the index test and the reference standard independent?

Yes. All of the tests were subjected to both the index test and reference standard for both phase 1 and phase 2 experiments.

The samples were selected based on the results of their initial real-time RT-PCR runs as well as the quality and remaining volume of the original samples. Undiluted samples were tested along with the diluted samples to ensure that same testing conditions were met for both undiluted and diluted samples. (page 14, par. 1)

The individual and pooled samples underwent nucleic acid extraction and SARS-COV-2 NAAT by qRT-PCR strictly according to the manufacturer's instructions for use and followed strict biosafety guidelines and good clinical laboratory practices. Results were recorded and encoded in electronic data collection forms. (page 18, par.3)

2.4. *Was "interpretation of the index test and the reference standard independent?*
No information. There was no explicit mention of blinding or independent interpretation of the results of both tests for phase 1 and phase 2. However, for phase 1, it is likely that interpretation is not independent since positive samples were deliberately diluted to simulate pooling. For phase 2, it is likely that interpretation was independent given that the experiment was conducted prospectively.

Fifty (50) uL aliquots from each sample were diluted (as described below) to simulate the different pool sizes at the worst possible pooling scenarios - where only 1 specimen is positive out of the pool (page 14, par. 2)

Swabbing was performed according to standard guidelines and procedures. After swabbing, the VTMs were transported back to the COVID-19 Testing Laboratory of PCMC and UPHDMC following biosafety standards and then stored in the reagent refrigerator until testing. (page 16, par. 3)

3. APPRAISING THE RESULTS

What were the likelihood ratios of the various test results?

The following data were presented by the study authors for phase 1 of the experiment:

EFFECT OF POOLING AS SIMULATED BY DILUTION ON TEST SENSITIVITY

	pool size = 5				
Specimen	**N**	**Pos**	**Sensitivity**	**95% CI**	
Fresh	22	17	77%	55%	92%
Mod to strong positive	4	4	100%	40%	100%
Weak positive	18	13	72%	47%	90%
Frozen	14	13	93%	66%	100%
Mod to strong positive	7	7	100%	59%	100%
Weak positive	7	6	86%	42%	100%
Total	**36**	**30**	**83%**	**67%**	**94%**

	pool size = 10				
Specimen	**N**	**Pos**	**Sensitivity**	**95% CI**	
Fresh	22	15	68%	45%	86%
Mod to strong positive	4	4	100%	40%	100%
Weak positive	18	11	61%	36%	83%
Frozen	14	11	79%	49%	95%
Mod to strong positive	7	7	100%	59%	100%
Weak positive	7	4	57%	18%	90%
Total	**36**	**26**	**72%**	**55%**	**86%**

	pool size = 20				
Specimen	**N**	**Pos**	**Sensitivity**	**95% CI**	
Fresh	22	15	68%	45%	86%
Mod to strong positive	4	4	100%	40%	100%
Weak positive	18	11	61%	36%	83%
Frozen	14	9	64%	35%	87%
Mod to strong positive	7	6	86%	42%	100%
Weak positive	7	3	43%	10%	82%
Total	**36**	**24**	**67%**	**49%**	**81%**

(page 23, Table 2)

From this data set, we also conducted an independent calculation of the sensitivity estimates, presented in the table below:

Pool Size	Specimen	Sensitivity	95% CI
5	Fresh (Moderate to strong positive)	100%	39.76-100
	Fresh (weak positive)	72.22%	46.52-90.31
	Fresh (overall)	77.27%	54.63-92.18
	Frozen (moderate to strong positive)	100%	59.04-100
	Frozen (weak positive)	85.71%	42.13-99.64
	Frozen (overall)	92.86%	66.13-99.82
	Overall	83.33%	67.19-93.63
10	Fresh (Moderate to strong positive)	100%	39.76-100
	Fresh (weak positive)	61.11%	35.75-82.70
	Fresh (overall)	68.18%	45.13-86.14
	Frozen (moderate to strong positive)	100%	59.04-100
	Frozen (weak positive)	57.14%	18.41-90.10
	Frozen (overall)	78.57%	49.20-95.34
	Overall	72.22%	54.81-85.80

20	Fresh (Moderate to strong positive)	100%	39.76-100
	Fresh (weak positive)	61.11%	35.75-82.70
	Fresh (overall)	68.18%	45.13-86.14
	Frozen (moderate to strong positive)	85.71%	42.13-99.64
	Frozen (weak positive)	42.86%	9.90-81.59
	Frozen (overall)	64.29%	35.14-87.24
	Overall	66.67%	49.03-81.44

Based on our independent calculations, we were able to obtain the same results as that of the study author's. Moreover, no likelihood ratios can be obtained as no data for specificity was presented for the phase 1 of the study.

For phase 2 of the study, the study authors were able to present the following results:

EFFECT OF POOLING ON TEST PERFORMANCE

Sensitivity

	Dorfman5-1					Dorfman10-5-1					Dorfman20-10-5-1				
	x	n	%	95% CI		x	n	%	95% CI		x	n	%	95% CI	
Site A	6	8	75%	35%	97%	3	8	38%	9%	76%	2	8	25%	3%	65%
Site B	4	4	100%	40%	100%	4	4	100%	40%	100%	4	4	100%	40%	100%
Total	10	12	83%	52%	98%	7	12	58%	28%	85%	6	12	50%	21%	79%

Specificity

	Dorfman5-1					Dorfman10-5-1					Dorfman20-10-5-1				
	x	n	%	95% CI		x	n	%	95% CI		x	n	%	95% CI	
Site A	212	212	100%	98%	100%	212	212	100%	98%	100%	212	212	100%	98%	100%
Site B	216	216	100%	98%	100%	216	216	100%	98%	100%	216	216	100%	98%	100%
Total	428	428	100%	99%	100%	428	428	100%	99%	100%	428	428	100%	99%	100%

Accuracy

	Dorfman5-1					Dorfman10-5-1					Dorfman20-10-5-1				
	x	n	%	95% CI		x	n	%	95% CI		x	n	%	95% CI	
Site A	218	220	99%	97%	100%	215	220	98%	95%	99%	214	220	97%	94%	99%
Site B	220	220	100%	98%	100%	220	220	100%	98%	100%	220	220	100%	98%	100%
Total	438	440	100%	98%	100%	435	440	99%	97%	100%	434	440	99%	97%	99%

Given the following data, we conducted independent calculations of sensitivity and specificity, as well as additional parameters such as PPV, NPV, and LRs:

Dorfman 5-1		**Reference Standard**		**TOTAL**
		(+)	**(-)**	
Index Test	**(+)**	10	0	10
	(-)	2	428	430
TOTAL		12	428	440

	Pt. estimate (95% CI)		**Pt. estimate (95% CI)**
Sensitivity	83.33% (51.59-97.91)	PPV	100% (-)
Specificity	100% (99.14-100)	NPV	99.53% (98.37-99.87)
LR(+)	357.50 (49.66-2673.79)	LR(-)	0.17 (0.05-0.59)

Dorfman 10-5-1		Reference Standard		TOTAL
		(+)	(-)	
Index Test	(+)	7	0	7
	(-)	5	428	433
TOTAL		12	428	440

	Pt. estimate (95% CI)		Pt. estimate (95% CI)
Sensitivity	58.33% (27.67-84.33)	PPV	100% (-)
Specificity	100% (99.14-100)	NPV	98.85% (97.77-99.41)
LR(+)	250.25 (33.35-1877.56)	LR(-)	0.42 (0.21-0.81)

Dorfman 20-10-5-1		Reference Standard		TOTAL
		(+)	(-)	
Index Test	(+)	6	0	6
	(-)	6	428	434
TOTAL		12	428	440

	Pt. estimate (95% CI)		Pt. estimate (95% CI)
Sensitivity	50% (21.09-78.91)	PPV	100% (-)
Specificity	100% (99.14-100)	NPV	98.62% (97.59-99.21)
LR(+)	214.50 (27.95-1646.07)	LR(-)	0.50 (0.28-0.88)

We were able to obtain similar results for the computations of sensitivity and specificity. In addition, based on computations of the likelihood ratios, we found that all three methods of the Dorfman pooling were able to show a strongly positive likelihood ratio indicating their usefulness for ruling in disease. On the other hand, the negative likelihood ratio for Dorfman 5-1 was better (moderately negative) as compared to the other two Dorfman pooling method which were found to be weakly negative.

4. APPLICABILITY

Are there biologic issues that may affect accuracy of the test? (Consider the influence of sex, co-morbidity, race, age, and pathology). Are there socio-economic issues that may affect accuracy of the test?

In terms of the patient samples used for phase 1 of the study, Ct values were found to have affected the sensitivity of pooled testing, with strongly moderate to strong positive specimens being easily detected by pooled testing. Furthermore, in phase 2 of the study, it was mentioned that only asymptomatic individuals were enrolled and that those who had symptoms, less than 18 years of age, and pregnant women were excluded. Aside from this, no other elaboration on patient characteristics such as age, sex, onset of disease, severity of disease, comorbidities and the like were presented for individual samples used in both phases of the experiment.

5. CONCLUSION

Overall, the study can be said to have a moderate internal validity given that an acceptable reference standard was used and that performance of the index and reference test was independent. Issues

encountered in the validity of the study were the independence of the definition of the index and reference test as well as lack of information to ascertain the independence of interpretation of the results. As for the results of the study, it was found that Dorfman pooling in 5-1, 10-5-1, 20-10-5-1 stages had strongly positive likelihood ratios but weak to moderately negative likelihood ratios. Lastly, it was difficult to ascertain factors that may have affected the test due to lack of information on certain patient characteristics.

Evaluation and Comparison of the Hologic Aptima SARS-CoV-2 and the CDC 2019 nCoV Real-Time RT-PCR Diagnostic Panel using a Four-Sample Pooling Approach

Mitchell et al, 2020

1. DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of patients (P), examination (E) used and disease or outcome (O) being diagnosed?

Research Question	
P	Frozen residual nasopharyngeal swabs collected in viral transport media from patients presenting to Pittsburgh-based UPMC medical facilities
I	CDC 2019 nCoV Real-Time RT-PCR Diagnostic Panel (CDC) four-sample pooling Hologic Aptima SARS-CoV-2 TMA Assay (TMA) four-sample pooling 35 positive pools (consisting of 1 positive and 3 negatives) and 20 negative pools (consisting of four uniquely negative samples) were tested.
C	Cepheid SARS-CoV-2 EUA (Cepheid)
O	Specificity, Sensitivity, #TP, #FN, #FP, #TN

2. APPRAISAL OF VALIDITY

2.1. Was the reference standard an acceptable one?

Yes. Reference standard utilized individual testing of RT-PCR using the same RT-PCR test kits. Moreover, prior testing was done using another brand of RT-PCR test kits (Cepheid) to determine the performance of CDC and TMA test kits in individual testing.

2.2. Was "definition" of the index test and the reference standard independent?

No. The definition of the index test and the reference standard are not independent given that pooling of samples was conducted after identifying each specimen to be positive or negative during individual testing.

"Samples were originally tested using the Cepheid SARS-CoV-2 EUA, which served as the reference method, and averaged Ct of N1 and E targets were used to classify samples into group 1 (Ct < 34, n=23), group 2 (Ct = 34-36, n=4), and group 3 (Ct ≥ 37, n=8). For pools, samples were thawed, mixed and 500 µl each of four samples were pooled. 35 positive pools, each consisting of 1 positive and 3 negatives, and 20 negative pools (four uniquely negative samples) were made." (page 2-3)

2.3. Was "performance" of the index test and the reference standard independent?

No. Performance of the index test and the reference standard was not independent since pools that tested positive led to individual re-testing.

"Positive samples were re-tested individually by both platforms; negatives were not." (page 3)

2.4. Was "interpretation" of the index test and the reference standard independent?

No information. We cannot determine if the interpretation of the index test and the reference standard was independent since no statement was given as to whether interpretation of the index test was blinded with respect to the interpretation of the reference standard.

3. APPRAISING THE RESULTS

What were the likelihood ratios of the various test results?

For CDC:

CDC Pooled Test Result	Reference Standard (Individual RT-PCR Test)		
	Positive	**Negative**	**Row Total**
Positive	34	0	34
Negative	1	20	21
Total	35	20	55

Statistic	Value	95% CI
Sensitivity	97.14%	85.08% to 99.93%
Specificity	100.00%	85.18% to 100.00%
Positive Likelihood Ratio	23.31	3.42 to 158.96
Negative Likelihood Ratio	0.03	0.00 to 0.20
Positive Predictive Value	100.00%	-
Negative Predictive Value	95.24%	74.34% to 99.28%

The use of the CDC test kit in pooled testing is seen to have a positive likelihood ratio of 23.31, implying that the test is strongly positive, thus the use of the CDC test kit in pooled testing may be useful in ruling in the disease. Meanwhile, it also has a negative likelihood ratio of 0.03, implying that the test is strongly negative, thus the use of the CDC test kit in pooled testing may also be useful in ruling out the disease.

For TMA:

TMA Pooled Test Result	Reference Standard (Individual RT-PCR Test)		
	Positive	**Negative**	**Row Total**
Positive	32	0	32
Negative	3	20	23
Total	35	20	55

Statistic	Value	95% CI
Sensitivity	91.43%	76.94% to 98.20%
Specificity	100.00%	85.16% to 100.00%
Positive Likelihood Ratio	19.20	2.83 to 130.37

Negative Likelihood Ratio	0.09	0.03 to 0.25
Positive Predictive Value	100.00%	-
Negative Predictive Value	86.96%	69.32% to 95.16%

The use of the TMA test kit in pooled testing is seen to have a positive likelihood ratio of 21.94, implying that the test is strongly positive and therefore it could be useful as a confirmatory test. Meanwhile, it also has a negative likelihood ratio of 0.09, implying that the test is strongly negative, and therefore it may also be useful as a screening test.

Sensitivity of CDC and TMA for Positive Subgroups

	CDC	**95% CI**	**TMA**	**95% CI**
Group 1 (Ct < 34)	100.00 %	85.18% to 100.00%	100.00%	85.18% to 100.00%
Group 2 (Ct = 34 to 36)	100.00 %	39.76% to 100.00%	100.00%	39.76% to 100.00%
Group 3 (Ct ≥ 37)	87.5%	47.35% to 99.68%	62.5%	24.49% to 91.48%

Given the computed sensitivity of the subgroups of the positive pools, it can be seen that both the CDC and TMA test kits have high sensitivity for positive pools with a Ct value less than 37, while sensitivity declines at a Ct value of 37 or higher.

4. APPLICABILITY

Are there biologic issues that may affect accuracy of the test? (Consider the influence of sex, co-morbidity, race, age, and pathology). Are there socio-economic issues that may affect accuracy of the test?

No biologic issues that may affect the accuracy of the test were explicitly mentioned in the study. However, the study only aimed "to use pooled testing as an approach for asymptomatic individuals requiring SARS-CoV-2 screening" (page 4). Other factors enumerated in the study that may affect accuracy of the test include "tracking individual specimens in pools, retrieving individual samples for confirmatory testing, automated and streamlined protocols, workflow, reporting, and billing." (page 4-5)

5. CONCLUSION

Overall, the study is considered to have low internal validity because while the reference standard is acceptable, the definition and performance of the index test and the reference standard are not independent of each other, which could lead to a higher risk of bias for the results of the study. Additionally, based on the computed likelihood ratios for both CDC and TMA, the use of both test kits in pooled testing may be useful as confirmatory tests or screening tests. Moreover, both test kits are shown to be more sensitive at Ct values less than 37.

Saliva sample pooling for the detection of SARS-CoV-2
Pasomsub et al, 2020

1. DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of patients (P), examination (E) used and disease or outcome (O) being diagnosed?

Research Question	
P	Saliva samples of patients under investigation for COVID-19 during the outbreak in Bangkok, Thailand
I	SARS-CoV-2 Nucleic Acid Diagnostic Kit (Sansure) five-sample pooling SARS-CoV-2 Nucleic Acid Diagnostic Kit (Sansure) ten-sample pooling Ten µL of the extracted RNA from each patient was pooled consecutively into pools of five samples and pools of ten samples.
C	SARS-CoV-2 Nucleic Acid Diagnostic Kit individual testing
O	Ct values of ORFlab gene, Ct values of N gene

2. APPRAISAL OF VALIDITY

2.1. Was the reference standard an acceptable one?

Yes. The reference standard used in the study is RT-PCR wherein samples were individually tested.

2.2. Was "definition" of the index test and the reference standard independent?

No. The index test and reference standard use the same criteria, targeting both the ORFlab and N gene fragments similarly.

2.3. Was "performance" of the index test and the reference standard independent?

Yes. The performance of the index test and the reference standard were independent as the samples were tested individually to investigate the effects of storage conditions. Additionally, the results of the individual testing did not affect the pooling of samples as samples were pooled in such a way that it could not be identified as to how many positive samples were placed in the pool.

"Ten µL of the extracted RNA from each patient was pooled consecutively into pools of five samples and pools of ten samples." (page 4)

"All Ct values of ORFlab and N genes in this study were compared with the Ct of immediate RT-PCR testing for SARS-CoV-2 performed in the previous study." (page 6)

2.4. Was "interpretation" of the index test and the reference standard independent?

No information. Given that the study is a retrospective study, we cannot ascertain that interpretation of the index test and reference standard are independent since there was no mention of blinding or independent interpretation of the two tests.

3. APPRAISING THE RESULTS

What were the likelihood ratios of the various test results?

a. For five-sample pools

Pooled Test Result	Reference Standard (Individual RT-PCR Test)		
	Positive	Negative	Row Total
Positive for 2 genes	11	0	11
Positive for 1 gene	2	0	2
Negative	0	27	27
Total	13	27	40

Statistic	Value	95% CI
Sensitivity	100.00%	75.29% to 100.00%
Specificity	100.00%	87.23% to 100.00%
Positive Likelihood Ratio	28.00	4.09 to 191.88
Negative Likelihood Ratio	0.07	0.01 to 0.47
Disease Prevalence	9.00%	
Positive Predictive Value	100.00%	-
Negative Predictive Value	100.00%	-
Accuracy	100.00%	91.19% to 100.00%

Based on the computed likelihood ratios, the use of five-sample pooling has a strongly positive likelihood ratio, implying that it may be useful as a confirmatory test. Meanwhile, its negative likelihood ratio is strongly negative, implying that it may also be useful as a screening test.

b. For ten-sample pools

Pooled Test Result	Reference Standard (Individual RT-PCR Test)		
	Positive	Negative	Row Total
Positive for 2 genes	12	0	12
Positive for 1 gene	1	0	1
Negative	0	7	7
Total	13	7	20

Statistic	Value	95% CI
Sensitivity	100.00%	75.29% to 100.00%
Specificity	100.00%	59.04% to 100.00%
Positive Likelihood Ratio	8.00	1.28 to 50.04
Negative Likelihood Ratio	0.07	0.01 to 0.47
Disease Prevalence	9.00%	-
Positive Predictive Value	100.00%	-
Negative Predictive Value	100.00%	-
Accuracy	100.00%	83.16% to 100.00%

The computed positive likelihood ratio of ten-sample pooling is moderately positive, while the negative likelihood ratio is strongly negative. Based on the computed likelihood ratios, the use of ten-sample pooling may be a useful screening test.

4. APPLICABILITY

Are there biologic issues that may affect accuracy of the test? (Consider the influence of sex, co-morbidity, race, age, and pathology). Are there socio-economic issues that may affect accuracy of the test?

The study mentions that "pool size should be selected according to the disease prevalence to save the test, and hence the cost, for each negative pool" (page 7). Moreover, the study also looks into the effect of the "storage condition, storage time, and freeze-thaw on the accuracy of the detection of the virus" (page 8). It is important to note that in resource-limited settings, facilities and testing capacity might not be available in those areas and thus, misclassification of patients, once the accuracy of the tests are affected as certain storage conditions may not be met, may be of concern. Hence it is recommended to conduct "immediate RT-PCR testing […] to minimize the effect of storage conditions that can decrease the sensitivity of the testing" (page 9). Another limitation of the pooling strategy includes the "inability to evaluate the adequacy of each specimen in a pool due to the loss of ability to detect the housekeeping gene from each sample" (page 9).

5. CONCLUSION

The study can be considered to have moderate internal validity given that it uses an appropriate reference standard and performance of the index test and reference standard are independent of each other. However, it must be noted that definition of the index test and reference standard utilize the same criteria, presenting some risk of bias. Moreover, there is no information to ascertain the independence of the interpretation of the index test and reference standard.

Based on the computed likelihood ratios, it is observed that the use of five-sample pools may be useful as either a confirmatory test or screening test. However, for ten-sample pools, it may be more useful as a screening test. It is also important to note that the study mentions other factors, such as storage conditions and storage time, that may affect the accuracy of the test.

Pooling of SARS-CoV-2 samples to increase molecular testing throughput
Perchetti et al, 2020

1. DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of patients (P), examination (E) used and disease or outcome (O) being diagnosed?

Yes.

Research Question	
P	SARS-CoV-2 positive and negative specimens (n=384) *"Initial water and VTM templates were used to confirm pipetting accuracy with 384 samples into a 96-well deep well plate."* (page 2, par, 2)
I	Pooled testing using CDC-based Washington State EUA SARS-CoV-2 RT-PCR assay targeting N1 and N2 genes 32 four-way pools tested; 1 pool = 1 positive + 3 negative samples *"We programmed a HAMILTON Microlab STARlet Automated Liquid Handler (Atlantic Lab Equipmant, Beverly, MA) to perform 4-way pooling on our CDC-based Washington state EUA SARS-CoV-2 RT-PCR assay targeting N1 and N2 as previously described."* (page 1, par. 4)
C	Individual testing using CDC-based Washington State EUA SARS-CoV-2 RT-PCR assay targeting N1 and N2 genes "These extracts include a. pdf print-out highlighting positive and inconclusive samples that need to be individually tested along with their rack location, a. json file with raw CT values that is loaded into our data warehouse, and a. csv file with the negative samples that can be imported into our Sunquest laboratory information system." (page 2, par. 4)
O	Sensitivity, Specificity, Difference in Ct values

2. APPRAISAL OF VALIDITY

2.1. Was the reference standard an acceptable one?

Yes. The reference standard used was RT-PCR which is the accepted standard test in detecting COVID-19 among patients.

"*Prior to pooling, neat samples were assayed by LDT and stored at 4 °C for < 24 h. HeLa cells were included as a negative extraction control and water as a negative PCR template on every run.*" (page 2, par. 2)

2.2. Was "definition" of the index test and the reference standard independent?

Likely no. Although not mentioned, there has been an overlap in the criteria for the use of the positive samples for the index test and the reference standard.

"We expanded this experiment to include 32 additional SARS-CoV-2 positive specimens by the CDC-based Washington state EUA assay pooled into negative samples."
page 2, par. 7

2.3. Was "performance" of the index test and the reference standard independent?

Yes. Both the individual and pooled testing using the CDC-based Washington state EUA SARS-CoV-2 RT-PCR assay were done on the specimens included.

"*Individual samples were pooled 1:4 through automated liquid handling, extracted, and assayed by our emergency use authorized CDC-based RT-PCR laboratory developed test." (page 1) "Prior to pooling, neat samples were assayed by LDT…"*
(page 2, par. 2)

2.4. Was "interpretation of the index test and the reference standard independent?

No. There was no mention whether any blinding was done during interpretation of results. In addition, the status of the specimen was previously known to the researchers before pooling as they designed the study to have a unique sample pooled with three distinct negative samples.

"*32 additional SARS-CoV-2 positive specimens by the CDC-based Washington state EUA assay pooled into negative samples. Each pool contains 4 specimens: 1 unique positive sample pooled into 3 distinct negative samples.*" (page 2, par. 7)

3. APPRAISING THE RESULTS

What were the likelihood ratios of the various test results?

The reported results are as follows:
Sensitivity (94%) and specificity (100%) values were given hence we can reconstruct the table and independently compute for the PPV, NPV and likelihood ratios:

Test for sensitivity (reconstructed 2x2 table)

Pooled RT-PCR	Individual RT-PCR		
	(+)	(-)	Total
(+)	30	0	30
(-)	2	32	34
Total	32	32	64

Computed values:

Statistic	Value	95% CI
Sensitivity	93.75%	79.19% to 99.23%
Specificity	100%	96.23% to 100.00%
Positive Predictive Value	100%	
Negative Predictive Value	94.12%	80.70% to 98.38%
LR+	30.94	4.48 to 213.60
LR-	0.06	0.02 to 0.24

Accuracy (*)	95.38%	87.10% to 99.04%

The LR+ is strongly positive, indicating that the test is useful for ruling in the disease and is a good confirmatory test. The LR- is strongly negative, indicating that the test is also useful for ruling out the disease and is a good screening test.

Based on our independent calculations, the computed sensitivity, specificity, PPV, and NPV matched the reported values in the study.

4. APPLICABILITY

Are there biologic issues that may affect accuracy of the test? (Consider the influence of sex, co-morbidity, race, age, and pathology). Are there socio-economic issues that may affect accuracy of the test?

No information. There is no reported biological nor socio-economic information about the specimen hence the effect on the accuracy of the test cannot be determined.

5. CONCLUSION

The study is considered to have moderate internal validity due to non-independence of definition and interpretation of results of the index test and reference standard (e.g. no mention of blinding). The results presented in the study were also consistent with our independent computation. Results showed strong likelihood ratios, indicating the index test may be useful as both confirmatory and screening tests.

Pooled testing for COVID-19 diagnosis by real-time RT-PCR: A multi-site comparative evaluation of 5- & 10-sample pooling

Praharaj et al., 2020

1. DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of patients (P), examination (E) used and disease or outcome (O) being diagnosed?

Yes. The study provides a direct enough answer to the clinical question.

Research Question	
P	Nasopharyngeal swab samples in the viral transport medium (VTM) previously tested positive and negative for SARS-CoV-2 *(page 90, par 2. column 1)*
I	5 sample pools (containing one positive and four negatives) and 10 (containing one positive and nine negatives) sample pools; Pooled testing using single-step real-time RT-PCR for SARS-CoV-2 targeting the E gene (E gene screening assay-NIV protocol, TIB Molbiol 2019 nCoV Kit (TIB Molbiol, Germany), Standard M nCoV Real-Time Detection Kit, (SD Biosensor Inc., Republic of Korea), PathoDetect COVID-19 Detection Kit (Mylab Discovery Solutions, Maharashtra)

	(page 90, par. 2, column 2)
C	Individual sample testing using single-step real-time RT-PCR for SARS-CoV-2 targeting the E gene; (E gene screening assay-NIV protocol, TIB Molbiol 2019 nCoV Kit (TIB Molbiol, Germany), Standard M nCoV Real-Time Detection Kit, (SD Biosensor Inc., Republic of Korea), PathoDetect COVID-19 Detection Kit (Mylab Discovery Solutions, Maharashtra) *(page 90, par. 2, column 2)*
O	Concordance between individual sample testing and pooled sample testing *(page 90, par. 2, column 2)*

2. APPRAISAL OF VALIDITY

2.1. Was the reference standard an acceptable one?

Yes. Individual RT-PCR testing is currently the acceptable standard for diagnosis of COVID-19.

2.2. Was "definition" of the index test and the reference standard independent?

No. All tests used for the reference standard and index standard utilize the E gene for detection of SARS-CoV-2, hence there is an overlap with the criteria for positive and negative results of the test.

Single-step real-time RT-PCR for SARS-CoV-2 targeting the E gene was performed on the extracted RNA from individual samples as well as the sample pools. (page 90, par 2, column 2)

2.3. Was "performance" of the index test and the reference standard independent?

Yes. Both the index test and the reference standard were conducted in all of the samples used in the study.

"RNA was extracted from both 5-sample and 10-sample pools, as well as the individual samples using the same RNA extraction kit. A volume of 200 µl of the pooled sample was used for RNA extraction. Participating laboratories were instructed to use the same extraction kits for individual samples, 5-sample pools, as well as 10-sample pools. The individual samples, as well as the pools, were included in the same extraction batch, and the same aliquot of sample was used for individual sample testing as well as creating 5- and 10-sample pools for RNA extraction. Each laboratory also included two negative pools of samples, one with five negative samples and another with 10 negative samples." (page 90, par. 3, column 1)

Single-step real-time RT-PCR for SARS-CoV-2 targeting the E gene was performed on the extracted RNA from individual samples as well as the sample pools. (page 90, par 2, column 2)

2.4. Was "interpretation of the index test and the reference standard independent?

No information. There was no statement on the study that indicates that blinding or independent interpretation of results for the index and reference test was done.

3. APPRAISING THE RESULTS

What were the likelihood ratios of the various test results?

The following data presented by the authors (page 91) show the percent concordance of results of individual and pooled testing for both 5-sample and 10-sample pool.

Table I. Concordance between pooled and individual sample testing

C_t value range for individual positive sample	Number of included positive samples	Concordance between pooled testing and individual testing	
		5-sample pooled testing versus individual testing, n (%)	10-sample pooled testing versus individual testing, n (%)
≤30 cycles	23	23/23 (100)	22/23 (95.6)
>30 and ≤33 cycles	44	42/44 (95.5)	35/44 (79.5)
>33 and <36 cycles	33	23/33 (69.7)	9/33 (27.3)
Overall	100	88/100 (88)	66/100 (66)

Using these data as true positives and false negatives, we computed the sensitivity of pooled testing and present the data with their respective 95% C.I.s below:

Sensitivity of 5-sample pooling:

Ct Value of Included Samples in the Pool	# TP	# FN	Sensitivity Pt. Estimate	95% C.I.
Overall	88	12	88%	79.98-93.64
≤30 cycles	23	0	100%	85.18-100
>30 and ≤33 cycles	42	2	95.45%	84.53-99.44
> 33 and <36 cycles	23	10	69.70%	51.29-84.41

Sensitivity of 10-sample pooling:

Ct Value of Included Samples in the Pool	# TP	# FN	Sensitivity Pt. Estimate	95% C.I.
Overall	66	34	66%	55.85-75.18
≤30 cycles	22	1	95.65%	78.05-99.89
>30 and ≤33 cycles	35	9	79.55%	64.70-90.20
> 33 and <36 cycles	9	24	27.27%	13.30-45.52

Based on our independent calculations, we have obtained the same point estimates as those of the reported values by the study authors. However, given that only sensitivity can be obtained, no likelihood ratios can be computed for this study.

4. APPLICABILITY

Are there biologic issues that may affect accuracy of the test? (Consider the influence of sex, co-morbidity, race, age, and pathology). Are there socio-economic issues that may affect accuracy of the test?

No information. As anonymized samples were used in the testing by the study authors, no information on age and gender corresponding to the samples were used during analysis. Furthermore, no patient characteristics of the included sample were discussed in the study.

5. CONCLUSION

Overall, the study can be considered to have moderate internal validity having satisfied the directness, acceptability of the reference standard, and independence of performance of the index test and

reference test. However, there was still limited information given on certain domains such as the independence of interpretation and lack of independence in the definition of the two tests which could have led to bias in the results of the study. In addition to this, we were not able to determine the likelihood ratio of the pooled testing strategy and its applicability due to limited information.

Proposal of RT-PCR Based Mass Population Screening for Severe Acute Respiratory Syndrome Coronavirus 2 (Coronavirus Disease 2019)
Sahajpal, 2020

1. DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of patients (P), examination (E) used and disease or outcome (O) being diagnosed?

Yes. The study provides a direct enough answer to the clinical question of the review.

Research Question	
P	**Positive samples = 6** **Negative samples= 934** *In mass screening, a 940 nasopharyngeal swab samples previously tested for SARS-CoV-2 were de-identified and assigned random numbers. The 940 samples contained 934 negative and 6 positive samples. From this, 94 pools of 10 samples each were generated and given a unique number. (page 3, par, 1)*
I	**10 sample pools; RT-PCR assay (PerkinElmer Inc.,)** *The assay is based on RNA extraction, followed by TaqMan based RT-PCR assay, to conduct in vitro transcription of SARS-CoV-2 RNA, DNA amplification, and fluorescence detection (PerkinElmer Inc., Waltham,MA). The assay targets specific genomic regions of SARS-CoV-2: nucleocapsid (N) gene and ORF1ab. (page 2, paragraph 5)*
C	**Comparator cannot be confirmed if RT-PCR.** *In mass population screening (Figure 1), under the institutional review board approved protocol, 940 nasopharyngeal swab samples previously tested for SARS-CoV-2 were de-identified and assigned random numbers (performed by N.S.S.). The 940 samples contained 934 negative and 6 positive samples. From this, 94 pools of 10 samples each were generated and given a unique number (performed S.A.,.7, who was blinded to the initial sample preparation). (page 3, paragraph 1)*
O	**Positive percent agreement, negative percent agreement**

2. APPRAISAL OF VALIDITY

2.1. Was the reference standard an acceptable one?

No information. The reference standard cannot be confirmed, except that samples were already identified as positive and negative.

2.2. Was "definition" of the index test and the reference standard independent?

No information. There was no given information on the criteria for interpreting results for both the index and reference standard.

2.3. Was "performance" of the index test and the reference standard independent?

No. The reference test (individual test) was only performed in the samples in the pools that turned out to be positive, otherwise, negative pool results were only confirmed by removing blinding and referring to the original individual results.

2.4. Was "interpretation of the index test and the reference standard independent?

Yes. The study employed blinding by assigning random numbers to the retroactively obtained nasopharyngeal swabs with known status. As such, pooled testing was done without the experimenters knowing the diagnostic status of the patient. After the experiment, and once the positive samples were identified, results were verified against the results of the individual samples performed beforehand.

3. APPRAISING THE RESULTS

The study authors reported the following results:

Therefore, 1000 samples resulted in an overall 91.6% positive percentage agreement and 100% negative percentage agreement compared with individual testing approach. (page 4, par. 1)

Based on the report of the study authors, we were able to reconstruct the 2x2 table and we were able to calculate the same results as reported. In addition, we calculated for the likelihood ratios of the test, presented below:

Test Result	**Reference Standard**		
	Disease present	Disease absent	Row total
Positive	11	0	11
Negative	1	988	989
Column Total	12	988	1000

Based on our calculations, the accuracy data are as follows:

Statistic	**Value**	**95% CI**
Sensitivity	91.67%	61.52-99.79
Specificity	100.00%	99.63- 100
Positive Likelihood Ratio	$9.1x10^5$	$0-7.5x10^{32}$
Negative Likelihood Ratio	0.08	0.01 to 0.54
Positive Predictive Value	100.00%	-
Negative Predictive Value	99.90%	99.34% to 99.98%

What were the likelihood ratios of the various test results?

The likelihood ratios computed were 0.08 for –LR (strongly negative), and $9.1x10^{32}$ for +LR (strongly positive). This shows that the test is good for both ruling in and ruling out disease.

4. APPLICABILITY

No information. Patient characteristics were not elaborated in the study, and therefore, the possible effect of sex, age, co-morbidities, pathology of disease and other socio-economic factors on the accuracy cannot be ascertained.

5. CONCLUSION

The study is deemed to have low internal validity due to the fact that the reference standard cannot be ascertained and there are concerns on the independence of the definition and performance of the index and reference test. In addition, there was no information on the patient characteristics used, hence applicability to certain population groups cannot be ascertained.

Novel multiple swab method enables high efficiency in SARS-CoV-2 screenings without loss of sensitivity for screening of a complete population
Schmidt et al., 2020

1. DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of patients (P), examination (E) used and disease or outcome (O) being diagnosed?

Yes. The study provides a direct enough answer to the clinical question.

Research Question	
P	Samples with predetermined concentrations of SARS-CoV-2 *(page 4, par. 1)*; Patients with clinical symptoms *(page 4, par. 2)*; and Asymptomatic residents of a nursing home *(page 5, par. 2)*
I	5 sample pool *(Four of the samples were incubated in a solution containing SARS-CoV-2, and one was incubated in solution not containing virus, page 4, par. 1)* 5 sample pool *(50 samples routinely sent in for SARS-CoV-2 testing of patients with clinical symptoms were randomly assigned to 10 five-sample mini-pools, page 4, par. 2)* 10 sample pool *(100 samples from asymptomatic residents of a nursing home were randomly assigned to 10 multiple-swab tubes containing 10 swabs each, page 5, par. 2)* Multiple swab method to NAT; Roche cobas SARS-CoV-2 *(page 2, Section 2.1)*
C	Proof of concept setup: predetermined status of SARS-CoV-2, but reference test not mentioned (page 4, par. 1) Patients with clinical testing: Individual NAT testing (page 4, par. 2); Roche cobas SARS-CoV-2 Asymptomatic residents of nursing home: no explicit mention of the reference test (page 5, par. 2); Roche cobas SARS-CoV-2
O	Number of positive pools that turn out positive in pooled testing

2. APPRAISAL OF VALIDITY

2.1. Was the reference standard an acceptable one?

For the evaluation of pooled testing in patients with clinical testing, the reference standard was acceptable given that the currently recognized gold standard for diagnosis of COVID-19 is individual NAT/RT-PCR testing. However, for proof of concept and asymptomatic residents of a nursing home, the reference standard was not explicitly mentioned.

"Next, we evaluated a five-sample minipool in a proof-of-concept setup. Samples from a proficiency panel test provider (INSTAND) with predetermined concentrations of SARS-CoV-2 were used." (page 4, par. 1)

"To evaluate the test in patients with a moderate likeli-hood of SARS-CoV-2 infection, 50 samples routinely sentin for SARS-CoV-2 testing of patients with clinical symp-toms were

randomly assigned to 10 five-sample mini-pools. Both the reference tube and the multiple-swabtube underwent NAT testing." (page 4, par. 2)

In a second real-life application, 100 samples from asymptomatic residents of a nursing home were ran-domly assigned to 10 multiple-swab tubes containing 10 swabs each. All 5 multiple-swab tubes containing a total of 8 positive swabs were correctly identified. All 5 multiple-swab tubes containing no positive swab sample were also true negative" (page 5, par. 2)

2.2. *Was "definition" of the index test and the reference standard independent?*

No information. There is no information provided as to the criteria (i.e. target gene, Ct value) for interpreting results of both the index test and reference standard.

2.3. *Was "performance" of the index test and the reference standard independent?*

Yesfor the evaluation of the test in patients with clinical symptoms. Both the reference tube which contains the individual sample and the multiple-swab tube underwent NAT testing.

"To evaluate the test in patients with a moderate likeli-hood of SARS-CoV-2 infection, 50 samples routinely sentin for SARS-CoV-2 testing of patients with clinical symp-toms were randomly assigned to 10 five-sample mini-pools. Both the reference tube and the multiple-swabtube underwent NAT testing." (page 4, par. 2)

No for the proof of concept study. Tt was mentioned that samples with predetermined concentrations of SARS-CoV-2 were used and the study did not mention whether these samples still underwent individual RT-PCR/NAT testing.

"Next, we evaluated a five-sample minipool in a proof-of-concept setup. Samples from a proficiency panel test provider (INSTAND) with predetermined concentrations of SARS-CoV-2 were used….Results of single-swab sample tubes and multiple-swab sample tubes were compared. " (page 4, par. 1)

No information for the evaluation of samples from asymptomatic residents of a nursing home. There was no mention of the test to confirm whether the negative pools contained negative samples. The study only mentioned that multiple-swab tubes containing no positive swab samples were also true negatives. For the positive samples, it can be inferred that individual and pooled testing were conducted in all samples given that Ct values were reported for both tests.

"All 5 multiple-swab tubes containing a total of 8 positive swabs were correctly identified. All 5 multiple-swab tubes containing no positive swab sample were also true negative." (page 5, par. 2)

2.4. *Was "interpretation of the index test and the reference standard independent?*

No information. The study did not mention whether blinding or independent interpretation of the index and reference test was done.

3. APPRAISING THE RESULTS

What were the likelihood ratios of the various test results?

The following results were presented by the study authors for the proof-of-concept setup:

"Four of the samples were incubated in a solution containing SARS-CoV-2, and one was incubated in a solution not containing virus. Each of the swabs was transferred to a five-sample multiple-swab tube, in accordance with the protocol described above. Respiratory swabs from SARS-CoV-2–negative volunteers were used to complete the pools. Results of single-swab sample tubes and multiple-swab sample tubes were com-pared. We determined that all multiple-swab tubes con-taining a SARS-CoV-2–positive sample were correctly identified in the multiple-swab protocol, independent of the virus concentration in the original sample. All multiple-swab tubes containing no SARS-CoV-2–positive sample were also true negative."

No computation of sensitivity was provided, however, based on the information presented, we calculated sensitivity, specificity, PPV, NPV, and likelihood ratios. Imputation was conducted using 0.001 as a correction factor.

Proof-of-concept setup (5-samples/pool)		**Reference Standard**		
Index test (Multiple swab NAT testing)		**Positive**	**Negative**	**Total**
	Positive	4	0	4
	Negative	0	1	1
	Total	4	1	5

Statistic	**Point estimate**	**95% C.I.**
Sensitivity	100%	39.76-100
Specificity	100%	2.50-100
PPV	100%	-
NPV	100%	-
LR (+)	1001.75	$0\text{-}8.02\text{x}10^{29}$
LR (-)	<0.01	$0\text{-}2.05\text{x}10^{23}$

The positive likelihood ratio was computed to be at 1001.75 and the negative likelihood ratio computed was less than 0.01 which indicates that the test is both strongly positive and negative indicating its utility for ruling in and ruling out disease.

The following results were presented by the study authors for the evaluation among symptomatic patients :

"To evaluate the test in patients with a moderate likelihood of SARS-CoV-2 infection, 50 samples routinely sent in for SARS-CoV-2 testing of patients with clinical symp-toms were randomly assigned to 10 five-sample mini-pools. Both the reference tube and the multiple-swabtube underwent NAT testing.

Each of the four pools con-taining a positive sample was correctly identified with the multiple-swab method. Multiple-swab tubes con-taining no positive sample were also correctly identified to be negative in multiple-swab tubes of five swabs."

Like with the previous scenario, no computation of accuracy measures was provided, however, based on the information presented, we calculated sensitivity, specificity, PPV, NPV and LRs. Imputation was conducted using 0.001 as a correction factor.

Symptomatic patients (5-samples/pool)		**Reference Standard**		
Index test (Multiple swab NAT testing)		**Positive**	**Negative**	**Total**
	Positive	4	0	4
	Negative	0	6	6
	Total	4	6	10

Statistic	**Point estimate**	**95% C.I.**
Sensitivity	100%	39.76-100
Specificity	100%	54.07-100
PPV	100%	-
NPV	100%	-
LR (+)	6000.5	$0\text{-}4.94\text{x}10^{30}$
LR (-)	<0.01	$0\text{-}2.05\text{x}10^{23}$

Based on the given data, the sensitivity, specificity, PPV, and NPV of 5-sample pooled testing among symptomatic patients is 100%. For sensitivity and specificity, this means that all pools with positive or negative samples were correctly identified by the multiple swab method as positive or negative respectively. Likewise, the 100% PPV and NPV means that pools who were flagged as positive by the multiple swab method are truly positive or containing a positive sample within the pool.

The positive likelihood ratio was computed to be at 1001.75 and the negative likelihood ratio computed was less than 0.01 which indicates that the test is both strongly positive and negative indicating its utility for ruling in and ruling out disease.

The following results were presented by the study authors for the evaluation among asymptomatic residents in a nursing home:

"In a second real-life application, 100 samples from asymptomatic residents of a nursing home were ran-domly assigned to 10 multiple-swab tubes containing 10 swabs each. All 5 multiple-swab tubes containing a total of 8 positive swabs were correctly identified. All 5 multiple-swab tubes containing no positive swab sample were also true negative."

Like with the two previous scenarios, no computation of accuracy measures was provided, however, based on the information presented, we calculated sensitivity, specificity, PPV, NPV and LRs.

Asymptomatic residents in a nursing home (10-samples/pool)		**Reference Standard**		
Index test (Multiple swab NAT testing)		**Positive**	**Negative**	**Total**
	Positive	5	0	5
	Negative	0	5	5
	Total	5	5	10

Statistic	**Point estimate**	**95% C.I.**
Sensitivity	100%	47.82-100
Specificity	100%	47.82-100
PPV	100%	-
NPV	100%	-
LR (+)	5001	$0\text{-}4.11\text{x}10^{30}$
LR (-)	<0.01	$0\text{-}1.64\text{x}10^{23}$

Based on the given data, the sensitivity, specificity, PPV, and NPV of 10-sample pooled testing among symptomatic patients is 100%. For sensitivity and specificity, this means that all pools with positive or negative samples were correctly identified by the multiple swab method as positive or negative respectively. Likewise, the 100% PPV and NPV means that pools who were flagged as positive by the multiple swab method are truly positive or containing a positive sample within the pool.

The positive likelihood ratio was computed to be at 5001 and the negative likelihood ratio computed was less than 0.01 which indicates that the test is both strongly positive and negative indicating its utility for ruling in and ruling out disease.

4. APPLICABILITY

Are there biologic issues that may affect accuracy of the test? (Consider the influence of sex, co-morbidity, race, age, and pathology). Are there socio-economic issues that may affect accuracy of the test?

No information. The study only mentioned that the pooled testing approach was done among symptomatic patients as well as asymptomatic residents in a nursing home. No other characteristics were given such as severity of disease, time of testing (e.g., time since onset of symptoms or contact), and presence of other comorbidities. However, it can be inferred that for the asymptomatic residents in a nursing home, the samples used were that of older age individuals due to the nature of the setting to which the samples were obtained. No subanalysis comparing these characteristics were provided hence, it is difficult to determine whether these factors possibly affect performance of pooled testing.

5. CONCLUSION

Overall, the study can be considered to have moderate internal validity given that the study provides a direct enough answer to the research question and several circumstances in the study have satisfied the criteria of acceptability of reference standard and performance of the index and reference test. However, other biases may have been present due to the lack of information to ascertain independence of definition and interpretation of the index test and reference standard in general. Furthermore, while the results show that sensitivity, specificity, PPV, and NPV are all 100% with moderately positive and strongly negative likelihood ratios, results were imprecise and have wide confidence intervals due to the very small sample pools used in the study.

Efficient high-throughput SARS-CoV-2 testing to detect asymptomatic carriers

Shental et al, 2020

1. DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of patients (P), examination (E) used and disease or outcome (O) being diagnosed?

Yes.

Research Question	
P	Four sets of samples (n=384) previously clinically tested for COVID-19 "*We tested P-BEST using four sets of 384 samples, each containing an increasing number of positive carriers ranging from two to five.*". (page 2 , par. 2)
I	Pooling-Based Efficient SARS-CoV-2 Testing (P-BEST) using clinically approved COVID-19 PCR-based diagnostic protocol Four sets of 48 pools tested, each set containing an increasing number of positive samples ranging from 2 -5 positive samples. (page 2,par. 1)
C	Individual RT-PCR Testing using Seegene COVID-19 diagnostic kit, which identifies three SARS-CoV-2 genes: E, RdRP, and N genes.

	(page 6, par. 2)
o	Number of true positives, false positives, true negatives, false negatives

2. APPRAISAL OF VALIDITY

2.1. Was the reference standard an acceptable one?

Yes. The reference standard used was individual RT-PCR. Currently, RT-PCR is the accepted standard test in detecting COVID-19 among patients.

"Naso- and oropharynx swabs were collected for analysis by the laboratory of clinical virology in SUMC, which is approved by the Israeli Ministry of Health to test for SARS-CoV-2 infections. The laboratory uses a clinically approved 2019-nCoV detection kit (Seegene, CA, USA) for both viral nucleic acid extraction and quantitative reverse transcription PCR (qRT-PCR)–based amplification. The kit identifies three SARS-CoV-2 genes: E, RdRP, and N genes. RNA extraction was performed using the STARMag 2019-nCoV kit (Seegene, CA, USA) on a liquid-dispensing robot (STARlet Hamilton, USA). All samples were analyzed individually, and positive and negative results were recorded by the SUMC diagnostic laboratory before our study."
(page 6, par. 2)

2.2. Was "definition" of the index test and the reference standard independent?

No. There is an overlap in the definition of a positive/detected result for the index test and the reference standard. A Ct value of <40 for both tests were considered positive.
"PCR samples with C(t) values of <40 were considered positive." (page 6, par. 5)

Table 8. Result interpretation, clinical specimens

Ct value	Result
≤ 40	Detected (+)
> 40 or N/A	Not detected (-)

(page 63, Seegene RT-PCR IFU)

2.3. Was "performance" of the index test and the reference standard independent?

Yes. The samples prior to the study were tested individually using 2019-nCoV Seegene kit, then pooled and tested using a clinically approved COVID-19 PCR-based diagnostic protocol.

"Naso- and oropharynx swabs were collected for analysis by the laboratory of clinical virology in SUMC, which is approved by the Israeli Ministry of Health to test for SARS-CoV-2 infections. The laboratory uses a clinically approved 2019-nCoV detection kit (Seegene, CA, USA) for both viral nucleic acid extraction and quantitative reverse transcription PCR (qRT-PCR)–based amplification. The kit identifies three SARS-CoV-2 genes: E, RdRP, and N genes. RNA extraction was performed using the STARMag 2019-nCoV kit (Seegene, CA, USA) on a liquid-dispensing robot (STARlet Hamilton, USA). All samples were analyzed individually, and positive and negative results were recorded by the SUMC diagnostic laboratory before our study." (page 6, column 2, paragraph 2)

Pooled samples were then tested by the clinical diagnostic laboratory of the Soroka University Medical Center (SUMC) using a clinically approved COVID-19 PCR-based diagnostic protocol that included an RNA extraction stage (page 2, column 1, paragraph 2)

2.4. Was "interpretation of the index test and the reference standard independent?

No information. There was no mention of whether there was any blinding done during interpretation of results. Also, given that the status of each sample was known to the authors prior pooling, the interpretation may not be independent.

"We tested P-BEST using four sets of 384 samples, each containing an increasing number of positive carriers ranging from two to five." (page 2, paragraph 2)

3. APPRAISING THE RESULTS

What were the likelihood ratios of the various test results?

There were no computed values presented (Sensitivity, specificity, PPV, NPV, LR+, LR-) however, the number of true positives, false positives, true negatives, false negatives were given. From these numbers we compute the Sensitivity, specificity, PPV, NPV, LR+, LR- values for each set of pools tested.

Number of TP, FP, TN, FN for each experiment, testing 8 pools of 48 samples each, with varying number of positive samples:

	2+/384	3+/384	4+/384	5+/384
# TP	2	3	4	5
#FP	0	0	0	1
#TN	382	381	380	378
#FN	0	0	0	0

Computed values:

	2+/384		3+/384		4+/384		5+/384	
Statistic	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI
Sensitivity	100%	15.81% to 100.00%	100%	29.24% to 100.00%	100%	39.76% to 100.00%	100%	47.82% to 100.00%
Specificity	100%	99.04% to 100.00%	100%	99.04% to 100.00%	100%	99.03% to 100.00%	99.74%	98.54% to 99.99%
PPV	100%		100%		100%		79.29%	35.09% to 96.44%
NPV	100%		100%		100%		100%	
LR+	383.00	54.09 to 2712.09	382.00	53.95 to 2704.99	381.00	53.81 to 2697.89	379.00	53.52 to 2683.69
LR-	0.00		0.00		0.00		0.00	
Accuracy	100%	99.04% to 100.00%	100%	99.04% to 100.00%	100%	99.04% to 100.00%	99.74%	98.56% to 99.99%

The LR+ for all sets of tests are considered strongly positive, ranging from 379-383, indicating that the test is useful for ruling in the disease and a good confirmatory test. Based on the reconstructed 2x2 table, the LR- is 0.00 which is strongly negative indicating that the test is also useful for ruling out the disease.

4. APPLICABILITY

Are there biologic issues that may affect accuracy of the test? (Consider the influence of sex, co-morbidity, race, age, and pathology). Are there socio-economic issues that may affect accuracy of the test?

No information. There is no reported biological nor socio-economic information about the specimen hence the effect on the accuracy of the test cannot be determined.

5. CONCLUSION

The study is considered to have **moderate internal validity** provided that there is an overlap in the criteria for the index test and reference standard. There is also insufficient information to determine the independence of interpretation of the index test and reference standard. There may be bias as the methods are designed such that the status of the samples was previously known to the researchers before the conduct of the test. Independent computation based on the given data shows that the positive likelihood ratios are strongly positive and the negative likelihood ratios are weak-moderately negative for all methods done in the study.

"Evaluation of pooled sample analysis strategy in expediting case detection in areas with emerging outbreaks of COVID-19: A pilot study"

Singh et al, 2020

1. DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of patients (P), examination (E) used and disease or outcome (O) being diagnosed?

Yes. The study provides a direct enough answer to the clinical question.

Research Question	
P	Suspected COVID-19 patients (page 3, para. 1)
I	Pooled testing (Pool of 5) *"For pooled analysis, 200 µl from each of 5 consecutive samples were collected in a single 1.5 ml centrifuge tube and processed for RNA extraction"*(page 3, para. 1) RNA extraction (QIAamp Viral RNA Mini kit (Qiagen, Hilden, Germany) Detection (Real-Time Fluorescent RT-PCR Kit, BGI, Hong Kong) (page 3, para. 1)
C	Individual testing RNA extraction (QIAamp Viral RNA Mini kit (Qiagen, Hilden, Germany) Detection (Real-Time Fluorescent RT-PCR Kit, BGI, Hong Kong) (page 3, para. 1)
O	Sensitivity, Specificity, PPV, and NPV (page 4-5, par. 2-5)

2. APPRAISAL OF VALIDITY

2.1. Was the reference standard an acceptable one?

Yes. The RT-PCR individualized testing was used as a reference test.

> *"For pooled analysis, 200µl from each of 5 consecutive samples were collected in a single 1.5 ml centrifuge tube and processed for RNA extraction using QIAamp Viral RNA Mini kit (Qiagen, Hilden, Germany) as per manufacturer's instructions. RNA extraction for individualized testing was also performed using the same kit. The extracted RNA samples were subjected to diagnosis using Real-Time Fluorescent RT-PCR Kit for Detecting SARS-CoV-2 (BGI, Hong Kong) as per the manufacturer's protocol on a BioRad CFX96 thermal cycler." (page 3, par. 1)*

2.2. Was "definition" of the index test and the reference standard independent?

No. There was a corresponding criterion used for both the index test and the reference standard.

> *"As recommended by the manufacturer, a sigmoidal curve with a Ct value ≤ 35 was considered as the criterion for considering a sample as positive for SARS-CoV-2." (page 3, par. 1)*

2.3. Was "performance" of the index test and the reference standard independent?

Yes. Parallel testing on the individual samples and their pools was done.

> *"A total of 545 samples were collected, with 140, 270 and 135 of them belonging to districts A1, B1 and B2, respectively. Both, the individual samples and their pools were processed in parallel for testing." (page 3-4, par. 1)*

2.4. Was "interpretation of the index test and the reference standard independent?

Yes. Comparison of the diagnostic performance of the qRT-PCR individualized and pooled sample was done in a blinded manner.

> *"The diagnostic performance of qRT-PCR on pooled sample was compared with that of individual samples in a blinded manner." (page 1, par. 1)*

3. APPRAISING THE RESULTS

What were the likelihood ratios of the various test results?

2x2 table of Reference Standard results from the study:

Pooled Test Result	**Reference Standard (qRT-PCR)**		
	Positive	**Negative**	**Row Total**
Positive	26	0	26
Negative	0	519	519

Total	26	519	545

Diagnostic characteristics of the pooled sample analysis strategy from the study:

Prevalence independent parameters		
Statistic	**Value**	**95% CI**
Sensitivity	75.0%	47.6 to 92.7%
Specificity	98.9%	94.2% to 100%
Positive Likelihood Ratio	69.8	9.7 to 500.1
Negative Likelihood Ratio	0.3	0.1 to 0.6

Prevalence dependent parameters			
Assumed Prevalence	**Statistic**	**Value**	**95% CI**
1%	PPV	41.3%	9% to 83.5%
	NPV	99.8%	99.4% to 99.9%
2%	PPV	58.7%	16.6% to 91.1%
	NPV	99.5%	98.8% to 99.9%
3%	PPV	68.3%	23.1% to 93.9%
	NPV	99.2%	98.2% to 99.7%
4%	PPV	74.40%	28.8% to 95.4%
	NPV	99%	97.6% to 99.6%
5%	PPV	78.6%	33.7% to 96.3%
	NPV	98.7%	97% to 99.4%

The study presented the following information on the diagnostic accuracy:

TP	**FP**	**FN**	**TN**	**Sensitivity (95% CI)**	**Specificity (95% CI)**
12	1	4	92	0.75 (0.48,0.93)	0.99 (0.94, 1.00)

Independent calculation on the diagnostic characteristics of the pooled sample analysis strategy

Statistic	Value	95% CI
Sensitivity	75.00%	47.62% to 92.73%
Specificity	98.92%	94.15% to 99.97%
Positive Likelihood Ratio	69.75	9.73 to 500.08
Negative Likelihood Ratio	0.25	0.11 to 0.59
Disease Prevalence	14.68%	8.63% to 22.74%
Positive Predictive Value	92.31%	62.60% to 98.85%
Negative Predictive Value	95.83%	90.78% to 98.17%
Accuracy	95.41%	89.62% to 98.49%

Independent calculation on the diagnostic characteristics of the pooled sample analysis strategy with assumed disease prevalence of 1%

Statistic	Value	95% CI
Sensitivity	75.00%	47.62% to 92.73%
Specificity	98.92%	94.15% to 99.97%
Positive Likelihood Ratio	69.75	9.73 to 500.08
Negative Likelihood Ratio	0.25	0.11 to 0.59
Disease Prevalence	1.00%	
Positive Predictive Value	41.33%	8.95% to 83.47%
Negative Predictive Value	99.75%	99.41% to 99.89%
Accuracy	98.69%	94.34% to 99.91%

Independent calculation on the diagnostic characteristics of the pooled sample analysis strategy with assumed disease prevalence of 2%

Statistic	Value	95% CI
Sensitivity	75.00%	47.62% to 92.73%
Specificity	98.92%	94.15% to 99.97%
Positive Likelihood Ratio	69.75	9.73 to 500.08
Negative Likelihood Ratio	0.25	0.11 to 0.59
Disease Prevalence	2.00%	
Positive Predictive Value	58.74%	16.57% to 91.08%
Negative Predictive Value	99.49%	98.81% to 99.78%
Accuracy	98.45%	93.96% to 99.86%

Independent calculation on the diagnostic characteristics of the pooled sample analysis strategy with assumed disease prevalence of 3%

Statistic	Value	95% CI
Sensitivity	75.00%	47.62% to 92.73%
Specificity	98.92%	94.15% to 99.97%
Positive Likelihood Ratio	69.75	9.73 to 500.08
Negative Likelihood Ratio	0.25	0.11 to 0.59
Disease Prevalence	3.00%	
Positive Predictive Value	68.33%	23.13% to 93.93%
Negative Predictive Value	99.22%	98.21% to 99.67%
Accuracy	98.21%	93.59% to 99.79%

Independent calculation on the diagnostic characteristics of the pooled sample analysis strategy with assumed disease prevalence of 4%

Statistic	Value	95% CI
Sensitivity	75.00%	47.62% to 92.73%
Specificity	98.92%	94.15% to 99.97%
Positive Likelihood Ratio	69.75	9.73 to 500.08
Negative Likelihood Ratio	0.25	0.11 to 0.59
Disease Prevalence	4.00%	
Positive Predictive Value	74.40%	28.84% to 95.42%
Negative Predictive Value	98.96%	97.60% to 99.55%
Accuracy	97.97%	93.23% to 99.71%

Independent calculation on the diagnostic characteristics of the pooled sample analysis strategy with assumed disease prevalence of 5%

Statistic	Value	95% CI
Sensitivity	75.00%	47.62% to 92.73%
Specificity	98.92%	94.15% to 99.97%
Positive Likelihood Ratio	69.75	9.73 to 500.08
Negative Likelihood Ratio	0.25	0.11 to 0.59
Disease Prevalence	5.00%	
Positive Predictive Value	78.59%	33.86% to 96.34%
Negative Predictive Value	98.69%	96.98% to 99.43%
Accuracy	97.73%	92.87% to 99.63%

The computed point estimates for sensitivity and specificity are the same in the study. The presented from the study values are as follows: sensitivity (75%, 47.6 to 92.7%), specificity (98.9%, 94.2% to 100%), positive LR (69.8, 9.7 to 500.1), and negative LR (0.3, 0.1 to 0.6). Also, the computed NPV, PPV in accordance with the prevalence dependent measures are consistent with the results of the study.

The independent calculation of the pooled samples shows that the accuracy is inversely proportional to the prevalence of the disease such that as the disease prevalence gets higher, the accuracy gets lower. In addition, PPV decreases as the degree of disease prevalence increases. With the 69.75 (LR > 10, strongly positive) positive likelihood ratio, the study espouses an idea that pooled testing is useful and reliable for ruling in disease.

4. APPLICABILITY

Are there biologic issues that may affect accuracy of the test? (Consider the influence of sex, co-morbidity, race, age, and pathology). Are there socio-economic issues that may affect accuracy of the test?

No information. Patients characteristics were not discussed in the study (the study conforms to the ethics criterion of anonymized biological samples), ergo, applicability cannot be determined.

> *"As per the guidelines, the institutional ethics committee can grant waiver of consent if the research is done on anonymized biological samples and/or the primary purpose of the research is refinement and improvement of the public health programs. As our study met both these criteria it was approved with the waiver of consent" (page 3, par. 1)*

> *"The relevant clinical and epidemiological details of the patients were entered in a standard form approved by the Indian Council of Medical Research, which is spearheading the nationwide laboratory network for COVID-19 testing" (page 3, par. 1)*

5. CONCLUSION

While there are issues on the applicability and the independence of the definition of the index and reference test, the study has a moderate internal validity since it provides a direct and an unbiased comparison of diagnostic performance between the individualized and pooled testing. In addition, the parameters that were reported in the study reflect the independent calculation.

"Evaluating the efficiency of specimen pooling for PCR-based detection of COVID-19"
Wacharapluesadee et al, 2020

1. DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of patients (P), examination (E) used and disease or outcome (O) being diagnosed?

Yes. The study provides a direct enough answer to the clinical question.

Research Question	
P	Nasopharyngeal and throat swab from COVID-19 patients under investigation (page 2194, par. 4)
I	Two (2) Pooling ratios using real-time PCR (qPCR): 1x ratio represents 10% infection rate (0.1 ml f NT-VTM from one SARS-CoV-2 positive sample + 0.9 ml of pooled negative NT-VTM sample) (page 2194, para. 6) 2x ratio represents 20% infection rate (0.2 ml f NT-VTM from one SARS-CoV-2 positive sample + 0.8 ml of pooled negative NT-VTM sample) (page 2194, para. 6)
C	Individual testing for positive samples- Real-time PCR (qPCR) Negative samples- qPCR amplifying the ORF1ab gene (BGI, Shenzhen, China) (page 2194, para. 6)
O	Sensitivity was determined using qPCR threshold cycles from SARS-CoV-2 testing (page 2197, para. 1)

2. APPRAISAL OF VALIDITY

2.1. Was the reference standard an acceptable one?

Yes. The reference standard was an acceptable one since individual testing using RT-PCR is currently the "gold standard" for diagnosing COVID-19 infection

" Real-time PCR (qPCR) for detection of SARS-CoV-2 was performed using a commercial kit that targets the ORF1ab gene" (page 2194, par. 7)

2.2. Was "definition" of the index test and the reference standard independent?

No. The definition of the index test and the reference standard are not independent.
The cutoff PCR cycle threshold (Ct) was set at 38 (page 2194). Hence, the index test and the reference standard used this value to determine viral load in the pooling strategies.

In addition, the previously positive SARS-COV-2 specimens with their corresponding Ct values were used as criteria in the pooling strategies.

> *"The protocol's stated limit of detection of ORF1ab real-time PCR was 100 copies/mL and the cutoff PCR cycle threshold (Ct) was 38." (page 2194, par. 6)*
>
> *"Previously positive specimens with high and low-concentrations of RNA, as determined by PCR Ct values at the time of detection, were selected to determine the effect of viral load on pooling to ensure that the sensitivity and accuracy of the assay were maintained." (page 2194, par. 7)*
>
> *"In this study, specimens with Ct values between 26 and 35 were considered to have low concentrations of viral RNA, while those with Ct values lower than 26 were considered to have of high-concentrations viral RNA. Ct values higher than 35 were considered weakly positive." (page 2195, par. 1)*
>
> *"The sensitivity of viral RNA detection for each pool was compared with the sensitivity of qPCR results for the individually tested positive specimen in that pool. For 2X ratio pools, the positive specimen with the lower Ct value (Positive NT 1), when individually tested, was used for comparison." (page 2195, par. 4)*
>
> *"All 1X ratio pools (Ct<35) were positive, with Ct value difference within a range of −1.36 to +1.66 when compared to individual (non-pooled) testing. All 2X ratio pools were positive, with Ct value difference within a range of −1.72 to +1.81 when compared to individual testing." (page 2197, par. 1)*

2.3. Was "performance" of the index test and the reference standard independent?

Yes. The 49 positive samples were tested individually using the "using the standard real-time quantitative PCR (qPCR) to ensure that detection accuracy is not compromised" (page 2194) while the 50 leftover negative samples (tested using qPCR amplifying the ORF1ab gene) were combined into a single sample and was retested to confirm the negative result using qPCR.

To compare the limit of detection of specimen pooling and individual testing, replication was performed:

> "*The fifteen 1X(L>35) pools were tested by performing duplicate (replicates I and II) qPCR assays to determine the limit of detection of specimen pooling when compared to individual testing." (page 2195, par. 3)*
>
> *"In pooled testing of 1X L>35, 13 of 15 of either replicate pools tested positive for SARS-CoV-2. Of the 13 positive pools, 4 pools had only 1 replicate that tested positive. The two false-negative pooled samples tested positive in only 1 replicate when individually tested." (page 2197, par. 2)*

2.4. Was "interpretation" of the index test and the reference standard independent?

No. It is likely not to be independent since there is already knowledge of the Ct values (results of the gold standard—individual qPCR test) before incorporating them into the different pooling strategies.

> *"NT specimens with PCR cycle threshold (Ct) greater than 35 were pooled to determine the limit of detection and sensitivity of pooling samples to test for SARS-CoV-2." (page 2194, par. 3)*
>
> *"Forty-nine PCR positive NT specimens yielding Ct ranging from 12.91 to 37.10 were selected for the study in five pooling ratios. Thirty-one of these had a 1X pooling ratio and 18 had 2X ratios." (page 2195, par. 2)*
>
> *"Among the 1X ratio, 12 had low viral concentrations, (L, Ct values from 27.90 to 34.86), 15 had weakly positive viral concentrations (L>35, Ct values from 35.23 to 37.10), and 4 had high viral concentrations (H, Ct values from 18.00 to 23.76)." (page 2195, par. 2)*
>
> *"The 2X ratio pools had two positive specimens each (Positive NT 1 and 2 in Table 1), with viral concentrations as follows: five pools had two low concentration specimens (L+L, Ct values from 29.82 to 35.52), five pools had two high concentration specimens (H+H, Ct values from 12.91 to 25.56), and eight pools had one high and one low concentration specimens (H+L, Ct values from 18.47 to 33.41)." (page 2195, par. 3)*

3. APPRAISING THE RESULTS

What were the likelihood ratios of the various test results?

The study presented the overall following information on the diagnostic accuracy of pooled testing:

TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
47	-	2	-	95.92%(0.86 to 0.1)	-

1X ratio pools results modelling 10% infection rate:

TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
29	-	2	-	93.55% (0.79 to 0.99)	-

2X ratio pools results modelling 20% infection rate

TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
18	-	-	-	100.00% (0.81 to 1)	-

In the study the negative homogenized pool of 50 was used as a dilution mixture for the NT positive specimens. The result of the calculation is consistent with the authors' conclusion that pooling "does not compromise the sensitivity of detecting SARS-CoV-2 provided the Ct value of the individually tested sample is lower than 35" (page 2198).

Independent computation shows that there is a high sensitivity on both pooling ratios (1x and 2x), However, given that only sensitivity can be obtained, no likelihood ratios can be computed for this study. Furthermore, it is important to note that 1x ratio pooling of specimens with Ct values higher

than 35 (weakly positive viral concentrations) yields two (2) false-negative results and low sensitivity.

4. APPLICABILITY

Are there biologic issues that may affect accuracy of the test? (Consider the influence of sex, co-morbidity, race, age, and pathology). Are there socio-economic issues that may affect accuracy of the test?

No information. There were no patient characteristics mentioned since the study was an "evaluation of laboratory techniques using archived clinical specimens" (page 2194). It was only mentioned that the NT specimens used in the study were collected from the patients under investigation for COVID-19 infection.

5. CONCLUSION

Based on the appraisal, the study had some issues in terms of the definition and interpretation of the index test and the reference standard. It can be inferred that the study has a moderate internal validity. Furthermore, the lack of information on the characteristics of the sample specimens, limits the appraisal on the applicability of the test. Further, based on the results of the study, the pooling strategies have a bearing on the calculated sensitivity.

EVALUATION OF COVID-19 RT-QPCR TEST IN MULTI SAMPLE POOLS

Yelin et al, 2020

1. DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of patients (P), examination (E) used and disease or outcome (O) being diagnosed?

Research Question	
P	Swabs from both nostrils and throat previously collected by healthcare providers and sent to the virology laboratory at the Rambam Health Care Campus, Haifa, Israel
I	AgPath-ID One-Step RT-PCR Reagents (Thermo Fisher Scientific) in a Bio-Rad CFX 96 qPCR machine with WHO primers and probe for analysis of pooled samples prior to RT-qPCR - Five positive samples and 67 negative samples were chosen arbitrarily. Sixty-six of the negative samples were mixed into pools of different sizes containing equal volumes of 1, 2, 4, 8, 16, and 32 unique samples. Negative pools of size 1 and 2 were prepared in duplicates made of different samples to determine whether different negative-sample composition in the pool affected the detection of positive samples. The final 67th sample was mixed with the pool of negative samples as control for the positive samples. Seegene Allplex 2019-nCov Assay in a Bio-Rad CFX 96 qPCR machine for analysis of pooled samples prior to RNA extraction - Transport swab buffers were taken from the collection tubes of 3 previously confirmed positive samples and mixed at equal volumes with the sample transport

	buffer from the collection tubes of 7 previously determined negative samples. A volume of 500 µL from the pooled tube was mixed with 2 mL lysis buffer for inactivation, and RNA was extracted [...] and eluted in 50 µL elution buffer.
C	Individual testing of samples using AgPath-ID One-Step RT-PCR Reagents (Thermo Fisher Scientific) in a Bio-Rad CFX 96 qPCR machine with WHO primers and probe
O	Sensitivity

2. APPRAISAL OF VALIDITY

2.1. Was the reference standard an acceptable one?

Yes. Individual testing using the same RT-PCR test kits was used as a reference standard.

2.2. Was "definition" of the index test and the reference standard independent?

No. The use of the same test kits for the index test and reference standard lead to an overlap of the criteria used in interpretation.

"Laboratory RT-qPCR procedure [for pooled samples] was performed according to the procedure for individual samples in the clinical laboratory, on an identical qPCR machine and program and with reagents used." (page 2)

2.3. Was "performance" of the index test and the reference standard independent?

No information. It was not mentioned in the study as to whether both the index test and reference standard were performed on all samples. It was only mentioned that only previously tested positive samples were tested individually, but there was no information regarding the individual testing of negative samples.

"First, by pooling RNA extracts of clinical samples, we tested previously confirmed positive samples alone and combined with an increasing number of previously confirmed negative samples." (page 1)

2.4. Was "interpretation" of the index test and the reference standard independent?

No. Interpretation of the index test and the reference standard are not independent since prior results from the individual testing of the sample were used as a basis in the pooling of samples for the index test.

"The negative pools were distributed in 6 rows of a 96-well plate, 5 µL per well, and 10 µL of the positive samples, and the 67th negative sample were distributed in the 7th row. Also, 5 µL of the positive samples were then diluted into the "pool" of 1 negative sample to make a ½ dilution, then the ½ dilution was diluted in the 2 samples pool to make a ¼ dilution, and so forth, up to 1/64." (page 2)

3. APPRAISING THE RESULTS

What were the likelihood ratios of the various test results?

Pooled Test Result	**Reference Standard (Individual RT-PCR Test)**		
	Positive	**Negative**	**Row Total**

Positive	9	0	9
Negative	1	2	3
Total	10	2	12

Statistic	Value	95% CI
Sensitivity	90.00%	55.50% to 99.75%
Specificity	100.00%	15.81% to 100.00%
Positive Likelihood Ratio	2.70	0.54 to 13.56
Negative Likelihood Ratio	0.10	0.02 to 0.64
Positive Predictive Value	100.00%	
Negative Predictive Value	66.67%	23.75% to 92.77%
Accuracy	91.67%	61.52% to 99.79%

For pooled testing using RT-qPCR, the positive likelihood ratio is considered to be weakly positive, implying that this may not be a good test for ruling in the disease. Meanwhile, the negative likelihood ratio is considered to be moderately negative, hence this could be of some use as a screening test.

4. APPLICABILITY

Are there biologic issues that may affect accuracy of the test? (Consider the influence of sex, co-morbidity, race, age, and pathology). Are there socio-economic issues that may affect accuracy of the test?

No biologic issues were raised in the study that may affect accuracy of the test and no socio-economic issues were also listed in the study that may affect the accuracy of the test. Given this, applicability of the study could not be determined.

5. CONCLUSION

Upon appraisal, the study may be considered to have low internal validity. The study utilizes an acceptable reference standard. However, definition of the index test and reference standard are not independent given that the same test kits were used, leading to an overlap in the criteria used for interpretation. Moreover, interpretation of the index test and reference standard were also not independent since prior results from the reference standard were used as a basis for pooling in the index test.

Based on the computed likelihood ratios, the index test has a weakly positive likelihood ratio, implying that this may not be a good test for ruling in the disease, while it also has a moderately negative likelihood ratio, implying that it may be of some use as a screening test.

Development and Evaluation of Novel and Highly Sensitive Single-Tube Nested Real-Time RT-PCR Assays for SARS-CoV-2 Detection

Yip et al., 2020

1. DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of patients (P), examination (E) used and disease or outcome (O) being diagnosed?

Yes. The study provides a direct enough answer to the clinical question.

Research Question	
P	SARS-CoV-2 positive specimens with low viral load and 49 SARS-CoV-2 negative specimens *(page 7, par. 1)*
I	4 pooled samples (1 positive sample + 49 negative samples); STN real-time RT-PCR assays (STN COVID-19-RdRp/Hel Assay and STN COVID-19-N Assay) *(page 5, Table 2; page 7, par. 1)*
C	non-nested COVID-19-RdRp/Hel assay; (QuantiNova Probe RT-PCR Kit (QIAGEN, Hilden, Germany) *(page 5, Table 2; page 7, par. 1)*
O	Number of true positive pools and false negative pools *"When the Cp value of the positive sample was approximately 32, both the non-nested COVID-19-RdRp/Hel assay and the STN RT-PCR assays flagged the pool positive. However, when the positive sample had a Cp value of approximately 33, the two STN RT-PCR assays, but not the non-nested COVID-19-RdRp/Hel assay, detected SARS-CoV-2 RNA in the pool. For the pools containing a positive sample with Cp > 34, the non-nested and STN RT-PCR assays flagged the pools as negative."* *(page 5, par. 1)*

2. APPRAISAL OF VALIDITY

2.1. Was the reference standard an acceptable one?

Yes. The reference standard was an acceptable one. Currently, RT-PCR is the recognized gold standard for detection of COVID-19 among patients.

"These pooled samples were evaluated by the non-nested COVID-19-RdRp/Hel assay and the STN real-time RT-PCR assays." (page 7, par. 1)

2.2. Was "definition" of the index test and the reference standard independent?

No information. There is no available information as to the criteria for interpreting the index and the reference tests. However, given that one of the PCR kits for pooled testing is the same as that of the individual testing, it is likely that there is overlap in the criteria for both tests.

2.3. Was "performance" of the index test and the reference standard independent?

No information. While the positive samples were both tested using the index tests and the reference standard, there was no information regarding whether the negative samples combined with the positive sample were tested with the reference standard.

"Furthermore, we prepared four pooled samples, with each pool consisting of one SARS-CoV-2-positive specimen with low viral load and 49 SARS-CoV-2-negative specimens. These pooled

samples were evaluated by the non-nested COVID-19-RdRp/Hel assay and the STN real-time RT-PCR assays." (page 7, par. 1)

2.4. *Was "interpretation of the index test and the reference standard independent?*
No information. There was no information whether interpretation of the index test and reference standard was independent or blinded. Given that there are also no criteria mentioned to determine a positive or negative test result, it is possible that interpretation was not independent given that study authors had knowledge of the viral load of the positive sample prior to pooling.

"Furthermore, we prepared four pooled samples, with each pool consisting of one SARS-CoV-2-positive specimen with low viral load and 49 SARS-CoV-2-negative specimens." (page 7, par. 1)

3. APPRAISING THE RESULTS

What were the likelihood ratios of the various test results?

The study authors reported the following information regarding the accuracy of the index tests:

"While the non-nested COVID-19-RdRp/Hel assay was positive in only one of four sample pools (25%), both of the STN assays were positive in two of four samples pools (50%)." (page 1, abstract)

Based on the information provided, we considered these values as the true positives and false negatives then independently calculated sensitivities and 95% C.I. for each index test presented. Note however that there is a very small number of pools tested, hence, it is expected that 95% C.I.s are wide. In addition, given that only data for TP and FN were available, no likelihood ratios can be calculated for this study.

Assay	# TP	# FN	Sensitivity Pt. Estimate	95% C.I.
Non-nested COVID-19-RdRp/Hel assay	1	3	25%	0.63-80.59
STN COVID-19-RdRp/Hel assay	2	2	50%	6.76-93.24
STN COVID-19-N assay	2	2	50%	6.76-93.24

4. APPLICABILITY

Are there biologic issues that may affect accuracy of the test? (Consider the influence of sex, co-morbidity, race, age, and pathology). Are there socio-economic issues that may affect accuracy of the test?
No information. Patient characteristics were not elaborated in the study, and therefore, the possible effect of sex, age, co-morbidities, pathology of disease and other socio-economic factors on the accuracy cannot be ascertained.

5. CONCLUSION

In general, the study can be considered to have low internal validity given that there is limited information to ascertain whether a criterion or domain has been satisfied and that the possibility of bias arising from the study cannot be eliminated.