SENSITIVITY OF GARGLE SAMPLES COMPARED TO SWABS FOR SARS-COV-2 DETECTION WITH NUCLEIC ACID AMPLIFICATION TESTING: A SYSTEMATIC REVIEW AND META-ANALYSIS

by

IRYNA KAYDA

B.Sc., The University of British Columbia, 2020

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Experimental Medicine)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

October 2023

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Sensitivity of gargle samples compared to swabs for SARS-CoV-2 detection with nucleic acid amplification testing: a systematic review and meta-analysis

submitted by Iryna Kayda in partial fulfilment of the requirements for

the degree of Master of Science

in Experimental Medicine

Examining Committee:

Dr. David Goldfarb, Clinical Associate Professor, Department of Pathology and Laboratory Medicine, The University of British Columbia

Supervisor

Dr. Pascal Lavoie, Professor, Department of Pediatrics, The University of British Columbia

Supervisory Committee Member

Dr. Audi Setiadi, Clinical Assistant Professor, Department of Pathology and Laboratory Medicine, The University of British Columbia

Additional Examiner

Abstract

Background: Challenges arose with diagnostic testing during the coronavirus disease 2019 (COVID-19) pandemic. Gargle sampling emerged as a novel method for COVID-19 testing, yet there remains uncertainty about its performance when compared with more conventional sampling methods.

Objective: We investigated the performance of self-collected gargle samples compared to traditional healthcare worker (HCW)-collected upper respiratory tract swabs for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) detection with nucleic acid amplification testing (NAAT). We aimed to (1) estimate sensitivity of gargle sampling, (2) estimate the difference in sensitivity between gargle and swab methods, and (3) understand how various testing contexts may impact gargle sensitivity.

Methods: MEDLINE, EMBASE, Web of Science, Global Index Medicus, and preprint servers were searched. Studies reporting primary data and investigating COVID-19 diagnostic performance of self-collected gargle samples compared to a HCW-collected swab, with at least five matched samples tested using NAAT, were included. Random effects meta-analysis was conducted to estimate both the pooled gargle sensitivity and mean difference in sensitivity between gargle and swab methods. Subgroup analyses were also performed to assess the potential impact of different sampling variables.

Results: Searches identified 327 results with 24 studies included in the review. 28 gargle-swab comparisons were pooled in meta-analysis. Gargle sensitivity was estimated to be 92.7% (95% confidence interval: 89.9% to 94.8%) and 3.2% (0.4% to 6.0%) less sensitive than swab collection. Gargle sensitivity was greater than 87.0% across diverse patient characteristics, settings, type or volume of gargle liquid, length of gargling time, wait time prior to gargling, and reference swab type or NAAT method used. Greatest sensitivities were observed when gargle sampling for 30 seconds or greater using 5-9 mL of saline. 92.9% (86.9% to 96.2%) gargle sensitivity was observed when there was no required wait time, and sensitivity was 90.4% (87.0% to 93.0%), even when compared to high quality combined nasopharyngeal and oropharyngeal (NPOP) swabs.

Conclusion: Gargle sampling is a sensitive, non-inferior method for SARS-CoV-2 detection across various testing contexts. Implementation of gargle sampling has potential to significantly reduce barriers associated with HCW-collected swabs and would facilitate testing and genomic surveillance for SARS-CoV-2 across diverse settings.

Lay Summary

Traditional methods for respiratory viral testing involve inserting a swab into a person's nose, which can be uncomfortable, and requires services of a healthcare worker (HCW). This issue was heightened by the high volume of testing needed during the coronavirus disease 2019 (COVID-19) pandemic. To address this, gargle sampling methods were developed as new, simple, more comfortable, self-collection methods for COVID-19 testing. We thoroughly reviewed the scientific literature and identified studies that would more directly help address the question: "how well do gargle sampling methods work to detect the COVID-19 virus?". Combining information from 24 studies, where enough data was available for analysis, we found that gargle sampling detected the COVID-19 virus as well as swab-based testing methods performed by HCWs. More widespread adoption of gargle sampling may decrease discomfort and barriers to COVID-19 testing. In the future, this method may also be applied for testing for other respiratory viruses.

Preface

This thesis is original, unpublished, and independent work by the author I. Kayda. Content described throughout Chapters 1-5 utilizes only secondary data and does not require ethical approval to be obtained. Stages of the systematic review requiring more than one reviewer, including screening, data extraction, and quality assessment, were completed in collaboration with co-reviewers K. Lechiile and M. Woo Kinshella, and statistical analyses and data synthesis were conducted with consultation from biostatistician J. Bone. The search strategy was developed by I. Kayda and reviewed by biomedical librarian D. Giustini prior to conducting searches. I. Kayda personally led and was involved in all aspects of this work, which was designed and carried out under the supervision of Dr. D. Goldfarb and Dr. P. Lavoie. Contents of this work has been presented by I.Kayda as abstracts, posters, and oral presentations at institutional conferences, and the author will develop a manuscript for journal publication from this work.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACE2: angiotensin-converting enzyme 2

ARDS: acute respiratory distress syndrome

BAL: bronchoalveolar lavage

cDNA: complementary deoxyribonucleic acid

CE-IVD: Conformité Européenne for In Vitro Diagnostic Medical Devices

CI: confidence interval

COVID-19: coronavirus disease 2019

Ct: threshold cycle

DFA: direct fluorescent antibody

E: envelope

EUA: Emergency Use Authorized

FDA: Food and Drug Administration

GLMM: generalized linear mixed model

HC: Health Canada

HCW: healthcare worker

HMPV: human metapneumovirus

I^2: I-squared heterogeneity

IPAC: infection prevention and control

LDT: laboratory developed test

MERS-CoV: Middle East respiratory syndrome-related coronavirus

MIS-C: multisystem inflammatory syndrome in children

MOOSE: Meta-Analysis of Observational Studies in Epidemiology

N: nucleocapsid

NAAT: nucleic acid amplification test

NP: nasopharyngeal

NPOP: nasopharyngeal and oropharyngeal

NPIs: non-pharmaceutical interventions

OP: oropharyngeal

ORF: open reading frame

PCR: polymerase chain reaction

PICU: pediatric intensive care unit

PPE: personal protective equipment

PRISMA-DTA: Preferred Reporting Items for A Systematic Review and Meta-Analysis of Diagnostic Test Accuracy Studies

PRISMA-P: Preferred Reporting Items for Systematic Review and Meta-analysis Protocols

PRISMA-S: Preferred Reporting Items for Systematic Review and Meta-analysis for Reporting Literature Searches

PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analysis

PROSPERO: International Prospective Register of Systematic Reviews

QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies 2

RADT: rapid antigen detection test

RdRp: RNA-dependent RNA polymerase

RNA: ribonucleic acid

RNase P: ribonuclease P

RSV: respiratory syncytial virus

RT-PCR: reverse transcription polymerase chain reaction

S: spike

SARS-CoV: severe acute respiratory syndrome coronavirus

SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

TAT: turnaround time

WHO: World Health Organization

τ^2: tau-squared heterogeneity

Acknowledgements

This journey has been challenging yet completely rewarding, and I have many to thank for believing in me every step of the way. I would like to first extend my thanks to the Experimental Medicine graduate program and my professors at UBC for the opportunity to conduct my graduate studies with you. The knowledge I've gained and skills I've acquired during my time in this program will undoubtedly serve me well as I move forward in my academic and professional life.

I give my sincerest thanks to Dr. David Goldfarb for serving as my supervisor and for believing in me since day one. From encouraging my independence and confidence as a researcher, taking time out of your busy day to answer my questions, to opening new opportunities for my professional development, and inspiring me to continue pursuing work that is thoughtful, creative, team-based, and community-informed. The lessons I've learned from working with you have made such a positive impact on my life and career, and I will carry them with me as I move into my next chapter. Thank you for being the absolute best mentor a young professional could ask for. There are not enough words to fully express my gratitude for your support and guidance these last few years, and I look forward to more collaborations together in the future.

I extend my gratitude to my incredible colleagues Kwana Lechiile, Maggie Woo Kinshella, and Jeff Bone for their expertise and collaboration throughout the development of my first ever systematic review and meta-analysis. Thank you for supporting me through every single step of this project. I have learned so much from each of you and this work would truly not exist without each of your contributions. I would also like to thank Dr. Pascal Lavoie for taking the time to serve on my supervisory committee. Thank you for offering thorough and

critical feedback that has furthered the quality of my work and strengthened my critical thinking as a researcher.

I would also like to thank my dedicated, hard-working colleagues in the clinical microbiology and virology laboratory of BC Children's and Women's Hospital. Thank you for taking the time to answer all of my questions, for showing me the ropes in the laboratory testing world, for demonstrating by example that kindness and team spirit truly overcome any challenge, and for making the lab the absolute best place to work. I have learned more valuable lessons from all of you than any textbook or paper could ever teach me. It has been an absolute honor and joy to work alongside you these last few years.

Thank you to my wonderful parents, sisters, and dogs for your unconditional love and support my entire life. To my parents, your courage, hard work, and sacrifices have given me the most beautiful life and you inspire me daily to pursue my dreams fearlessly – I am eternally grateful for you. I also extend my thanks to my incredible friends and my dance community for the laughs, support, and adventures. And thank you for showing me that it's okay to take a break sometimes and just dance it out – in fact, you'll come back with a clearer head and better ideas.

My heartfelt gratitude goes out to everyone in my closest circle – thank you for walking with me through the many obstacles and for being there to celebrate all of the wins. Finally, thank you for always reminding me of just how many joys there are to appreciate in every single day.

For my family, friends, and microbiology laboratory colleagues.

Chapter 1: Introduction

1.1. Background

1.1.1 Introduction to respiratory virus SARS-CoV-2 and its impact

Coronaviruses are a group of commonly circulating viruses known to infect diverse animals, such as mammals and birds. They are composed of a single-stranded, positive-sense ribonucleic acid (RNA) genome surrounded by a protein capsid and envelope. Coronaviruses also have one of the longest genomes among RNA viruses measuring at a length of 26-32 kilobases [1]. Coronavirus infection in humans, including coronaviruses responsible for the common cold, typically lead to mild to moderate upper respiratory illness with potentially some gastrointestinal symptoms [1-3]. However, some specific coronaviruses have notably been highly pathogenic, and their spillover into the human population during the last two decades have led to severe illness and resulted in public health emergencies. These include the severe acute respiratory syndrome coronavirus (SARS-CoV) in 2002 and the Middle East respiratory syndrome-related coronavirus (MERS-CoV) in 2012 [3, 4]. Most recently, a novel coronavirus emerged in December 2019. Due to its high viral transmissibility and widespread global travel in the present day, it quickly eclipsed case numbers and geographic range reported in previous SARS-CoV and MERS-CoV outbreaks. The International Committee on Taxonomy of Viruses named this novel coronavirus the “severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)” in February 2019. Given its rapid spread and the escalating number of infected cases being reported across the globe, the World Health Organization (WHO) officially declared the SARS-CoV-2 outbreak a pandemic on March 11, 2020 [1, 2].

SARS-CoV-2 is a novel betacoronavirus that belongs to the family *Coronaviridae* and it has similar characteristics to other known coronaviruses. In addition to its RNA genome, the complete, infectious SARS-CoV-2 viral particle is composed of several viral proteins allowing for successful infectivity. These include envelope (E) proteins comprising the viral envelope, surface spike (S) proteins which bind to surface human angiotensin-converting enzyme 2 (ACE2) and mediate viral entry into human host cells, and RNA-dependent RNA polymerase (RdRp) which facilitates successful viral replication to produce new SARS-CoV-2 virions capable of continuing infection [1, 3, 5]. Given its high affinity for ACE2 and the widespread presence of these receptors on epithelial cells of the human respiratory tract, including the nasal, oral, and throat regions, SARS-CoV-2 successfully infects the upper respiratory tract of humans. In some cases, SARS-CoV-2 migrates to infect regions of the lower respiratory tract, such as the alveolar epithelial cells of the lungs as well [2, 3, 6]. Viral replication of SARS-CoV-2 results in cell damage and cell lysis, as well as initiation of immune responses, leading to illness and widespread inflammation of human host tissues [1, 2].

SARS-CoV-2 infection and viral replication causes an infectious disease known as "coronavirus disease 2019 (COVID-19)". It manifests clinically as an array of symptoms ranging from mild flu-like symptoms to severe, life-threatening illness, however asymptomatic COVID-19 cases have also been reported [3]. COVID-19 symptoms are similar to symptoms of other common respiratory infections and include fatigue, chills, fever, headaches, cough, sputum production, dyspnea, and sore throat, although other symptoms including nausea, vomiting, diarrhea, chest pain, and olfactory and gustatory sense disturbances have also been reported in some COVID-19 patients [1-3, 6]. While asymptomatic and mildly symptomatic infections occur, some COVID-19 patients present with severe symptoms requiring immediate, intensive

medical attention. SARS-CoV-2 infection in some COVID-19 patients escalates to life-threatening conditions including acute respiratory distress syndrome (ARDS), septic shock, pneumonia, coagulation dysfunction, respiratory failure, severe systemic inflammation, and multiorgan failure, with several cases resulting in death [1-3]. While previous research demonstrates that the entire population, irrespective of age, is susceptible to SARS-CoV-2 infection, the severity of resulting COVID-19 illness appears to differ across age groups [2]. The median age of infection is 50 years old and older age has been associated with greater COVID-19 morbidity, greater admission rates to intensive care, progression to ARDS, and more severe outcomes, including greater rates of mortality [2, 3]. Children may present asymptomatically or with mild COVID-19 disease including fever, fatigue, cough, and/or gastrointestinal symptoms, with their recovery occurring within one to two weeks following disease onset. However, several reports have described that a subset of children have also been hospitalized with severe COVID-19 illness. These pediatric patients required medical attention for respiratory distress and severe inflammatory conditions associated with a SARS-CoV-2 infection, such as "multisystem inflammatory syndrome in children (MIS-C)". The clinical care of these young patients was reported to require respiratory support, vasopressor support, and admission to a pediatric intensive care unit (PICU) to aid their recovery [7-11]. Although many individuals recover from an acute COVID-19 illness, a portion of the population including the elderly, some children, and those living with comorbidities or chronic conditions, such as recent organ transplant, diabetes mellitus, hypertension, immunodeficiency, or chronic cardiovascular or respiratory disease, are at increased risk for severe illness or death as a result of a SARS-CoV-2 infection [12]. Epidemiological findings by the WHO reveal that the COVID-19 pandemic has resulted in over 700 million confirmed cases with over 6.9 million deaths reported globally across over 230

countries, territories, or areas, since its initial emergence three years ago and new infections continue to occur in the era of COVID-19 vaccines and emerging therapeutics [3, 13].

Ongoing research and clinical case reports describe that many recovered individuals struggle with ongoing, debilitating symptoms as a result of their prior SARS-CoV-2 infection, a condition referred to as "long COVID". This condition is believed to be due to damage caused by inflammation and immune responses to a prior SARS-CoV-2 infection, it adversely affects numerous organ systems, and no effective treatment for long COVID has yet been approved. While acute SARS-CoV-2 viral infection is often considered to be a predominantly respiratory illness, over 200 diverse symptoms have been reported to be associated with subsequent long COVID illness including neurological (i.e. "brain fog", memory loss, dizziness, tinnitus, light sensitivity), cardiovascular (i.e. endothelial dysfunction, myocarditis, an increased risk of deep vein thrombosis), respiratory (i.e. dyspnea), and gastrointestinal symptoms (i.e. abdominal pain, nausea, constipation, heartburn) [14]. Onset of long COVID symptoms follow acute SARS-CoV-2 infection and are estimated to impact at least 10% of all infected individuals, resulting in at least 65 million individuals believed to be struggling with long COVID. Notably, long COVID has been reported to affect individuals of all ages including children, as well as non-hospitalized individuals who had only mild acute SARS-CoV-2 infection, and at least 10-12% of infected individuals who were vaccinated against COVID-19. However, these estimates likely underestimate the true disease burden of long COVID given the novel nature of the disease, as well as the wide array of clinical presentations, challenges with diagnosis, and likelihood of unreported cases [14]. This existence of long-term, post-infection health consequences of the SARS-CoV-2 virus, beyond the initial acute infection, further underscore the importance of identifying new infections and continuing to minimize viral transmission.

During the early stages of the COVID-19 pandemic, vaccines and therapeutics were not yet available, and countries around the world focused instead on mass testing and non-pharmaceutical interventions (NPIs) to control SARS-CoV-2 transmission [15]. Mass respiratory viral diagnostic testing was offered to the public at medical centers and designated testing sites to identify those infected with the SARS-CoV-2 virus, however, guidelines on who was eligible for testing differed across regions and temporally throughout the course of the pandemic. Some guidelines allowed testing for all, while others prioritized targeted testing for only symptomatic individuals, those with known exposure to a SARS-CoV-2 positive individual, frontline workers, or those believed to be at high risk for severe COVID-19 illness due to age or existing health status [16]. Those identified as being infected with SARS-CoV-2 were quarantined through physical isolation, and contact tracing was undertaken by public health teams to identify those they'd encountered while unknowingly contagious [3, 16]. To curb viral transmission in communities, strict public health measures were implemented and included community NPIs such as social distancing, which required individuals to keep a certain physical distance from others in social settings where they would normally be in close proximity, mass closures of public facilities and gatherings, including workplaces, schools, shops, restaurants, recreation centers, religious centers, entertainment venues, and large-scale events, as well as stay at home orders from local governments and restrictions on local and international travel [2, 16, 17]. Personal NPIs were also encouraged as preventative measures to minimize viral transmission including frequent hand hygiene, covering coughs and sneezes, staying home when feeling unwell, and wearing face masks in public places [3, 16-18]. While mandatory to curb transmission of SARS-CoV-2 and maintain hospital capacity at overwhelmed health centers, these policies resulted in significant strains on modern society [16, 19]. The pandemic presented

economic challenges due to workplace closures, absence of employees from work, and restrictions on travel leading to downstream supply chain and operational issues, as well as reduction in overall worldwide employment [19]. In addition, the pandemic significantly disrupted education due to closures of schools and learning institutions, with estimates of 1.6 million students across 200 countries being affected, in what is considered to be the largest education disruption in recorded history [20]. Furthermore, healthcare systems globally were challenged by supply and staff shortages amid high demands in healthcare due to influx of COVID-19 patients and mass testing needs. Intensive infection prevention and control (IPAC) measures were introduced, standard hospital operations such as non-emergent surgeries and clinical visits required rescheduling to divert resources to care of COVID-19 patients, and general visitor bans were put in place at health centers to minimize risk of SARS-CoV-2 transmission [16, 21]. Although the development of COVID-19 vaccines and therapeutics have since prompted public health measures to be removed and decreasing cases numbers prompting the WHO to announce the end of COVID-19 as a global health emergency, long-term consequences of the SARS-CoV-2 virus continue to impact society in a multitude of ways. Ongoing efforts are needed to address these impacts and to remain vigilant for the emergence of novel SARS-CoV-2 strains and spillover of newly emerging viruses [2, 3].

1.1.2 Diagnostic testing for respiratory viruses and existing sample types

Even prior to the onset of the COVID-19 pandemic, the circulation of respiratory viruses presented a high cost to the healthcare system and a significant global disease burden, particularly for young children and the elderly in long-term care facilities [22-25]. In young children globally, respiratory viral infections are a significant cause of morbidity and mortality,

with high rates of hospital admission occurring seasonally due to infection with commonly circulating viruses such as influenza A and B, parainfluenza, respiratory syncytial virus (RSV), enterovirus, adenovirus, rhinovirus, and coronaviruses [23, 24]. In addition to the patient's clinical symptoms and other relevant history, laboratory testing is commonly used to diagnose which pathogen is responsible for the presenting illness. Timely and accurate identification of the etiological pathogen is important as it allows for patient care decisions to be made regarding use or discontinuation of antibiotics, eligibility for certain antiviral medications (i.e. oseltamivir for influenza infection), as well as informing implementation of IPAC measures and large-scale public health surveillance efforts, which obtain regional, national, and international data from clinical isolates and results routinely submitted by clinical labs [22, 24, 26, 27]. Furthermore, certain population groups, such as febrile infants under 3 months old, patients who are immunocompromised, have history of prematurity, or have certain comorbidities, may be at an increased risk of virus-specific complications or death from infection with certain pathogens. Therefore, respiratory viral testing to identify the specific pathogen responsible for infection in these patients may be an especially important use of the clinical laboratory if it informs rapid initiation of appropriate clinical care [24].

Traditional laboratory methods for identifying respiratory viruses include direct fluorescent antibody (DFA) testing, lateral flow chromatography used in rapid antigen detection tests (RADTs), and virus isolation in cell cultures [23, 27-29]. While these have been used historically and have proven to be effective, there are several drawbacks to these methods. Viral culture methods can be challenging in that some viruses cannot be cultured successfully in the laboratory setting, either at all or in an appropriate amount of time. It is vital that diagnostic tests have a fast turnaround time (TAT), from sample collection to result reporting, to quickly inform

decision-making for patient management and outbreak identification [27]. However, viral culture is a time-consuming detection method where results may take many days to obtain and not arrive in time to impact patient care. In contrast, virus-specific RADTs detect presence of viral antigen, are simple to perform, and offer results rapidly within minutes, however certain viruses do not have a commercially available RADT [28]. Furthermore, diagnostic tests require certain performance characteristics to ensure accurate and reliable results are produced by the test. The ability of a test to correctly identify the presence of a disease when it is present, known as "sensitivity", and to correctly identify the absence of disease when it is absent, known as "specificity", are both highly important properties of a diagnostic test. These are both commonly quantified as percentages, where 100% sensitivity indicates all cases were correctly identified and no false negatives were reported, and 100% specificity indicates no false positive results were reported [30]. While many currently approved tests vary in their performance, those with sensitivities and specificities in the 90% range tend to be accepted for clinical use. Although useful in certain clinical scenarios, some RADTs have been shown to have lower sensitivities for detection of respiratory infections, such as RSV, with false negative results being reported and presence of infection being missed in some infected individuals [31]. Finally, these traditional methods often only target a single pathogen for detection during the test, with some immunofluorescence-based methods allowing detection of up to only five pathogens at once [32]. However, many circulating viruses may be the potential etiological agent, or co-infection with a number of different respiratory viruses may be responsible for the presenting illness. Previous research has found presence of more than one respiratory virus in 10% of clinical samples, with higher proportions found in samples obtained from vulnerable populations such as immunocompromised individuals and young children [32]. Given that co-infections may

potentially increase severity of illness, such as in infants with severe bronchiolitis who are co-infected with RSV and human metapneumovirus (HMPV), the capability of efficiently identifying multiple pathogens with a single test would be beneficial [32, 33]. The limitations described for many of the traditional laboratory methods have fueled development of newer diagnostic technology that optimizes molecular biology to identify infectious diseases.

The development of molecular diagnostic methods such as nucleic acid amplification tests (NAATs), which rely on detection of genetic material for pathogen identification, have addressed many of the limitations of older, non-molecular methods [27]. Previous research has demonstrated that molecular methods detected up to 50% more respiratory pathogens in adults and children than non-molecular methods [28]. NAATs, including real time reverse transcription polymerase chain reaction (RT-PCR) tests, have become routinely used in clinical settings for the targeted detection of common respiratory viruses with an RNA genome such as influenza A and B, RSV, HMPV, rhinovirus, and human coronaviruses. When compared to non-molecular methods, respiratory viral testing with RT-PCR has been shown to be more sensitive and specific than RADTs, DFA testing, and viral cultures. In addition, RT-PCR provides results faster than viral cultures, with test results available under one hour for some commercial RT-PCR tests such as the Cepheid GeneXpert [23, 27, 29, 32, 34]. Furthermore, RT-PCR testing has capacity for targeted detection of multiple pathogens simultaneously from a single clinical sample, known as "multiplexing", with some able to test for up to 15-20 respiratory pathogens at one time [32]. Clinical laboratories with sufficient capacity and reagents can design their own laboratory developed tests (LDTs) using multiplex RT-PCR targeting selected pathogens of interest. However, several commercial RT-PCR tests, such as the FilmArray Respiratory Panel 1.7, are readily available and allow testing for many common viruses and some bacteria known to be

responsible for upper respiratory tract infections [28, 32, 35]. As molecular diagnostic testing advances, emerging commercial RT-PCR tests are being designed as fully automated, "sample in and result out" tests, where all reaction and analysis steps take place in a closed system with minimal sample handling required on the part of laboratory technicians [28]. These advances simplify the respiratory viral diagnostic process, lower risks for contamination with less sample handling needs, and increase testing throughput and efficiency in the busy clinical laboratory.

The process of RT-PCR testing requires specialized instrumentation, chemical reagents, and a biological sample from an individual with suspected viral infection. RT-PCR uses reagents, including specially designed primers, which target known genetic sequences specific to the virus of interest. RT-PCR testing involves an *in vitro* chemical reaction where targeted viral RNA, that is potentially present in a clinical sample, is first reverse transcribed to synthesize a complementary deoxyribonucleic acid (cDNA). Using a thermocycler machine, this cDNA then undergoes many cycles of amplification by polymerase chain reaction (PCR) to produce more copies of the cDNA. Fluorescent probes added to the reaction lead to fluorescence signals being emitted as PCR continues and more cDNA copies are produced, and these signals may be recorded in real time as the reaction progresses. If the amount of fluorescence signal surpasses a certain threshold within a pre-set number of amplification cycles, the sample is considered to contain viral material and to be "positive" for the specific virus targeted in the RT-PCR reaction [36-40]. The amplification cycle at which the fluorescence signal passes the threshold is considered the "threshold cycle (Ct)". In addition to qualitative positive or negative results, the Ct value may serve as a semi-quantitative result of RT-PCR testing for comparison between tested samples, so long as all reaction and analysis steps were completed with the same reagents, assay, and instruments. As less amplification cycles are needed to pass the threshold for

detection, lower Ct values tend to be associated with higher amounts of starting viral material in the sample [23, 36]. Finally, clinical RT-PCR tests also tend to include internal controls that target known human genes, such as human ribonuclease P (RNase P), in addition to viral genes, to serve as an indicator of sample quality for the collected biological sample submitted for testing [41, 42].

A key component in diagnostic testing for upper respiratory viral infections, even prior to testing with RT-PCR, is the collection and pre-analytical processing of a high-quality biological sample from an individual with suspected infection. Several factors influence the detection of a respiratory pathogen including the anatomic site sampled, type of sample collected, sample transport and storage, and the pre-analytical processing of the sample (i.e. RNA extraction) ahead of RT-PCR testing [27]. Samples may be collected from several anatomic regions of the upper respiratory tract including the nasopharynx, oropharynx, anterior nares, mid-turbinate nares, oral cavity, or a combination of these regions. Sample types include swabs, washes, and aspirates of these regions to collect respiratory epithelial cells and any potential viral material that may be present [27, 29, 43, 44]. However, in some clinical cases, lower respiratory tract samples such as bronchoalveolar lavage (BAL) or induced sputum may also be collected [27]. One of the most common sample types recommended for RT-PCR respiratory viral testing is the nasopharyngeal (NP) swab collected by a healthcare worker (HCW), although combined nasopharyngeal and oropharyngeal (NPOP) swabs have also been used as the gold standard [43-45]. In addition to the collection site, the type of swab influences the sample quality. The swab should ideally be designed to allow for efficient uptake of respiratory epithelial cells and release of collected cells into a liquid collection media [22]. Previous research has demonstrated that nylon flocked swabs, which are composed of nylon fibers sprayed on and arranged radially, were

superior to swabs composed of rayon or cotton tips, yielding two- to threefold more total and infected respiratory cells. [22, 27, 23, 43]. In addition, swabs composed of materials such as cotton, calcium alginate, or wood should be avoided as these materials may inhibit the PCR reaction, while swabs composed of a plastic or aluminum shaft are recommended for use [46, 47]. Immediately after collection, respiratory swab samples are often eluted in liquid viral transport media in a sterile container and refrigerated to minimize degradation of viral RNA prior to testing. Samples are recommended to be tested within 48-72 hours of collection, or frozen at -70 to -80 degrees Celsius if testing is delayed [27, 29, 47, 48]. In preparation for RT-PCR testing, collected samples typically undergo a pre-analytical processing step known as "RNA extraction" to separate and concentrate the RNA from mucous, proteins, lipids, DNA, and other cellular materials also present in the heterogeneous biological sample. Once extracted, RT-PCR is performed on the purified RNA for target detection of viral RNA [49, 50]. Previous work has found that 46%-68.2% of errors in the testing process occur during this pre-analytical stage and include issues such as improper sample collection, improper storage and transport of samples, or cross contamination during sample handling [51]. Proper sample collection from the appropriate anatomic site and pre-analytical processing are both highly important for downstream analysis with RT-PCR testing, the accuracy of test results, and minimizing false negative results that may occur due to poor sample quality or pre-analytical errors [46, 48, 52, 53]. Overall, the implementation of newer molecular methods, such as RT-PCR, in combination with high-quality sample collection in routine testing have improved capacity for diagnosing respiratory viral infections in a rapid, sensitive, specific, and efficient manner.

1.1.3 COVID-19 testing challenges and the emergence of novel sample types

Accessible, rapid, and sensitive diagnostic testing has been vital during the COVID-19 pandemic for identifying new SARS-CoV-2 infections, managing outbreaks, the clinical care of COVID-19 patients, and minimizing viral transmission [47, 54]. However, the rapid spread of SARS-CoV-2 and the high testing demand have significantly challenged clinical laboratories and existing health infrastructure. Like molecular diagnostic testing for other respiratory viruses, the gold standard method for COVID-19 testing involves upper respiratory tract swab collection by a HCW and subsequent RNA extraction and testing with an RT-PCR assay targeting SARS-CoV-2 genetic material. Various sample types have been clinically validated for use in COVID-19 testing globally including the collection of NP swabs, oropharyngeal (OP) swabs, or NPOP swabs, although their individual use has varied by health authority and country [44, 47, 52, 54-56]. Once the initial SARS-CoV-2 sequencing data became available, numerous novel RT-PCR assays were developed and validated for SARS-CoV-2 detection, including LDTs developed by clinical labs and commercially manufactured tests. These RT-PCR tests targeted various SARS-CoV-2 genes for detection including the E, S, RdRp, nucleocapsid (N), and several open reading frame (ORF) genes, or any combination of these. They also targeted various genes to control for sample quality such as the human RNase P gene [42, 46, 48, 54, 57]. In Canada, the collection of NP swabs from the nasopharynx by a trained HCW was commonly used for clinical diagnosis of COVID-19, particularly early in the pandemic before the development of rapid antigen tests [41, 58, 59]. Although standard respiratory viral testing workflows were quickly adapted for COVID-19 testing, numerous challenges arose given the high volume of swab sample collections and tests that needed to be performed during the pandemic.

Challenges related to traditional swab collection were reported early in the pandemic and limited testing capacity globally. These included supply chain issues, resource shortages, and low acceptability from patients. COVID-19 testing was being performed on an unprecedented scale and required HCWs to don personal protective equipment (PPE), such as gloves, gowns, eye protection, and masks, to collect upper respiratory tract samples by swabbing hundreds of potentially infectious individuals daily. There were increasingly high global demands for the specialized swabs, viral transport media, and PPE required for respiratory sample collections, which led to widespread supply chain shortages of these resources [47, 52, 53, 60-64]. Furthermore, shortages of HCWs were being reported globally for several reasons and this also impacted efficient sample collection and testing. At the time, respiratory samples were required to be manually collected by a trained HCW for quality assurance and no sample types that could be reliably self-collected by individuals were available [53]. Throughout the pandemic, the high workload of HCWs was overburdened and challenged by staff shortages, partly due to staff absences due to illness and quarantine requirements. HCWs were also required to juggle increasing patient care responsibilities with the growing demand for COVID-19 sample collections, and the number of patients presenting to clinical sites daily for testing far outnumbered the amount of HCWs properly trained and available to perform swab collections. This challenge was further amplified in regions where HCWs were already in limited number prior to the pandemic such as rural, remote, or resource-constrained communities [47, 59]. From the patient perspective, individuals were required to present to a designated COVID-19 testing site and wait for a HCW to collect their swab sample. This requirement posed difficulties for those who did not have access to a personal vehicle or had to travel long distances to reach a testing site, particularly individuals in rural and remote areas, where reports demonstrated that

members of these communities were even less likely to receive testing [61, 63]. In addition, the above-mentioned challenges taken together often led to increasingly long lines at COVID-19 testing sites and burdensome wait times of several hours for sample collection by a HCW [65-69]. The experience of having an upper respiratory swab collected was often described as uncomfortable or painful by patients of all ages, with young children having particular difficulty tolerating the procedure. Undergoing multiple NP swab collections was also notably challenging for those requiring repeat testing for SARS-CoV-2, either for clinical care decisions or COVID-19 screening for work or travel purposes [41, 52, 59, 61, 70]. Although a generally safe procedure, there have been reports of complications related to upper respiratory swab collection including severe epistaxis, cerebrospinal fluid leakage requiring endoscopic surgery, and broken swabs being retained in the upper respiratory tract prompting endoscopic removal [71]. Over the course of the pandemic, NP swab collection has received low acceptability ratings from patients and resulted in testing fatigue, or even aversion to undergo COVID-19 testing, amongst the public [41, 59, 61]. In addition, NP swab collection requires close proximity between the potentially infectious patient and the HCW collecting the sample. As the collection may trigger a sneeze or cough from the patient, there are also potential risks of infection to the HCW performing an NP swab collection [48, 72].

To address these challenges with standard swab collection, alternative sample collection methods for COVID-19 testing have been investigated. Research has explored the use of self-collection of samples by patients to reduce reliance on and risk of exposure for HCWs, preserve PPE, lessen requirement for presentation to a designated testing site, and minimize sample collection bottlenecks at testing sites. In addition, collection of samples from less invasive anatomic sites have also been investigated to lessen patient discomfort and improve

acceptability of testing [41, 58, 72, 73]. The self-collection of mid-turbinate nares, anterior nares, oral swabs, or combined oral-nasal swabs, which could be collected reliably by a patient without HCW involvement, have been proposed. Although these methods still rely on availability of specialized swabs and viral transport media [58, 74-76]. Self-collection of swab-independent sample types have also been explored to circumvent supply chain issues and the need for HCW collection at testing sites, including the self-collection of saliva [41, 74, 77-81]. While systematic reviews and meta-analyses have reported acceptable performance of saliva compared to NP swabs for SARS-CoV-2 detection, there have been some drawbacks with saliva collection as well [82, 83]. These include the potential difficulties for some individuals, such as those with sicca syndrome or xerostomia, to produce a sufficient volume of saliva for testing [74, 81]. Furthermore, the heterogeneous composition and high viscosity of saliva poses challenges for laboratories, as it may clog automated liquid handlers and may require additional pre-analytical processing steps, that would not normally be required for swabs in viral transport media, to sufficiently prepare the saliva sample for RT-PCR testing [41, 81]. Over the course of the pandemic, several groups globally have sought a novel sample type for COVID-19 testing that ideally possessed some of the following characteristics including being 1) swab-independent, 2) simple and reliable to self-collect, 3) minimally invasive, 4) stable for transport, 5) amenable for high throughput testing, and 6) maintains acceptable performance for SARS-CoV-2 detection with RT-PCR when compared to traditional upper respiratory swabs.

Gargle sampling has emerged as a novel method for respiratory sample collection that possesses these characteristics. Two studies published in October 2017 and January 2019, prior to onset of the COVID-19 pandemic, reported that self-collected sterile water or saline gargle samples are an acceptable sample type for RT-PCR testing for common respiratory viruses,

including influenza A, influenza B, and RSV [84, 85]. Several groups around the world have since developed variations of this gargle collection method during the pandemic and independently investigated its performance for COVID-19 testing when compared to a gold standard upper respiratory swab [86-97]. This less invasive, self-collect, swab-independent method involves gargling a small volume of liquid, such as saline or water, for a brief period of time to facilitate collection of a sample from an individual's mouth and throat. This sample, also referred to as "throat washing" or "gargle lavage", may then be forwarded for downstream molecular testing with RT-PCR, with little to no additional pre-analytical steps or transport media required. Gargle sampling has received high acceptability ratings from patients, with greater acceptability than saliva being reported in head-to-head comparisons of the two swab-independent, self-collect methods [41, 81]. Self-collection of gargle samples has also been demonstrated to have good performance on various RT-PCR tests detecting several SARS-CoV-2 genes, even when the collection is performed independently and unobserved by a HCW [59]. Furthermore, clinical saline gargle samples have been demonstrated to be stable at room temperature, with no added preservative or transport media, for at least two days [41]. While gargle samples artificially spiked with SARS-CoV-2 and mixed with a preservative buffer have been reported to be stable at room temperature with detectable viral RNA when tested 31 days later [45]. Furthermore, self-collection of gargle samples allows for the delocalization of sample collection from designated testing sites, which may improve accessibility and convenience for those seeking to be tested for SARS-CoV-2 infection [59, 63, 77, 98]. Given these benefits of gargle sample collection, several regions have notably introduced gargle-based COVID-19 testing in non-clinical, community settings during the pandemic. This has allowed for mass screening programs and sample collections for COVID-19 testing to take place at home, in

workplaces, schools, or other public settings, before being submitted to a clinical laboratory for RT-PCR testing [99-106]. Overall, these beneficial characteristics of novel gargle sample collection have potential to lower barriers to COVID-19 testing and address resource constraints, particularly in settings where trained HCWs or specialized swabs are minimally available.

1.1.4 Assessing diagnostic performance of novel sample types

Prior to the implementation of a novel diagnostic sample, its performance must first be assessed and meet an acceptable standard for clinical use. Clinical laboratories commonly achieve this on the individual level by conducting a validation study, where the novel approach is compared to an existing, clinically accepted method. Clinical validations of a novel sample may be performed by collecting paired samples from patients undergoing routine diagnostic testing for the condition of interest. The performance of the novel sample collection method may then be assessed by head-to-head comparison to the clinically accepted method, where all variables including patient population sampled, laboratory processing, and diagnostic testing approach remain the same between methods. Systematic reviews and meta-analyses of these primary studies may then be conducted to consolidate evidence and further offer summarized conclusions regarding diagnostic performance of the novel sample with greater power than any single validation study alone. In the COVID-19 testing literature, several clinical validations of alternative samples such as saliva, nasal swabs, and gargle samples had their performance directly compared against HCW-collected upper respiratory tract swabs, such as NP or NPOP swabs, as these were the accepted collection method for COVID-19 diagnosis [41, 56, 58, 77, 87, 88]. Furthermore, systematic reviews and meta-analyses investigating diagnostic performance of

novel samples, such as saliva, against routine swab collection for COVID-19 testing have also been reported [82].

These investigations often focused on estimating clinical diagnostic sensitivity, as an indicator of diagnostic performance, to understand the proportion of SARS-CoV-2 positive cases detected by the novel method compared to those detected by standard swab collection in a clinical setting. This sensitivity estimate is quantified as a percentage where 100% sensitivity indicates all positive cases with true disease were correctly identified, with no false negative results reported. Importantly, this sensitivity estimate is distinct from the estimate of "analytical sensitivity", which instead focuses on a test's ability to detect the lowest amount of material for positive detection, with lower detection concentrations being associated with higher analytical sensitivity. Notably, high analytical sensitivity may not be associated with acceptable clinical diagnostic sensitivity [107]. This is exemplified in cases where individuals may have recovered from a previous infection yet have traces of potentially non-viable viral material detected by upper respiratory swab collection and NAAT for a period of time following initial infection and symptom resolution, which has been reported in some COVID-19 patients [108]. If the analytical sensitivity is high and allows for detection of small amounts of residual viral material, this result may not necessarily be clinically appropriate if individuals who may not be actively infected are classified as having disease. Hence, the important distinction between analytical sensitivity and clinical diagnostic sensitivity and our focus on estimation of the latter in this work.

Several approaches may be used to calculate an estimate of clinical diagnostic sensitivity. In the absence of a perfect reference test, it is common practice to use a "composite reference standard" where a patient is considered positive for disease if a positive result is obtained on either sample type collected from them [109]. Which in this context would be an individual

being deemed SARS-CoV-2 positive if they obtain a positive result on either a gargle sample, a swab sample, or both. As there have been reports of false negative results occurring with HCW-collected upper respiratory swabs, previous validation studies and meta-analyses have deemed it appropriate and utilized a composite reference standard in their estimation of sensitivity [56, 82, 87]. This approach would allow estimation of sensitivity for both the novel method and the swab method against the composite reference standard, rather than assuming 100% sensitivity for swab collection, as well as allow calculation of the difference in sensitivity between the two methods. This approach would also account for the possibility of false negative results occurring on both sample types. However, use of the composite reference standard with the assumption of any positive result being a true positive result does not allow for accurate estimation of specificity. Given that comparison is being performed between two sample types with NAAT, rather than RADTs or serological tests, and that it is unexpected that sampling of two different upper respiratory anatomic sites would yield false positive results, there is less of a focus on estimating clinical specificity in this context. Instead, there is a greater concern for missed detections and false negative results being reported with a novel sample collection method. Thus, estimation of sensitivity is prioritized is this case to determine if the diagnostic performance of the novel sample is comparable to the clinically accepted collection method. This approach has been similarly used in previous clinical validation studies comparing gargle sampling to swab collection, and prior systemic reviews and meta-analyses of self-collected saliva as an alternative to HCW-collected swabs for SARS-CoV-2 detection [56, 82, 87]. It has also been utilized for clinical diagnosis of diverse conditions including pulmonary tuberculosis, where an individual is considered to have the disease if a positive result is obtained by collection of either induced sputum or gastric aspirate samples [109, 110]. Through estimating both the overall clinical

diagnostic sensitivity of a novel method and the difference in sensitivity between the two methods using the composite reference standard, decisions can be made regarding the potential implementation of a novel method based on the comparability of its performance to the method currently accepted for clinical use.

1.2. Rationale

While there has been much work conducted globally on the use of the gargle sampling method for COVID-19 testing, there has also been some variation in its reported performance with some validation studies concluding sensitivity of 60.7% and others reporting sensitivity as high as 100.0% [92, 98]. However, these individual studies sampled varying populations around the world with various types and volumes of gargle liquid, differing gargle collection instructions, and utilized different reference swabs and analytical methods for COVID-19 testing based on clinical guidance in their respective region. In addition, these were often single-center studies calculating gargle sensitivity across relatively small numbers of SARS-CoV-2 positive samples. These small study sample sizes were likely due to feasibility and limitations in collecting and testing study samples during an ongoing pandemic. For greater clarity on gargle sample performance compared to conventional sampling methods, it would be beneficial to consolidate all the available evidence through rigorous systematic review and meta-analysis. Given the novelty of the topic, consolidated evidence is minimally available and only one systematic review and meta-analysis on this topic currently exists. This previous work estimated that gargle sampling achieved high sensitivity (91%) for SARS-CoV-2 detection when compared to NP or NPOP swab collection [111]. However, several questions remain unanswered, and an updated systematic review and meta-analysis is warranted. While performance of gargle

sampling appears to be high, this prior work did not explore the difference in sensitivity between the two sampling methods, did not utilize other possible swab methods for comparison, and included only studies published up to September 2021. Additional studies on this topic have since been conducted and reported, which would allow for further analyses investigating several unexplored sampling variables on gargle performance such as the type of swab material, swab sampling site, and analytical method used. Furthermore, several studies have implemented mandatory wait times since an individual last had something by mouth, including eating, drinking, dental care, and/or smoking. Both the utility of this waiting period and the optimal amount of time needed for greatest gargle sensitivity is currently unclear. This remains an unexplored yet important practical consideration for implementation of this novel collection method in real-world settings. While gargle sampling has promise, further investigation is needed to address remaining questions. Overall, a greater understanding of gargle sample performance in detecting SARS-CoV-2 will inform decisions surrounding its potential use as a novel alternative to traditional upper respiratory swab collection. In addition, a more thorough examination of how gargle samples perform across various contexts will help strengthen implementation guidance and may contribute to more efficient, less resource-intensive, self-collect testing practices for SARS-CoV-2 and other respiratory viruses in the future.

1.3. Objective

Our objective is to systematically review the existing medical literature for quantitative evidence and determine the performance of gargle samples, compared to upper respiratory tract swab samples collected by healthcare workers, for SARS-CoV-2 detection by nucleic acid amplification testing in individuals with lab-confirmed or suspected COVID-19.

1) The primary outcome of interest is the overall mean sensitivity of gargle sampling for SARS-CoV-2 detection.

2) An additional outcome of interest is the mean difference in sensitivity between gargle and swab sampling methods for SARS-CoV-2 detection.

3) We also aim to explore variation within studies to investigate how performance of gargle sampling may differ due to real-world differences in sample collection and COVID-19 testing strategies.

Chapter 2: Methods

To explore and collate the available literature on performance of gargle sample collection for COVID-19 testing, we carried out a systematic review and meta-analysis. Rigorously conducted systematic reviews and meta-analyses, that follow pre-specified guidelines aimed at minimizing bias and thoroughly summarizing all available literature, are widely recognized as producing important evidence for clinical decision-makers [112, 113]. By first searching for and selecting available research studies in a systematic way, then combining their data in statistical analyses, we could explore similarities and differences across studies, generate new hypotheses for future research, increase the sample size, and summarize the results of multiple studies to produce an overall estimate with greater power than may be possible with a single study alone [112, 113]. High quality systematic reviews and meta-analyses are expected to include several characteristics, which are often decided upon *a priori* during the development of a protocol. These include 1) clearly defined, pre-specified aims, 2) a thorough search of the medical literature for relevant studies, 3) systematic inclusion and/or exclusion of studies based on pre-specified criteria, 4) assessment of study quality for all included studies, 5) appropriate selection of statistical methods for pooling quantitative data into an overall estimate, 6) and the use of stratified and sensitivity analyses to explore variation across studies and test the robustness of pooled results. It is also recommended that selection, data extraction, and quality assessment of studies is performed by more than one reviewer to minimize selection bias and errors in data collection and analysis [112, 113].

Our work was developed with these characteristics in mind and carried out in accordance with several published guidelines to prioritize quality and reproducibility, while minimizing bias in the results. To assist with outlining the scope of our systematic review and meta-analysis, we

first developed a PICO question related to diagnostic tests outlining the population (P), intervention (I), comparator (C), and outcome (O) we were interested in [113-115]. Our systematic review and meta-analysis focused on the population of children and adults providing samples for COVID-19 testing, used self-collected gargle samples including water or saline as the intervention under study, used HCW-collected swabs as the comparator, and focused on COVID-19 diagnosis via NAAT as the outcome. Subsequently, we developed and reported our protocol in accordance with the "Preferred Reporting Items for Systematic Review and Meta-analysis Protocols (PRISMA-P)" guideline and publicly registered it with the "International Prospective Register of Systematic Reviews (PROSPERO)" on Feb 24, 2022 (CRD42022312628) prior to beginning the review as recommended [116, 117]. If changes occurred during the review process, amendments to the protocol were planned to also be reported on PROSPERO. To foster transparency and reproducibility in the conduct and presentation of our work, the following is reported in accordance with the general "Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)" guideline [118]. Given that our work specifically consolidates evidence from observational studies related to diagnostic test accuracy, we also consulted the "Preferred Reporting Items for A Systematic Review and Meta-Analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA)" guideline, as well as the "Meta-Analysis of Observational Studies in Epidemiology (MOOSE)" guideline [119, 120].

2.1. Search strategy

Our search strategy was designed and reported in accordance with the "Preferred Reporting Items for Systematic Review and Meta-analysis for Reporting Literature Searches (PRISMA-S)" guideline [121]. Online databases (Ovid MEDLINE, Ovid EMBASE, Web of

Science, and Global Index Medicus) and online preprint servers (medrxiv, biorxiv, and Research Square) were searched from 14 March 2022 to 5 May 2022 by one reviewer (IK) using advanced search terms, subject headings, and combinations of keywords related to "SARS-CoV-2", "COVID-19", "diagnostic testing", "gargle", "throat lavage", and "swab". These platforms were specifically selected because previous research has demonstrated that optimal search strategies, with sufficient coverage of the medical literature, included searches of MEDLINE, EMBASE, and Web of Science [122]. As online preprint servers were particularly important for the sharing of new COVID-19 related research during the pandemic, these platforms were also selected to search for any relevant studies that may have become available prior to their publication in traditional peer reviewed journals. In addition, previous research has demonstrated that involvement of a librarian in the development of a search strategy is beneficial and correlated with improved reproducibility of searches [121]. Therefore, our search strategy was also reviewed by a biomedical librarian prior to searching. A search strategy specific to the Ovid MEDLINE database is available in Figure 1, and a complete search strategy for all platforms searched can be found in Appendix A. Search dates extended from inception of the online database/preprint server to the present time, and there were no limits placed on language, dates, or study design when conducting searches for potentially relevant studies. All results retrieved in Global Index Medicus, medrxiv, biorxiv, and Research Square searches were reviewed for relevance by one reviewer (IK). Potentially relevant studies were selected after scanning the first 100 results from each individual search on these specific platforms. When a peer reviewed journal publication and its earlier preprint were both found in searches, the preprint was removed as a duplicate and the publication was selected for inclusion. Reference chaining was also used to locate potentially relevant studies, and all references were managed using EndNote software.

Final search results from all databases and preprint servers were imported into Covidence software for subsequent de-duplication, screening, and data extraction stages.

Ovid MEDLINE	
	Ovid MEDLINE(R) and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations, Daily and Versions <1946 to May 05, 2022>
1	exp COVID-19/
2	exp SARS-CoV-2/
3	(covid-19 or covid19 or covid 19 or coronavirus disease 2019 or coronavirus disease-19 or disease 2019, coronavirus or 2019 novel coronavirus disease* or 2019 novel coronavirus infection* or 2019 ncov disease* or 2019 ncov infection* or 2019-ncov disease* or 2019-ncov infection* or covid 19 virus disease* or covid 19 virus infection* or covid-19 virus disease* or covid-19 virus infection* or coronavirus disease 19 or disease, 2019-ncov or disease, covid-19 virus or infection, 2019-ncov or infection, covid-19 virus or infection, sars-cov-2 or sars cov 2 infection* or sars coronavirus 2 infection* or sars-cov-2 infection*).tw,kf.
4	(2019 novel coronavirus or 2019 novel coronaviruses or 2019-ncov or covid 19 virus or covid-19 virus or covid-19 viruses or coronavirus 2, sars or coronavirus, 2019 novel or coronavirus disease 2019 virus or novel coronavirus, 2019 or sars cov 2 virus or sars coronavirus 2 or sars-cov-2 virus or sars-cov-2 viruses or severe acute respiratory syndrome coronavirus 2 or virus, covid-19 or virus, sars-cov-2).tw,kf.
5	1 or 2 or 3 or 4
6	Mouthwashes/
7	(gargl* or (saline adj5 gargl*) or (gargle adj5 lavag*) or (gargle adj5 wash*) or (gargle adj5 rins*) or (throat adj5 lavag*) or (throat adj5 wash*) or (throat adj5 rins*) or mouthwash* or (mouth adj5 wash*) or (mouth adj5 lavag*) or (mouth adj5 rins*) or (oral adj5 rins*) or (oral adj5 wash*) or (oral adj5 lavag*)).tw,kf.
8	6 or 7
9	Diagnosis/
10	exp Nucleic Acid Amplification Techniques/
11	covid-19 nucleic acid testing/
12	(diagnos* or detect* or test* or nucleic acid amplification or nucleic acid or naat or polymerase chain reaction or pcr or rt-pcr or rt pcr).tw,kf.
13	"Sensitivity and Specificity"/
14	9 or 10 or 11 or 12 or 13
15	(swab* or nasopharyngeal or (nasopharyngeal adj5 swab*) or np or (np adj5 swab*) or oropharyngeal or (oropharyngeal adj5 swab*) or nasal or (nasal adj5 swab*) or mid turbinate or (mid turbinate adj5 swab*) or mid-turbinate or (mid-turbinate adj5 swab*) or throat or (throat adj5 swab*) or oral or (oral adj5 swab*)).tw,kf.
16	5 and 8 and 14 and 15

Figure 1. Ovid MEDLINE search strategy

2.2. Screening of search results and selection criteria

Title/abstract and full text screening of search results was carried out by two reviewers (IK and KL), working independently, with discrepancies resolved by a third reviewer (MWK). Included studies were selected if they met all selection criteria outlined below, and excluded studies had the reasoning for their exclusion recorded and reported in Appendix B. Study designs reporting primary data and investigating COVID-19 diagnostic performance of self-collected gargle samples compared to a HCW-collected reference swab sample were included. Eligible studies included, but were not limited to, randomized clinical trials, case series, prospective or retrospective cohort studies, case-control studies, and/or cross-sectional studies. A study required at least five matched samples (gargle and swab) tested with a nucleic acid amplification testing (NAAT) platform within 10 days of diagnosis (if study participants had been diagnosed with COVID-19 prior to study sampling). Eligible study populations included all individuals having matched samples collected for COVID-19 testing, with no restrictions on age or presence of symptoms. The study must have also involved a form of "gargling", "throat lavage", or "throat wash" during the gargle sample collection. Both peer reviewed publications and non-peer reviewed preprints were included. International publications written in languages read by the review team (English, French, or German) were eligible for inclusion. Studies that conducted multiple gargle-swab comparisons with discrete differences in methodology (i.e. different populations sampled, different liquids, tested on different assays etc.) and reported these results separately within the same article, were considered as separate "comparison groups" for the purposes of our analysis. All studies reporting secondary data, study protocols, review articles, opinion papers, editorials, case reports, or studies that did not aim to evaluate the diagnostic performance of gargle samples compared to matched swabs for COVID-19 testing were

excluded. Studies with less than five matched sample pairs, or studies that used "oral rinse" only, with no gargling component, during sample collection were also excluded.

2.3. Data extraction from included studies

The final set of included studies underwent data extraction into a custom data extraction form that can be found in Appendix C. Data extraction for each included study was carried out independently by two reviewers (IK and KL), who first piloted data extraction by doubly extracting three articles before a consensus check together. After completing data extraction, a final consensus check of all the extracted data was also carried out by discussion between them to finalize the data included in the review and meta-analysis. The data extraction form contained fields to capture information related to study information (title, year of publication, first author name, country, language, publication type, main study aim and conclusion, dates of study duration, study design, funding, conflicts of interest, meeting inclusion or exclusion criteria), population characteristics (age, sex distribution, presence of symptoms at time of collection, setting, and whether HCWs or lab-confirmed COVID-19 patients were included or not), reference swab collection (type of swab, sampling site, transport media used), gargle sample collection (type of gargle liquid, volume of liquid, method and time of sampling, wait time since anything by mouth before collection), lab testing protocols (NAAT method, device, test kit, gene targets, criteria for positive result), type and number of samples used for both index and reference testing, and the type and number of samples reported positive for SARS-CoV-2 via gargle, swab, both, or either sample type. When information of interest was not provided in the study report, we contacted study corresponding authors directly by email. When unsuccessful after three attempts, we indicated that the information is unavailable and proceeded with the

existing data where possible. Individual studies were excluded from being pooled in the meta-analysis when the unavailability of data prevented calculation of a study estimate.

2.4. Quality assessment of included studies

Given the nature of the topic, quality assessment of included studies was carried out at the individual study level using an adapted version of the published "Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2)" tool, which is available in Appendix D [123]. This tool is recommended for assessing the methodological quality of diagnostic accuracy studies and has been commonly utilized in previous meta-analyses with similar aims of comparing diagnostic performance of two sample types [83, 111, 123, 124]. As recommended, this published tool was adapted slightly to align with the scope of this specific review and meta-analysis. Quality assessment was carried out independently by two reviewers (IK and KL), who piloted on the first few studies together, and discrepancies were resolved by a third reviewer (MWK). Individual studies were assessed for "risk of bias" and "concern(s) regarding applicability" in four domains related to the study's design and execution: patient selection, reference standard (swab sample), index test (gargle sample), and the flow of participants through the study and timing of tests. A review-specific guidance document was also developed and used by the review team when assessing quality of included studies and is available in the Appendix E. Risk of bias plots (summary bar and traffic light plots) were developed and reported to visualize the quality of the studies included in this review and meta-analysis. For the purposes of visualization with the traffic light and summary bar plots, studies were given the domain-specific assessment of "some concerns" when evidence reported in the article was unclear.

2.5. Data synthesis and meta-analysis

R statistical software (version 4.2.2) and *meta*, *metafor*, *dplyr*, and *MKinfer* packages were used to estimate effect measures, heterogeneity, and to synthesize results from all included studies. The primary analysis estimated the overall mean sensitivity of gargle sampling for SARS-CoV-2 detection with corresponding 95% confidence interval (CI). We also estimated the pooled mean difference in sensitivity (with 95% CI) between gargle and swab sampling methods for SARS-CoV-2 detection. Given the absence of a gold standard test in terms of sensitivity, a composite reference standard was used to define a positive result with either sample (gargle or HCW-collected respiratory swab) testing positive being considered a positive reference standard. This analysis did not allow for an estimation of specificity [109].

Given that it is conventional and recommended for meta-analysis of clinical studies, a random-effects model was selected as the meta-analytical model for our analysis, rather than a fixed-effect model. This model was selected given that its assumptions expect and allow some between-study heterogeneity. In addition, it is considered more applicable for real-world applications when pooling results of studies that are expected to differ because of slight methodological differences [112, 113, 125]. To estimate the mean difference in sensitivity between sampling methods, the Wilson method, which accounts for the paired nature of the sample collections, was used to first estimate the difference in sensitivity, standard error, and 95% CIs for each study [82, 126]. Individual study estimates were then pooled together using random-effects meta-analysis with inverse variance for weighting, where studies with larger sample sizes and greater precision were given greater weight in the overall pooled estimate. The Sidik-Jonkman estimator was selected to estimate tau-squared (τ^2), a statistical measure of between-study variance calculated in the random-effects model. A Hartung-Knapp adjustment

was also applied, which slightly increased the confidence interval for a more conversative estimate of precision in our pooled estimate. These were selected for this meta-analysis to account for the relatively small number of studies pooled, potentially high heterogeneity expected across included studies, and to minimize the possibility of falsely positive estimates of effect size [125, 127-130]. To estimate the overall mean sensitivity of gargle sampling, random-effects meta-analysis with generalized linear mixed model (GLMM) was conducted, which previous research has recommended for meta-analysis of proportions, such as sensitivity [82, 131]. A Hartung-Knapp adjustment was similarly applied for the reasons listed above, while a Sidik-Jonkman estimator was not applicable for this GLMM model.

As recommended, subgroup analyses were also conducted to investigate sources of heterogeneity across the included studies [112, 113, 127]. As individual studies varied in their methodologies, subgroup analysis allowed us to explore any changes in gargle sensitivity estimates that may be attributable to real-world differences in sample collection and COVID-19 testing strategies across studies. Studies were stratified into subgroups based on variables believed to be clinically and/or practically important, given that they may impact gargle sample performance for SARS-CoV-2 detection. These subgroups were defined by the following *a priori* study characteristics of interest such as the type of reference swab(s) used, patient characteristics (age group, presence of symptoms, HCWs included in study etc.), sample collection method used for gargle (type of gargling liquid, length of gargling time, time limit since anything by mouth before collection etc.), type of analytical method used (testing method, gene targets etc.). Due to the variation in laboratory testing protocols observed during data extraction from the included studies, *post hoc* subgroup analyses were also conducted to explore if gargle performance varied depending on whether LDTs or commercial assays were used for

testing study samples. Subgroup analyses were performed if at least three comparable studies/comparison groups were available for pooling statistically.

Data used for classification decisions was obtained from study reports directly or provided from the study authors upon request. Subgroups were defined by the following criteria in this analysis. Studies were stratified by age group based on age eligibility for study participation, including the youngest age eligible. Studies that included any participants under 18 years old were considered to sample "both adults and children", while those that reported excluding individuals under 18 years were considered "adults only". Studies that did not report participant age ranges were excluded from this subgroup categorization. Stratification by presence of symptoms was based on studies reporting inclusion of only participants with symptoms of potential SARS-CoV-2 infection for "symptomatic only". Studies that sampled both symptomatic and asymptomatic individuals, including those that sampled mostly symptomatic participants, were considered to sample both "symptomatic and asymptomatic" for this subgroup classification. Studies that did not report participant information regarding symptoms were excluded. Studies were categorized by participant setting if they reported sampling "outpatients only" or "inpatients only", and those that did not report such information were excluded. Studies that reported including HCWs as part of their sampled study population were categorized into a subgroup, and those that did not report presence of HCWs in their study were excluded. Stratification was also performed based on the presence of a laboratory-confirmed COVID-19 diagnosis in the study population sampled. Studies that reported including participants with lab-confirmed COVID-19 were included in this subgroup, and those who reported sampling only individuals with suspected infection and no known COVID-19 diagnosis were categorized as "no lab-confirmed COVID-19 patients included". Studies that did not detail

the COVID-19 status of participants upon enrolment to their study were excluded from these subgroups. Stratification by reference swab used for comparison was based on studies reporting their use of “flocked swabs”, swabs composed of nylon material, “nylon flocked swabs”, “nasopharyngeal”, “combined nasopharyngeal/oropharyngeal swabs”, or “combined nasal/oropharyngeal swabs”. For the purposes of these subgroup classifications, throat swabs were considered to be oropharyngeal swabs, and mid-turbinate and anterior nares swabs were both considered to be nasal swabs. Studies where information regarding the swab type used was unclear or missing were excluded. Stratification by type of liquid media used for gargle collection was performed based on studies reporting use of “saline” or “water”, and those that were missing information or used other liquids were excluded. Studies were stratified into three groups (“under 5 mL”, “5 mL-9 mL”, or “10 mL and greater”) based on their reported volume of liquid used for gargle collection. Those that used volumes that overlapped these discrete volume-based subgroups or did not report an exact volume (i.e. participant took a “few sips”) were excluded. Stratification by gargle collection time was also divided into three groups (“less than 15 seconds”, “15-29 seconds”, or “30 seconds or greater”) based on the gargle collection instructions reported by the study. Studies that used gargle collection times that overlapped these discrete time-based subgroups or did not report exact times in their gargle collection instructions (i.e. participant gargled “briefly”) were excluded. Studies were further stratified by reported length of time their participants were instructed to wait since anything by mouth prior to gargle sample collection, which included activities such as eating, drinking, smoking, and/or dental care. Studies were subdivided into three groups (“No wait”, “15-45 min wait”, or “60 min wait”) and those that did not report any information regarding a required wait time were excluded from the subgroups. Stratification was also performed based on whether study participants were

reported to have had any other paired samples collected from their mouth/throat consistently ahead of gargle collection during the study, which could have included saliva, chewing cotton pads, and/or self-collection or HCW-collection of swab samples. Studies that collected solely gargle samples from the participant's mouth/throat, or that alternated the collection order of mouth/throat samples paired with the gargle sample, were categorized as "no other samples collected". Studies that reported other mouth/throat samples being collected during the same sampling period as the gargle sample but did not specify collection order were excluded from these subgroups. Stratification was also conducted based on studies reporting use of "RT-PCR" (with extraction), "direct RT-PCR" (without extraction), LDTs, or commercial assays for COVID-19 testing of study samples. Studies where paired gargle and swab samples were tested on different assays were excluded from subgroup analyses. Where information was reported, studies were further subdivided based on their use of assays targeting one, two, or three different SARS-CoV-2 viral genes for COVID-19 testing, and those where details on viral genes targeted by the assay were unavailable were excluded. For the purposes of analysis, ORF1a/b was considered to be one viral gene target rather than two. Studies utilizing commercial assays were additionally stratified by whether the assay was publicly reported to be Food and Drug Administration (FDA) Emergency Use Authorized (EUA), approved by Health Canada (HC), and/or Conformité Européenne for In Vitro Diagnostic Medical Devices (CE-IVD) marked based on assay-specific searches of online manufacturer details and the Health Canada website. Studies where this approval information was missing or unclear were excluded from these subgroups.
To test the robustness of our estimates, sensitivity analyses were also performed based on study quality as recommended [112, 113]. We repeated the primary analysis including only higher

quality studies and excluded studies/comparison groups with "high" or "unclear" risk of bias in two or more of the four domains of the QUADAS-2 tool.

In accordance with recommendations, tau-squared (τ^2) values, I-squared (I^2) values, and forest plots were used to visualize heterogeneity of included studies. Heterogeneity in meta-analysis is associated with variation in the results of individual studies that are being pooled together to produce the summarized estimate. This variation may be attributed to sampling error in individual studies, to differences in the studies themselves including differences in populations included or other methodological differences, or reasons for variation in the results of studies may be unclear [112, 125]. Forest plots graphically depict the results of individual studies with their individual point estimates and confidence intervals, where overlapping confidence intervals and clustering of point estimates in the plot indicate similarity of study results, lower heterogeneity, and greater consistency across individual study results. τ^2 values are also reported for transparency. While more difficult to interpret, increasing, non-zero values of τ^2 typically indicate greater heterogeneity across study results. Additionally, I^2 values are also reported to quantify heterogeneity across studies, and I^2 is a percentage value with greater interpretability compared to other measures, such as τ^2. No heterogeneity is depicted by an I^2 value of 0%. Whereas, increasing I^2 values indicate increasing heterogeneity, with tentative interpretation of I^2 values of 25%, 50%, and 75% being considered as relatively low, moderate, and high amounts of heterogeneity [112, 113, 127, 132].

Additional planned analyses exploring the type of publication were no longer applicable as all included studies were peer reviewed publications. Planned meta-regression based on median age of participants and study design were not performed due to unavailability of data and all included studies being considered prospective cohort studies. While commonly used to assess

for presence of publication bias in a meta-analysis of randomized controlled trials, previous work has demonstrated that conventional funnel plots may be misleading when used for meta-analysis of studies of diagnostic test accuracy [133]. Therefore, we were unable to assess for publication bias statistically in these analyses.

Chapter 3: Results

3.1. Search and screening results

Searches run from 14 March 2022 to 5 May 2022 resulted in a total of 327 studies identified. The PRISMA flow diagram in Figure 2 outlines the review and screening process. 125 studies were identified from Ovid MEDLINE, 146 studies were identified from Ovid EMBASE, 42 studies were identified from Web of Science, and no studies were identified from Global Index Medicus. After searching preprint servers, nine studies were identified simultaneously from medrxivv and biorxiv, while three studies were identified from Research Square. One study was identified via a Google Scholar alert, and one study was identified via recommendation. 146 of these studies were removed as duplicates, leaving 181 remaining for initial title/abstract screening. 135 studies were deemed irrelevant to the review topic, leaving 46 studies for more rigorous full text screening. Of those 46 studies, 22 were excluded based on our pre-specified selection criteria and were primarily excluded for having the wrong study objective for the review (Appendix B). Additional reasons for exclusion included no gargling reported during sample collection (*n*=4), being the wrong type of article (*n*=3), or study participants being sampled greater than 10 days since diagnosis (*n*=3), for those studies including participants with a lab-confirmed COVID-19 diagnosis. Therefore, a total of 24 studies met criteria to be included in the review. These 24 studies included 32 gargle-swab comparison groups for analysis (Appendix F). We contacted study corresponding authors for additional information and to ask for clarification about their study methodology, and a total of 13 (61.9%) replied and provided the missing data. Three studies (four comparison groups) were missing information needed to calculate study estimates and information was unavailable from their authors. Thus, they were

excluded [72, 92, 96]. In the end, 28 of the 32 gargle-swab comparison groups (24 studies) had sufficient data for pooling in the meta-analysis.

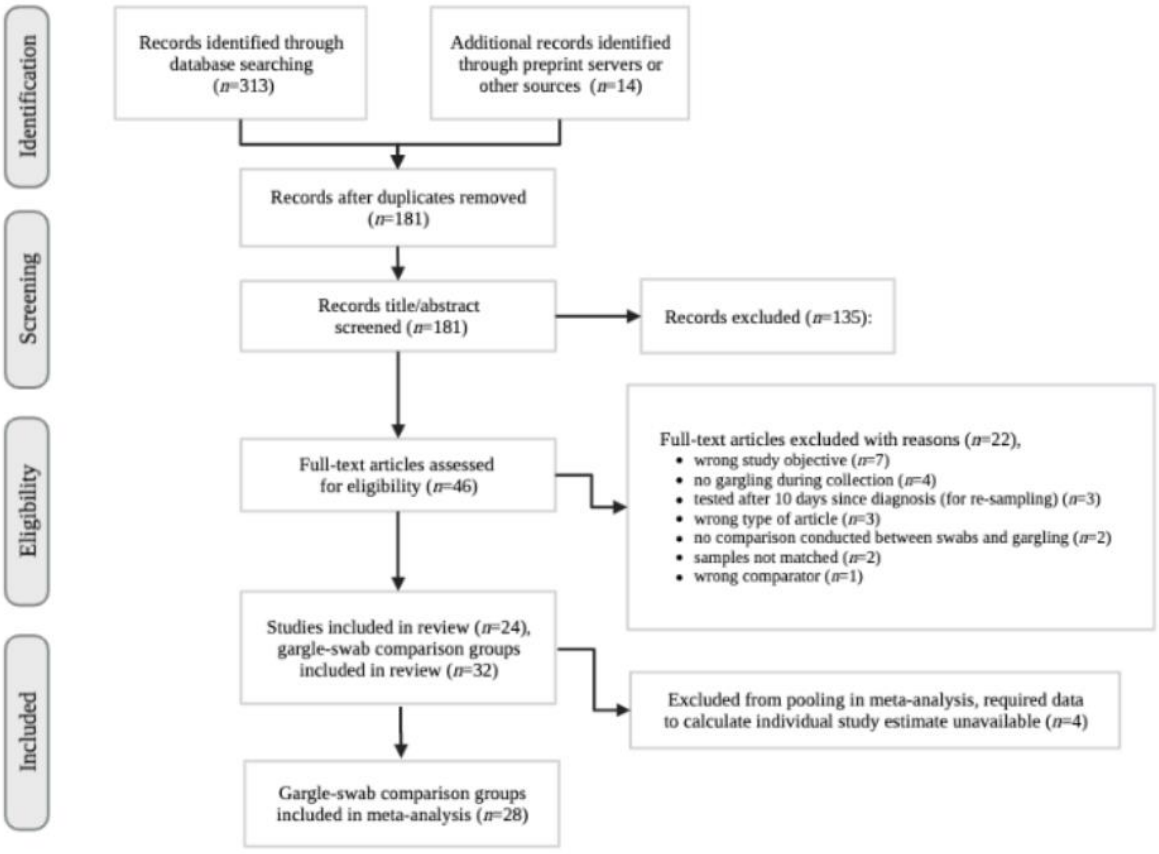

Figure 2. PRISMA flow diagram

3.2. Characteristics of included studies

The characteristics of all included studies (*n*=24)/comparison groups (*n*=32) are summarized in Appendix F. Our final set of studies included gargle-swab comparison groups from Europe (*n*=10), North America (*n*=11), South America (*n*=1), and Asia and the Middle East (*n*=10) with sample collection dates ranging March 2020 to July 2021, covering the original SARS-CoV-2 strain as well as the subsequent emergence of several novel variants. The study populations sampled varied and included adults, both adults and children (< 18 years old), both symptomatic and asymptomatic individuals, outpatients, inpatients, healthcare workers, individuals with a suspected infection and newly presenting for COVID-19 testing, as well as those with a previous lab-confirmed COVID-19 diagnosis who were resampled for study purposes. Various swab sampling methods were used for comparison to gargle sampling including combined nasopharyngeal/oropharyngeal swabs (*n*=10), combined nasal/oropharyngeal swabs (*n*=4), oropharyngeal (or throat) swabs (*n*=2), and nasopharyngeal swabs (*n*=12). Gargle collection methods also varied with use of saline, water, or premade gargling solution, varying volumes of liquid ranging from 2.5 mL to 20 mL, and varying collection time instructions ranging from 5-10 seconds to 30 seconds. Several studies enacted precautionary, required wait times since individuals last had something by mouth (i.e. eating, drinking, smoking, dental care) prior to gargle sample collection, which ranged from at least 15 minutes to 60 minutes. However, multiple studies did not require individuals to complete a waiting period and collected gargle samples irrespective of time since anything by mouth. RT-PCR was commonly used as the NAAT analytical method for SARS-CoV-2 detection (*n*=25), however, a few gargle-swab comparison groups were tested with modified, extraction-free, "direct" RT-PCR methods (*n*=3) instead. Numerous, different assays, including both laboratory-

developed (*n*=9) and commercial (*n*=18) assays, were utilized for COVID-19 testing in combination with several devices. These assays also targeted various combinations of several different SARS-CoV-2 viral genes.

3.3. Primary analysis results

After pooling across all 24 studies/28 comparison groups for which sufficient data were available, for a total of 1861 SARS-CoV-2 positive samples, the primary analysis found that overall gargle sample sensitivity was 92.7% (95% CI: 89.9% to 94.8%) with I^2 and τ^2 values of 31.7% and 3346 respectively (Figure 3). Gargle sampling was estimated to be 3.2% less (0.4% less to 6.0% less) sensitive than standard swab collection by HCWs for SARS-CoV-2 detection (Figure 4). Heterogeneity based on the I^2 and τ^2 values were 58.9% and 36 respectively.

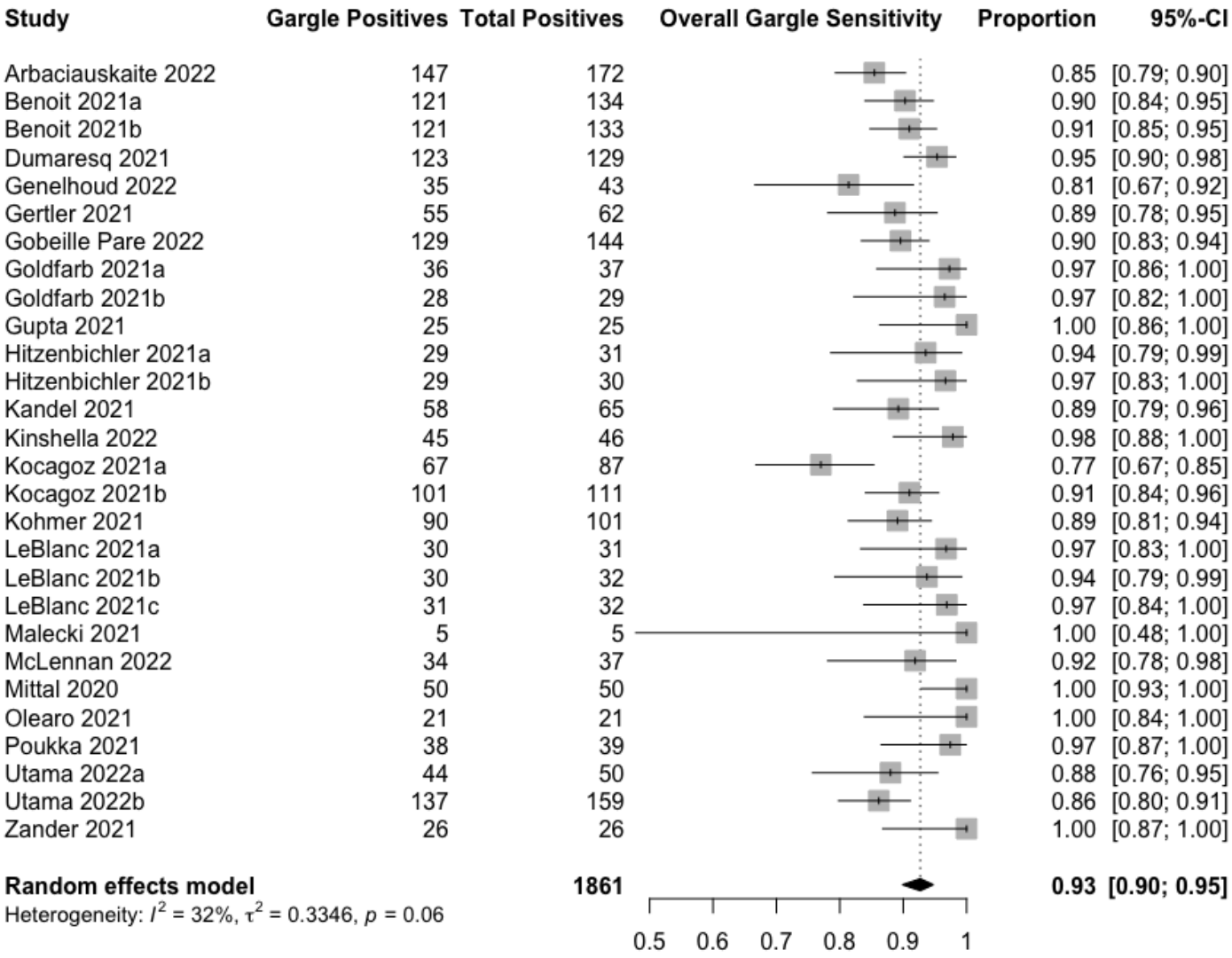

Figure 3. Estimate of overall gargle sample sensitivity across all included comparison groups (*n*=28)

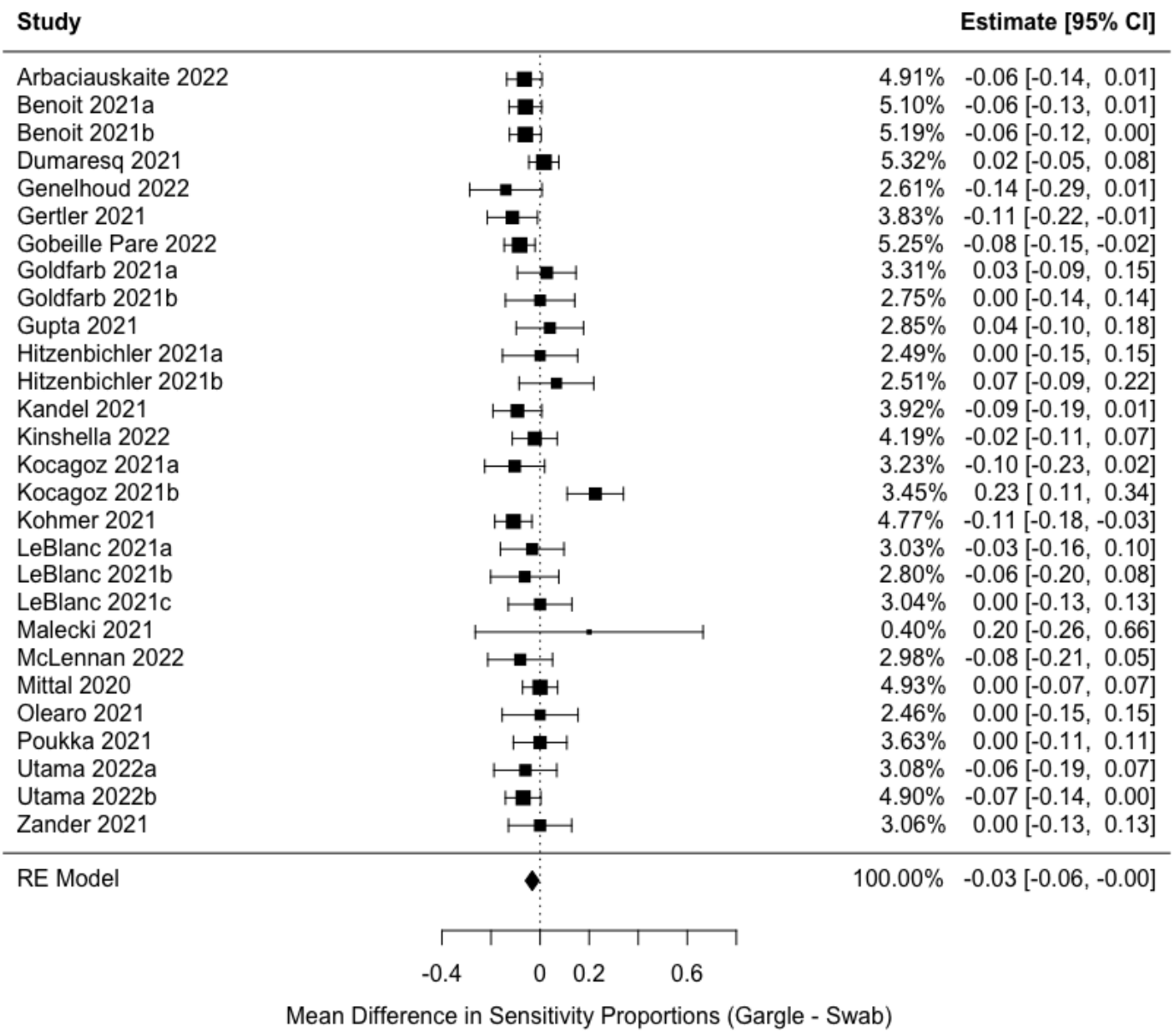

Figure 4. Estimate of mean sensitivity difference between gargle and swab sampling across all included comparison groups (*n*=28)

3.4. Subgroup analyses results

All estimates of overall gargle sensitivity, mean difference between gargle and swab sensitivity, and the associated heterogeneity values from numerous subgroup analyses are summarized in Table 1. Heterogeneity across subgroups ranged from "none" ($I^2 = 0.0\%$) to "high" ($I^2 > 75.0\%$). Gargle sensitivity was estimated to be greater than 87.0% across all subgroups stratified by diverse participant characteristics, settings, type or volume of gargle liquid, length of gargle sampling time, wait time prior to gargling, and reference swab type or NAAT analytical method used. Although point estimates of sensitivity varied slightly, confidence intervals were largely overlapping and indicated non-significant differences in gargle sensitivity across subgroups. In addition, gargle sample sensitivity was found to be no more than 8.0% lower than swab sensitivity across all subgroups, with some mean difference estimates found to be statistically significant non-zero differences between gargle and swab sensitivity.

When stratified by participant characteristics and study setting, gargle sensitivity estimates were similar between studies including adults only (92.1%; 87.4% to 95.1%) and studies including both adults and children (93.3%; 89.1% to 95.9%). Gargle sensitivity was also similar between studies sampling only symptomatic individuals (91.7%; 87.3% to 94.6%) and studies sampling both symptomatic and asymptomatic individuals (93.8%; 88.9% to 96.6%). Gargle sensitivity was comparable when sampling outpatients only (91.9%; 88.8% to 94.2%) or sampling inpatients only (94.3%; 81.8% to 98.4%). Studies that included HCWs as participants had an overall gargle sensitivity of 90.1% (86.4% to 92.9%). Studies that included individuals with a prior lab-confirmed COVID-19 diagnosis (≤ 10 days) had a gargle sensitivity of 95.4% (91.4% to 97.6%), and studies that included only participants with suspected SARS-CoV-2

infection who were newly presenting for testing estimated a gargle sensitivity of 89.9% (86.3% to 92.7%).

As sampling site varied and quality of the reference swab used in the study increased, overall gargle sensitivity point estimates were observed to decrease slightly when gargle sample collection was compared to combined nasal/oropharyngeal (96.7%; 85.2% to 99.3%) versus nasopharyngeal (93.4%; 87.3% to 96.7%) versus combined nasopharyngeal/oropharyngeal (90.4%; 87.0% to 93.0%) swabs. However, these slight differences were non-significant as confidence intervals overlapped across all three groups. Notably, sensitivity of gargle sampling was 90.4% (87.0% to 93.0%) and was 5.2% less (2.4% less to 8.0% less) sensitive, even when compared to SARS-CoV-2 detection by high quality combined nasopharyngeal/oropharyngeal swabs. In studies that used high quality nylon flocked swabs for comparison, gargle sensitivity was 95.7% (90.3% to 98.2%).

The greatest point estimate sensitivities were observed when gargle sampling for 30 seconds or greater (96.2%; 91.7% to 98.4%), using 5-9 mL (94.5%; 90.4% to 96.9%) of saline (95.6%; 90.6% to 98.0%). While the point estimate for gargle sensitivity was greatest when enacting a 60-minute wait time prior to gargle collection (96.6%; 91.4% to 98.7%), performance was comparable to subgroups using shorter wait times. Notably, gargle sensitivity was 92.9% (86.9% to 96.2%) in studies where there was no required wait time since the participant last had something by mouth, with gargle sampling 3.5% less (7.9% less to 0.9% more) sensitive than standard swab collection in this context. Gargle sensitivity in studies where participants had other samples collected from the mouth/throat consistently ahead of gargle collection (91.1%; 88.3% to 93.4%), was comparable to studies where solely gargle samples were collected (93.6%; 86.8% to 97.0%). Although point estimates varied slightly, gargle sensitivity estimates were

comparable overall, with overlapping confidence intervals, across the varied gargle sample collection methods used across included studies.

When stratified by analytical method used for COVID-19 testing in the study, gargle sensitivity remained comparable across groups. The point estimate of gargle sensitivity was slightly higher in studies using RT-PCR for testing at 92.9% (89.8% to 95.1%) compared to studies using direct RT-PCR methods, without prior extraction, at 89.7% (56.6% to 98.3%). However, confidence intervals overlapped representing a non-significant difference, and there were a lower number of studies/comparison groups included in the direct RT-PCR subgroup warranting caution in the interpretation. Sensitivity of gargle sampling was similar across studies employing numerous, different assays that targeted either one, two, or three viral genes for SARS-CoV-2 detection. This comparable performance of gargle samples was similarly observed across studies utilizing LDTs or commercial assays, including commercial assays which were deemed to be FDA EUA, HC approved, and/or CE-IVD marked.

3.5. Quality assessment and sensitivity analysis results

Quality assessment found that all included studies were at “low” concern regarding applicability for this review. Several studies were deemed to be at risk of bias due to “high” or “unclear” assessments in the QUADAS-2 domains of “patient selection”, “index testing”, and “reference testing” (Figure 5). These particular studies reported including individuals with prior lab-confirmed COVID-19, rather than solely individuals with suspected infection. Further, these studies did not report blinding of index and reference test results during study sample testing, although this may be less of a concern as testing was performed with an objective molecular test (Figure 6). A total of four studies/five comparison groups were included in sensitivity analysis

based on their having overall “low” risk of bias assessments and/or less than two domains with a “high” or “unclear (some concerns)” risk of bias assessment [45, 58, 59, 97]. After repeating the primary analysis with only these high-quality studies/comparison groups and across 346 SARS-CoV-2 positive samples total, we estimated that overall gargle sample sensitivity was 91.6% (78.7% to 97.0%) with I^2 and τ^2 values of 0.0% and 2171 respectively. Furthermore, gargle sampling was estimated to be 5.3% less (1.0% less to 9.6% less) sensitive than swab collection. Heterogeneity based on the I^2 and τ^2 values were 12.7% and 4 respectively.

Table 1. Subgroup analyses based on study characteristics of interest

Study Characteristics	Comparison groups/studies (*n*)	Gargle sensitivity			Difference in sensitivity (Gargle – Swab)		
		Estimate [95% CI] (%)	I^2 (%)	τ^2	Estimate [95% CI] (%)	I^2 (%)	τ^2
All studies/comparison groups	28	92.7 [89.9 to 94.8]	31.7	3346	-3.2 [-6.0 to -0.4]*	58.9	36
Participants and Setting							
Adults only	17	92.1 [87.4 to 95.1]	23.3	3995	-2.8 [-7.4 to 1.9]	65.6	57
Adults and children	11	93.3 [89.1 to 95.9]	29.5	1810	-3.6 [-6.5 to -0.7]*	30.3	9
Symptomatic only	13	91.7 [87.3 to 94.6]	53.8	2500	-2.9 [-8.0 to 2.2]	65.9	50
Symptomatic and asymptomatic	15	93.8 [88.9 to 96.6]	0.0	4777	-3.5 [-6.8 to -0.1]*	50.9	26
Outpatients only	17	91.9 [88.8 to 94.2]	11.3	1129	-5.0 [-7.7 to -2.4]**	47.9	20
Inpatients only	7	94.3 [81.8 to 98.4]	49.0	10553	2.4 [-7.5 to 12.3]	68.9	81
Healthcare workers included	7	90.1 [86.4 to 92.9]	0.0	0	-6.6 [-11.3 to -2.0]*	55.6	30
No lab-confirmed COVID-19 patients included	14	89.9 [86.3 to 92.7]	49.3	1761	-3.9 [-9.0 to 1.3]	76.5	64
Lab-confirmed COVID-19 patients included	14	95.4 [91.4 to 97.6]	0.0	3397	-2.6 [-5.8 to 0.6]	25.9	12
Swab Characteristics							
Flocked swabs used	15	94.5 [90.7 to 96.8]	0.0	2429	-2.9 [-5.7 to -0.2]*	36.4	12
Nylon swabs used	12	95.9 [91.1 to 98.1]	0.0	4977	-2.6 [-6.0 to 0.9]	33.4	14
Nylon flocked swabs used	11	95.7 [90.3 to 98.2]	0.0	5136	-2.8 [-6.6 to 1.0]	36.2	16
Nasopharyngeal/oropharyngeal swabs used	10	90.4 [87.0 to 93.0]	6.1	790	-5.2 [-8.0 to -2.4]**	31.7	7
Nasopharyngeal swabs used	12	93.4 [87.3 to 96.7]	52.3	6294	-1.9 [-8.0 to 4.2]	64.8	63
Nasal/oropharyngeal swabs used	4	96.7 [85.2 to 99.3]	0.0	0	-1.4 [-8.2 to 5.5]	8.7	4
Gargle Characteristics							
Saline used	14	95.6 [90.6 to 98.0]	16.4	7798	-3.1 [-6.1 to -0.1]*	35.3	18
Water used	10	90.8 [86.7 to 93.7]	56.1	1697	-3.3 [-10.2 to 3.7]	81.3	77
Under 5 mL liquid	3	87.2 [75.8 to 93.7]	0.0	0	-7.4 [-11.0 to -3.8]*	0.7	0
5-9 mL liquid	14	94.5 [90.4 to 96.9]	14.4	3131	-3.2 [-5.8 to -0.6]*	28.3	9
10 mL or greater liquid	7	88.8 [84.6 to 92.0]	0.0	0	-5.5 [-12.5 to 1.4]	58.6	47
Less than 15 seconds gargle collection	6	87.2 [78.6 to 92.7]	58.3	1599	-0.2 [-14.4 to 13.9]	79.4	141
15-29 seconds gargle collection	8	92.4 [88.4 to 95.0]	0.0	461	-4.5 [-8.7 to -0.3]*	51.5	15
30 seconds or greater gargle collection	9	96.2 [91.7 to 98.4]	0.0	2043	-2.4 [-5.5 to 0.6]	8.6	4

Study Characteristics	Comparison groups/studies (*n*)	Gargle sensitivity			Difference in sensitivity (Gargle – Swab)		
		Estimate [95% CI] (%)	I^2 (%)	τ^2	Estimate [95% CI] (%)	I^2 (%)	τ^2
Gargle Characteristics Continued							
No wait time prior to gargle collection	10	92.9 [86.9 to 96.2]	36.1	3582	-3.5 [-7.9 to 0.9]	42.6	21
15-45 min wait prior to gargle collection	6	89.2 [85.7 to 91.9]	0.0	0	-6.9 [-8.0 to -5.7]***	0.5	0
60 min wait prior to gargle collection	6	96.6 [91.4 to 98.7]	0.0	0	-1.4 [-4.5 to 1.6]	3.7	1
No other samples collected prior to gargle	11	93.6 [86.8 to 97.0]	51.4	6613	-1.1 [-7.1 to 4.9]	62.9	54
Yes other samples collected prior to gargle	15	91.1 [88.3 to 93.4]	15.8	680	-5.1 [-8.1 to -2.2]**	42.6	16
Analytical Methods							
RT-PCR used	25	92.9 [89.8 to 95.1]	0.0	2564	-4.5 [-6.6 to -2.4]***	38.8	16
Direct RT-PCR used	3	89.7 [56.6 to 98.3]	87.6	4744	4.6 [-36.1 to 45.2]	90.4	232
1 viral gene targeted by assay	6	95.2 [85.8 to 98.5]	0.0	2471	-2.3 [-7.9 to 3.4]	22.7	10
2 viral genes targeted by assay	16	92.3 [88.4 to 95.0]	0.0	1927	-5.2 [-7.8 to -2.7]***	39.7	17
3 viral genes targeted by assay	3	93.8 [52.6 to 99.5]	71.6	5730	-1.1 [-14.1 to 11.9]	45.1	14
Laboratory developed test (LDT) assay used	9	95.6 [90.2 to 98.1]	3.9	3798	-2.8 [-6.8 to 1.3]	34.6	14
Commercial assay used	18	91.1 [87.4 to 93.8]	36.6	2400	-3.1 [-7.3 to 1.0]	66.4	50
FDA EUA commercial assay used	8	92.9 [85.1 to 96.7]	0.0	2045	-4.2 [-7.2 to -1.3]*	16.2	5
Health Canada approved commercial assay used	8	92.6 [87.3 to 95.8]	37.0	1576	-3.5 [-6.5 to -0.4]*	18.5	4
CE-IVD marked commercial assay used	8	92.1 [89.0 to 94.3]	0.0	0	-4.8 [-9.3 to -0.4]*	53.5	24

Non-zero significant difference, *p*-value < ***0.001; **0.01; *0.05
Heterogeneity = I^2, τ^2
I^2 = 25% = low heterogeneity; I^2 = 50% = moderate heterogeneity; I^2 = 75% = high heterogeneity; see Methods for additional details regarding interpretation

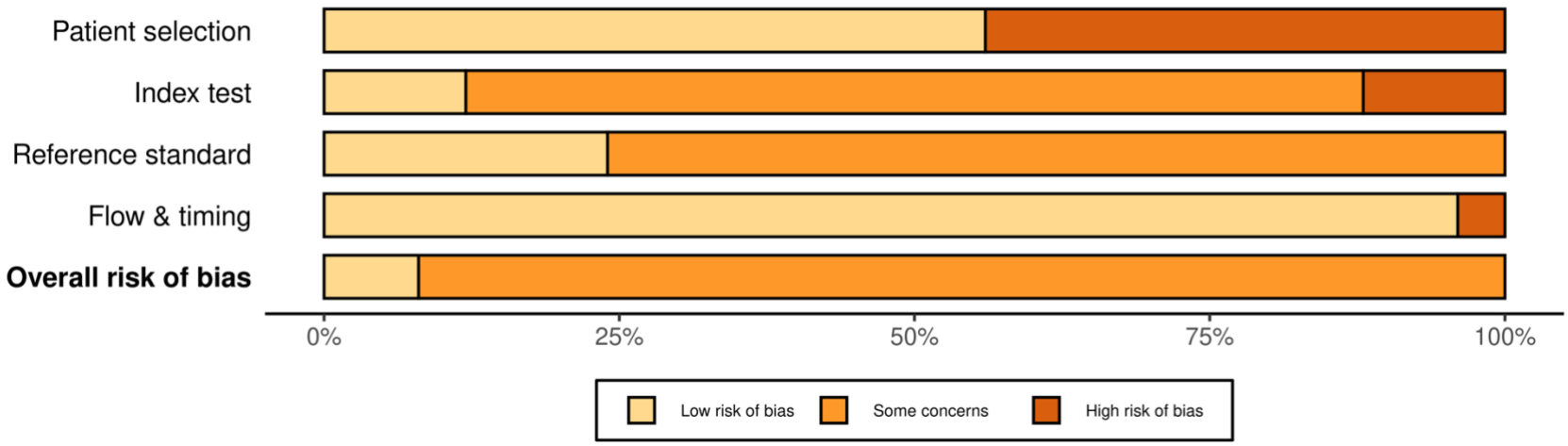

Figure 5. Study quality assessment summary bar plot across all included comparison groups (*n*=32)

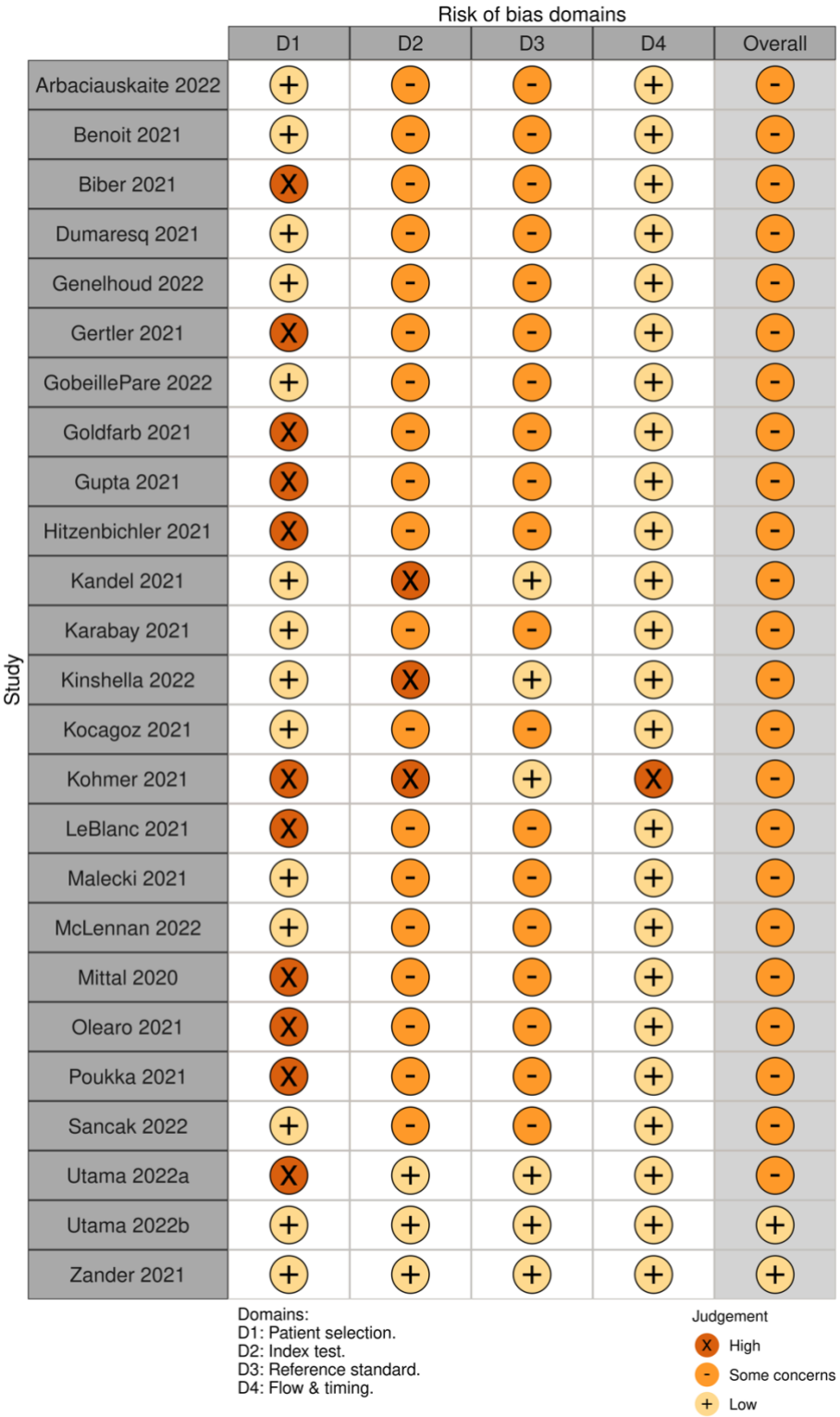

Figure 6. Study quality assessment traffic light plot across all included comparison groups (*n*=32)

Chapter 4: Discussion

4.1. Summary of findings

After systematically reviewing the literature and pooling data from 24 studies, which included 28 gargle-swab comparison groups, in a meta-analysis, we found a relatively high overall sensitivity for gargle sampling when compared to standard swab collection for SARS-CoV-2 detection with NAAT. We focused on studies employing NAAT methods only, as it is the gold-standard diagnostic method for respiratory viral testing, and we excluded RADTs or other testing methods. We also performed comprehensive stratified analyses to explore a near exhaustive list of factors that may potentially impact SARS-CoV-2 detection performance of a novel sampling method including the gargle collection method, type of swab comparator, NAAT assay used, as well as population and setting characteristics. These analyses provide a closer look at gargle sample performance across the variety of testing approaches implemented globally, as well as provide a framework to inform optimization of gargle sample collection to achieve the highest sensitivity for SARS-CoV-2 detection. At 92.7% gargle sensitivity overall, and greater than 87.0% sensitivity across all subgroups explored, its performance considerably exceeds the performance seen with rapid antigen-based tests [134]. Further, we found that gargle self-collection was only a few percentage points less sensitive than HCW-collected swabs, which has not been previously estimated. Although this was estimated to be a statistically significant non-zero difference, we do not believe it represents a meaningful performance difference from a clinical or practical perspective. Firstly, it is generally accepted that *p* values, which are estimated to be significant by statistical analysis, should be interpreted with caution and do not necessarily equate to differences that have biological or clinical significance. Instead, it is important to critically assess the statistical significance in combination with other available

information when making interpretations about findings and their clinical implications [135]. Furthermore, clinically used tests commonly vary in their diagnostic performance to some degree. Previous research has demonstrated that combined NPOP swabs yielded a slightly greater detection of respiratory viruses (influenza A and B, HMPV, RSV) when compared to detection by NP swabs alone [136]. Nevertheless, both these methods are accepted for clinical use, vary in their usage based on context, region, and clinical guidance, yet they are deemed to be comparable in performance for respiratory viral detection overall [41, 56, 58, 89]. With gargle sampling yielding an overall sensitivity of greater than 90% in our analysis, the estimated loss of 3% sensitivity is not expected to result in impaired diagnostic performance or significantly more missed SARS-CoV-2 diagnoses. For these reasons, this small, yet statistically significant difference in sensitivity would likely not preclude implementation of this sample type as an option for molecular COVID-19 testing in certain real-world contexts. In addition, after stratifying our results by only the highest quality studies, we found that overall gargle sensitivity and the mean difference in sensitivity did not change significantly, which further supports the robustness of our estimates. Overall, our findings support the reliability and non-inferiority of gargle sampling as an alternative to swabs collected by HCWs for COVID-19 testing.

4.2. Interpretation and implications

Our result (92.7%) is also comparable to the 91% overall gargle sensitivity estimated by an earlier gargle systematic review and meta-analysis [111]. Notably, we observed comparable gargle performance in studies sampling adults only and studies sampling both adults and children as young as 4 years old (92.1% vs 93.3%), which aligns with findings from the prior gargle meta-analysis that similarly found comparable performance in studies with and without children

[111]. While we were unable to explore gargle performance in children alone due to unavailability of data, the lack of a significant performance difference between groups in our analysis supports the use of gargle collection for those under 18 years old, so long as they are capable and comfortable following gargle collection instructions. Further, we found that gargle sensitivity was comparable across studies sampling symptomatic only or symptomatic and asymptomatic individuals, which corroborates findings from the previous gargle meta-analysis that also stratified by presence of symptoms [111]. We are the first to stratify gargle performance by inpatient only versus outpatient only populations and found comparable sensitivities for SARS-CoV-2 detection when sampling these groups. These findings support gargle sampling as an appropriate testing approach for both admitted patients and those in the community with and without symptoms. Interestingly, we found gargle sensitivity in studies that reported inclusion of HCWs as participants to be 90.1%, which did not differ significantly from the gargle sensitivities also estimated in other subgroups. This finding may dispel assumptions that validation studies sampling HCWs may overestimate the sensitivity of gargle sampling due to HCWs being potentially better at self-collecting a gargle sample than the general population. However, we could not directly explore sensitivity in studies that included only non-HCWs due to a lack of available data, and future research may be needed to directly compare gargle sample performance in HCWs versus non-HCWs to investigate this. Furthermore, the relatively high gargle sensitivity and low percentage difference in sensitivity compared to HCW-collected swabs estimated in studies sampling only individuals with suspected infection supports gargle collection as a viable option for COVID-19 diagnostic purposes.

Given that different studies utilized different swab comparators as their gold-standard, we found that gargle sensitivity point estimates did vary slightly, although non-significantly, when

varying the swabs used for comparison. These trends may be a result of slight performance differences between swab methods, rather than gargle sampling itself, with swabbing of certain anatomic sites potentially yielding a slightly higher sensitivity and quality of sample than others [29, 43, 44]. The NPOP swab is considered to be one of the highest quality swab methods used for respiratory viral diagnostics, potentially yielding higher estimates of sensitivity than NP or OP swabs alone [46]. However, it also requires the services of a trained HCW, with little to no potential for sufficient self-collection, and is an invasive sample to collect for patients, particularly young children [77]. Notably, gargle performance was greater than 90.0% across swab comparators used, even when compared to high-quality NPOP swabs. This finding supports gargle sampling as a reliable alternative to upper respiratory tract swab collections that may be highly invasive or challenging, including collection of the NPOP swab. These findings also align with the prior gargle meta-analysis that similarly reported slightly differing, yet non-inferior gargle sensitivities, across NP and NPOP swabs [111]. Furthermore, the type of material that the swab is made of may also impact the quality of sample yielded for laboratory testing and comparison. Nylon flocked swabs are composed of nylon fibers which are uniquely arranged to allow for efficient uptake and elution of collected sample, producing a specially designed swab known to yield respiratory samples of high quality [22, 23]. We are the first to explore gargle performance compared to this type of swab material, and we found that gargle sampling also performed comparably for SARS-CoV-2 detection in studies utilizing nylon flocked swabs.

Gargle sampling with saline, rather than water, demonstrated a slightly higher yet comparable sensitivity of 95.6%, which corroborates findings from a prior gargle meta-analysis [111]. Similar sensitivity was still observed when gargling with water, which may be useful for testing in settings where medical grade saline is not readily available. Performance was

comparable when gargle sampling with differing volumes of liquid, although there was a trend toward greater sensitivity when using 5-9 mL of liquid in our analyses. Although they were only able to stratify by two groups, a prior gargle meta-analysis found near identical estimates to ours when using under 5 mL (87% vs 87.2%) and 5 mL or greater (92% vs 94.5%) of liquid for gargling [111]. Upon stratifying by three groups in our analysis, we observed the highest sensitivity point estimate was when gargling with an intermediate amount of liquid (5-9 mL), although this was non-significant and comparable sensitivities were observed when gargling with other volumes. Our data supports the possibility that sensitivity may begin to decrease as gargle volume increases to a limit, potentially due to dilution of viral material during sampling. When considering practical implementation of this method, use of an intermediate volume of liquid may allow enough liquid to effectively perform gargling, while not being too much liquid such that dilution effects begin to negatively impact sensitivity. However, additional research is needed to explore this further and improve guidance regarding optimal gargle sample volume. We also observed a trend where longer gargle sampling time, at 30 seconds or greater, produced the highest sensitivity point estimate, which aligns with a prior gargle meta-analysis which found higher sensitivity when using over 10 seconds [111]. While comparable performance was observed with other times, a longer sampling time may result in capturing a greater amount of potential viral material from the mouth/throat. While this recommendation is slightly longer than the 10 second collection of a swab sample, given the non-invasive nature of the gargle sample, we do not expect that this will be an overly burdensome amount of sample collection time for the patient. We also observed that gargle testing guidance varied across studies, with several imposing a required wait time since the individual last had anything by mouth including drinking, eating, chewing gum, smoking, or dental care. In contrast, multiple studies did not

require this and sampled all individuals irrespective of a pre-specified waiting period. We are the first to explore the comparative performance of gargle sampling as it relates to precautionary wait times. We observed that gargle sensitivity estimates were comparable across different wait times, with a trend toward greater sensitivity in studies employing a 60-minute wait. However, depending on the testing setting and patient population, this may not always be a feasible request. Interestingly, we observed comparable gargle sensitivity in studies where there was no required wait time. These findings support the robustness of gargle sample performance and demonstrate that lengthy wait times, implemented out of precaution, may potentially be shortened, or removed without sacrificing test sensitivity. However, practical judgement is still needed to assess appropriateness of the collection method and review submitted samples for quality to ensure that they are not overtly contaminated with something the patient may have consumed. Further, in situations where there is concern regarding gargle sample quality or potential contamination, a confirmatory HCW-collected swab may still be collected from a patient with high clinical suspicion of SARS-CoV-2 infection who initially tested negative on a gargle sample [89]. Overall, this finding is particularly useful for practical implementation in point-of-care settings, such as emergency departments or drop-in clinics, where ill patients may be encouraged to take oral fluids regularly and a lengthy required wait time for gargle testing, beyond a few minutes, may be a significant barrier to uptake of this method.

Studies utilized various approaches for COVID-19 testing related to guidance in their respective region, and we observed non-inferior gargle sensitivity across the variety of testing approaches occurring in various laboratories around the world. Gargle sensitivity was observed to be comparable to swabs in studies utilizing RT-PCR, which is the gold-standard method and commonly employed for clinical testing for respiratory viruses, and this aligns with findings

from a prior gargle meta-analysis [111]. Gargle samples also performed well when extraction-free, direct RT-PCR was used for COVID-19 testing. We are the first to estimate gargle sensitivity as it relates to this newer testing method, which may allow for higher-volume, higher throughput testing than standard RT-PCR in unique testing situations where this may be warranted. However, given the small sample size in the subgroup, results should be interpreted with caution and additional research is needed to further understand performance of gargle sample collection in combination with direct RT-PCR testing. We observed no performance differences across studies employing assays targeting one, two, or three SARS-CoV-2 viral genes for detection. Furthermore, gargle performance was comparable to swabs when tested with a laboratory developed test or a commercial assay, including conventionally used commercial assays that are FDA Emergency Use Authorized, Health Canada approved, or CE-IVD marked for COVID-19 testing.

When compared to other non-invasive, swab-independent, self-collect sample types that have been deemed to be acceptable for COVID-19 testing such as saliva, gargle sampling appears to have similarly acceptable performance. A recent systematic review and meta-analysis estimated saliva sensitivity to be 3.4% lower (9.9% lower to 3.1% higher) than NP swabs, while we estimated gargle sampling to be 1.9% lower (8.0% lower to 4.2% higher) than NP swabs in our work [82]. Interestingly, saliva sensitivity was estimated to be 86.9% overall (82.3% to 90.4%), which is comparable yet slightly lower than our point estimate of overall gargle sensitivity being 93.4% (87.3% to 96.7%) when compared to NP swabs [82]. While gargle sampling exhibits similar performance to saliva sampling for COVID-19 diagnosis, it also maintains several additional benefits such as being a more homogenous, less viscous sample type that may be easily introduced into the COVID-19 lab testing workflow. In contrast, saliva

samples may be challenging for patients to collect a sufficient volume, and the highly viscous nature of saliva often requires additional sample handling steps and reagents to prepare the sample for RT-PCR testing that would not necessarily be required for processing of gargle samples [41, 74, 81]. Given its non-inferior performance when processed with the many varying approaches that are commonly used for COVID-19 testing of swabs in several regions, our findings demonstrate that gargle samples are a versatile, highly amenable, alternative sample type for SARS-CoV-2 molecular detection in the clinical lab.

4.3. Strengths, limitations, and recommendations for future research

Our work has several strengths including our consultation of the appropriate guidelines in the design and execution of this systematic review and meta-analysis. We searched multiple platforms for relevant studies and our search strategy was reviewed by an experienced librarian to aid in locating all available evidence related to the topic and minimize chances of inappropriately excluding studies. Our study selection, data extraction, and quality assessment were all conducted in duplicate by two reviewers as recommended to minimize bias and improve reliability. We included studies from around the world that performed this gargle to swab comparison using a variety of liquids, volumes, comparator swabs, collection methods, testing assays, and sampled varying patient populations including outpatients, inpatients, symptomatic, asymptomatic, adults, and children, over several months of the COVID-19 pandemic to prioritize generalizability and understand the performance of this novel sample type across many potential real-world contexts. In addition to performing the primary analysis, we also assessed studies for quality and stratified our estimates by only the highest quality studies to further test the robustness of our results. Our work builds on a previous gargle systematic review and meta-

analysis with inclusion of additional studies and greater data allowing for further stratified analyses of gargle performance across various sampled populations, swab and gargle collection methods, and testing approaches. We add to the body of knowledge surrounding ideal liquid medium, volume, and duration of gargle sampling. We also offer new insights to support practical implementation of gargle sampling by providing data surrounding use of a required wait time and gargle performance across various COVID-19 testing assays that are routinely used in clinical labs.

Our work also has several limitations. First, we used an imperfect reference standard with the swab collection method, and we were not able to estimate specificity of gargle sampling because we considered any positive result to be a true positive. However, this method is commonly used in contexts similar to ours. We included studies with known heterogeneity that varied in the gargle collection, swab collection, and testing methods they used; nonetheless, this reflects the real-world differences in testing approaches employed across regions and individual labs. We attempted to explore these variabilities and their impacts on gargle performance through our numerous subgroup analyses. Minimal data was available regarding time from collection to testing in individual studies, which may impact sample quality and sample performance, and we were unable to explore this in our analyses. Although, clinical swab samples were often transported in stabilizing media and untreated gargle samples have been demonstrated to be stable at room temperature for at least two days, so this may not be a significant concern [41]. According to the QUADAS-2 quality assessment, included studies were considered to be at risk of bias due to lack of information or unclear information regarding the blinding of paired sample results. Given that testing is performed through an objective molecular test with pre-specified or manufacturer specified cut-offs for result interpretation, rather than

subjective decision-making, this may be less of a concern. In addition, we performed sensitivity analysis and found that our estimates did not change significantly when pooling only high-quality studies. Several studies included individuals with confirmed SARS-CoV-2 infections, which may introduce selection bias. However, we were able to stratify and explore gargle performance in studies with and without lab confirmed COVID-19 patients in our subgroup analyses and found comparable results. We were unable to use funnel plots for diagnostic accuracy studies to statistically estimate the effect of publication bias in our review. We attempted to minimize risk of missing relevant studies through several rounds of searches of multiple databases and preprint servers, although the impact of unpublished studies with null or negative results on our findings remains unclear. Included studies were performed in high- and middle-income countries, with no representation of data from resource-limited settings. While the reported benefits of gargle collection may be especially useful in resource-limited settings, we cannot fully generalize our results and additional research specific to gargle sampling in this setting is warranted prior to its implementation. In addition, sample collection and testing across included studies were conducted prior to the emergence of Omicron and other SARS-CoV-2 variants and prior to widespread COVID-19 vaccination, which may alter the biology of SARS-CoV-2 infection and disease. Thus, our findings should be interpreted with caution and may not be fully applicable for detection of newly emerging variants or diagnosis in vaccinated individuals. Future research exploring performance of gargle sampling in adults versus children of various ages, in HCWs versus non-HCWs, in COVID-19 vaccinated individuals, in resource-limited settings, or in combination with direct RT-PCR testing are all warranted to further understand the performance of this novel sampling method for SARS-CoV-2 detection. Ongoing research and follow-up systematic reviews and meta-analyses, that include updated studies, are

also encouraged to explore gargle performance in these evolving circumstances and to continue optimization of this method.

Chapter 5. Conclusion

Overall, results from our work support the use of gargle sampling for SARS-CoV-2 detection as a comparable alternative sample type to high quality swabs collected by HCWs. Gargle sampling is a less invasive, reliable, swab-independent, self-collect method that performs comparably to swabs when tested with already existing laboratory methods and devices, allowing this sample type to be readily implemented in the clinical laboratory workflow with minimal disruption. The stable performance of gargle samples across various testing contexts and populations supports its use for COVID-19 testing in many settings. Given that it is more acceptable to patients, it may be useful for situations where frequent, repeated testing is warranted, such as travel or workplace COVID-19 screening, or for testing those that have difficulty with invasive swab collection, such as children. Its high acceptability may also help overcome testing fatigue or testing aversion associated with upper respiratory tract swab collection. Given its room temperature stability and ability for swab-independent self-collection, there is potential for individuals to collect their gargle sample outside clinical settings and subsequently "mail in" or "drop off" the sample to a centralized lab testing facility. This may be especially beneficial in remote or resource-limited settings where laboratory services or trained HCWs are difficult to access for standard swab testing. While gargle samples were observed to have slightly lower sensitivity than swabs by a few percentage points, it does not represent a practically significant difference given its overall sensitivity being over 90%. There are currently no collection methods that are 100% sensitive, with HCW-collected swabs being reported to have missed infections, particularly if challenges arise during swabbing that lead to a lower quality swab sample being collected [46, 56]. Missed infections can also occur if individuals are not tested to begin with, particularly if faced with the existing barriers of swab collection such as

limited access to trained HCWs for swabbing, unavailability of swabs, and overall invasiveness and discomfort leading to testing aversion. Thus, implementation of gargle sampling may offer a solution to these barriers and significantly improve testing access for populations where standard swab collection is not feasible.

In conclusion, our findings recommend this sample type as an alternative to HCW-collected swabs for COVID-19 testing across various contexts, from inpatient to community-based or at-home settings. Gargle samples offer comparable sensitivity for SARS-CoV-2 detection that give it potential as an option for COVID-19 testing, particularly in situations when swab collection by a HCW may not be possible. Widespread implementation of gargle sampling would also support ongoing testing efforts and genomic surveillance for SARS-CoV-2 across diverse settings. Finally, given its usefulness for COVID-19 testing and amenability for lab processing, future research is warranted to evaluate the potential benefits of gargle sampling for other commonly circulating respiratory viruses, such as influenza A, influenza B, RSV, and rhinovirus. Continued innovation with gargle sampling and other alternative sampling methods, particularly as it relates to self-collection, is greatly encouraged, and has potential to strengthen respiratory viral diagnostics and efficiency of disease management going forward.

References

1. Yang Y, Xiao Z, Ye K, He X, Sun B, Qin Z, Yu J, Yao J, Wu Q, Bao Z, Zhao W. SARS-CoV-2: characteristics and current advances in research. Virology journal. 2020 Dec;17(1):1-7.

2. Hu B, Guo H, Zhou P, Shi ZL. Characteristics of SARS-CoV-2 and COVID-19. Nature Reviews Microbiology. 2021 Mar;19(3):141-54.

3. Harrison AG, Lin T, Wang P. Mechanisms of SARS-CoV-2 transmission and pathogenesis. Trends in immunology. 2020 Oct 14.

4. Llanes A, Restrepo CM, Caballero Z, Rajeev S, Kennedy MA, Lleonart R. Betacoronavirus genomes: how genomic information has been used to deal with past outbreaks and the COVID-19 pandemic. International journal of molecular sciences. 2020 Jun 26;21(12):4546.

5. Aftab SO, Ghouri MZ, Masood MU, Haider Z, Khan Z, Ahmad A, Munawar N. Analysis of SARS-CoV-2 RNA-dependent RNA polymerase as a potential therapeutic drug target using a computational approach. Journal of translational medicine. 2020 Dec;18(1):1-5.

6. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, Niemeyer D, Jones TC, Vollmar P, Rothe C, Hoelscher M. Virological assessment of hospitalized patients with COVID-2019. Nature. 2020 May 28;581(7809):465-9.

7. Cao Q, Chen YC, Chen CL, Chiu CH. SARS-CoV-2 infection in children: Transmission dynamics and clinical characteristics. Journal of the Formosan Medical Association. 2020 Mar;119(3):670.

8. Chao JY, Derespina KR, Herold BC, Goldman DL, Aldrich M, Weingarten J, Ushay HM, Cabana MD, Medar SS. Clinical characteristics and outcomes of hospitalized and critically ill children and adolescents with coronavirus disease 2019 at a tertiary care medical center in New York City. The Journal of pediatrics. 2020 Aug 1;223:14-9.

9. Ganguly M, Giri PP, Basu S. A single-center experience from Eastern India depicting the epidemiology and phenotypic variations of Multisystem inflammatory syndrome in children (MISC) associated with SARS-CoV2 seen after first wave and second wave of COVID 19. Journal of Pediatric Critical Care. 2023 Mar 1;10(2):49-55.

10. Centers for Disease Control and Prevention [Internet]. [place unknown]: Centers for Disease Control and Prevention; [date unknown]. For Parents: Multisystem Inflammatory Syndrome in Children (MIS-C) associated with COVID-19; 2023 Jan 3 [cited 2023 Aug 10]. Available from: https://www.cdc.gov/mis/mis-c.html

11. Jonat B, Gorelik M, Boneparth A, Geneslaw AS, Zachariah P, Shah A, Broglie L, Duran J, Morel KD, Zorrilla M, Svoboda L. Multisystem inflammatory syndrome in children associated with coronavirus disease 2019 in a children's hospital in New York city: patient characteristics

and an institutional protocol for evaluation, management, and follow-up. Pediatric critical care medicine. 2021

12. Shaikh FS, Aldhafferi N, Buker A, Alqahtani A, Dey S, Abdulhamid S, AlBuhairi DA, Alkabour RS, Atiyah WS, Bachar Chrouf S, Alshehri A. Comorbidities and risk factors for severe outcomes in COVID-19 patients in Saudi Arabia: a retrospective cohort study. Journal of Multidisciplinary Healthcare. 2021 Aug 12:2169-83.

13. World Health Organization [Internet]. [place unknown]: World Health Organization; [date unknown]. WHO Coronavirus (COVID-19) Dashboard; 2023 Aug 10 [cited 2023 Aug 10]. Available from: https://covid19.who.int/

14. Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. Nature Reviews Microbiology. 2023 Mar;21(3):133-46.

15. Centers for Disease Control and Prevention [Internet]. [place unknown]: Centers for Disease Control and Prevention; [date unknown]. Community NPIs: Flu Prevention in Community Setting; 2022 Oct 7 [cited 2023 Aug 11]. Available from: https://www.cdc.gov/nonpharmaceutical-interventions/community/index.html

16. Füszl A, Ebner J, Van den Nest M, Bouvier-Azula L, Schahawi DE, Presterl E. COVID-19 patient and personal safety–lessons learnt for pandemic preparedness and the way to the next normal. Antimicrobial Resistance & Infection Control. 2023 Dec;12(1):1-9.

17. Combden S, Forward A, Sarkar A. COVID-19 pandemic responses of Canada and United States in first 6 months: A comparative analysis. The International journal of health planning and management. 2022 Jan;37(1):50-65.

18. Centers for Disease Control and Prevention [Internet]. [place unknown]: Centers for Disease Control and Prevention; [date unknown]. Personal NPIs: Everyday Preventive Actions; 2022 Oct 7 [cited 2023 Aug 11]. Available from: https://www.cdc.gov/nonpharmaceutical-interventions/personal/index.html

19. Tisdell CA. Economic, social and political issues raised by the COVID-19 pandemic. Economic analysis and policy. 2020 Dec 1;68:17-28.

20. Pokhrel S, Chhetri R. A literature review on impact of COVID-19 pandemic on teaching and learning. Higher education for the future. 2021 Jan;8(1):133-41.

21. Global News [Internet]. [place unknown]: Global News; c2021. ‘We’re living in daily pain’: Frustration as COVID-19 drives B.C. to cancel surgeries again; 2021 Dec 23 [cited 2023 Aug 11]. Available from: https://globalnews.ca/news/8471850/bc-covid-surgery-cancelled-reaction/

22. Daley P, Castriciano S, Chernesky M, Smieja M. Comparison of flocked and rayon swabs for collection of respiratory epithelial cells from uninfected volunteers and symptomatic patients. Journal of clinical microbiology. 2006 Jun;44(6):2265-7.

23. DeByle C, Bulkow L, Miernyk K, Chikoyak L, Hummel KB, Hennessy T, Singleton R. Comparison of nasopharyngeal flocked swabs and nasopharyngeal wash collection methods for respiratory virus detection in hospitalized children using real-time polymerase chain reaction. Journal of virological methods. 2012 Oct 1;185(1):89-93.

24. Gill PJ, Richardson SE, Ostrow O, Friedman JN. Testing for respiratory viruses in children: to swab or not to swab. JAMA pediatrics. 2017 Aug 1;171(8):798-804.

25. Childs A, Zullo AR, Joyce NR, McConeghy KW, van Aalst R, Moyo P, Bosco E, Mor V, Gravenstein S. The burden of respiratory infections among older adults in long-term care: a systematic review. BMC geriatrics. 2019 Dec;19(1):1-0.

26. Mayo Clinic [Internet]. [place unknown]: Mayo Clinic; [date unknown]. Influenza (flu); 2023 Sept 13 [cited 2023 Sept 15]. Available from: https://www.mayoclinic.org/diseases-conditions/flu/diagnosis-treatment/drc-20351725

27. Ginocchio CC, McAdam AJ. Current best practices for respiratory virus testing. Journal of clinical microbiology. 2011 Sep;49(9_Supplement):S44-8.

28. Busson L, Bartiaux M, Brahim S, Konopnicki D, Dauby N, Gérard M, De Backer P, Van Vaerenbergh K, Mahadeb B, De Foor M, Wautier M. Prospective evaluation of diagnostic tools for respiratory viruses in children and adults. Journal of virological methods. 2019 Apr 1;266:1-6.

29. Spencer S, Thompson MG, Flannery B, Fry A. Comparison of respiratory specimen collection methods for detection of influenza virus infection by reverse transcription-PCR: a literature review. Journal of Clinical Microbiology. 2019 Sep;57(9):e00027-19.

30. CDC Division of Laboratory Systems [Internet]. [place unknown]: CDC Division of Laboratory Systems; [date unknown]. Diagnostic Sensitivity and Specificity for Clinical Laboratory Testing; [date unknown] [cited 2023 Aug 17]. Available from: https://www.cdc.gov/labtraining/docs/job_aids/additional_resources/Sensitivity_and_Specificity_Final_5_23_2022_508.pdf

31. Zuurbier RP, Bont LJ, Langedijk AC, Hamer M, Korsten K, Drysdale SB, Snape MD, Robinson H, Pollard AJ, Martinón-Torres F, Rodríguez-Tenreiro Sánchez C. Low sensitivity of BinaxNOW RSV in infants. The Journal of infectious diseases. 2020 Nov 1;222(Supplement_7):S640-7.

32. Olofsson S, Brittain-Long R, Andersson LM, Westin J, Lindh M. PCR for detection of respiratory viruses: seasonal variations of virus infections. Expert review of anti-infective therapy. 2011 Aug 1;9(8):615-26.

33. Semple MG, Cowell A, Dove W, Greensill J, McNamara PS, Halfhide C, Shears P, Smyth RL, Hart CA. Dual infection of infants by human metapneumovirus and human respiratory syncytial virus is strongly associated with severe bronchiolitis. The Journal of infectious diseases. 2005 Feb 1;191(3):382-6.

34. Cepheid [Internet]. [place unknown]: Cepheid; [date unknown]. GeneXpert® Systems; [date unknown] [cited 2023 Aug 18]. Available from: https://www.cepheid.com/en-CA/systems/genexpert-family-of-systems/genexpert-system.html

35. Biomérieux [Internet]. [place unknown]: Biomérieux; [date unknown]. FilmArray™ Respiratory Panel; [date unknown] [cited 2023 Aug 18]. Available from: https://www.biomerieux-nordic.com/product/filmarray-respiratory-panel

36. Heid CA, Stevens J, Livak KJ, Williams PM. Real time quantitative PCR. Genome research. 1996 Oct 1;6(10):986-94.

37. Bio-Rad [Internet]. [place unknown]: Bio-Rad; [date unknown]. What is Real-Time PCR (qPCR)?; [date unknown] [cited 2023 Aug 21]. Available from: https://www.bio-rad.com/en-ca/applications-technologies/what-real-time-pcr-qpcr?ID=LUSO4W8UU

38. Kralik P, Ricchi M. A basic guide to real time PCR in microbial diagnostics: definitions, parameters, and everything. Frontiers in microbiology. 2017 Feb 2;8:108.

39. Mackay IM. Real-time PCR in the microbiology laboratory. Clinical microbiology and infection. 2004 Mar 1;10(3):190-212.

40. Espy MJ, Uhl JR, Sloan LM, Buckwalter SP, Jones MF, Vetter EA, Yao JD, Wengenack NL, Rosenblatt JE, Cockerill III FR, Smith T. Real-time PCR in clinical microbiology: applications for routine laboratory testing. Clinical microbiology reviews. 2006 Jan;19(1):165-256.

41. Goldfarb DM, Tilley P, Al-Rawahi GN, Srigley JA, Ford G, Pedersen H, Pabbi A, Hannam-Clark S, Charles M, Dittrick M, Gadkar VJ. Self-collected saline gargle samples as an alternative to health care worker-collected nasopharyngeal swabs for COVID-19 diagnosis in outpatients. Journal of clinical microbiology. 2021 Mar 19;59(4):e02427-20.

42. Gadkar VJ, Goldfarb DM, Young V, Watson N, Hoang L, Lee T, Prystajecky N, Al-Rawahi GN, Srigley JA, Tilley P. Gargle-direct: extraction-free detection of SARS-CoV-2 using real-time PCR (RT-qPCR) of saline gargle rinse samples. MedRxiv. 2020 Jan 1.

43. Frazee BW, de la Guardia AR, Alter H, Chen CG, Fuentes EL, Holzer AK, Lolas M, Mitra D, Vohra J, Dekker CL. Accuracy and discomfort of different types of intranasal specimen collection methods for molecular influenza testing in emergency department patients. Annals of emergency medicine. 2018 Apr 1;71(4):509-17.

44. Gadenstaetter AJ, Mayer CD, Landegger LD. Nasopharyngeal versus nasal swabs for detection of SARS-CoV-2: a systematic review. Rhinology. 2021 Oct 1;59(5):410-21.

45. Utama R, Hapsari R, Puspitasari I, Sari D, Hendrianingtyas M, Nurainy N. Self-collected gargle specimen as a patient-friendly sample collection method for COVID-19 diagnosis in a population context. Scientific reports. 2022 Mar 8;12(1):1-8.

46. Yan Y, Chang L, Wang L. Laboratory testing of SARS-CoV, MERS-CoV, and SARS-CoV-2 (2019-nCoV): Current status, challenges, and countermeasures. Reviews in medical virology. 2020 May;30(3):e2106.

47. Cheng MP, Papenburg J, Desjardins M, Kanjilal S, Quach C, Libman M, Dittrich S, Yansouni CP. Diagnostic testing for severe acute respiratory syndrome–related coronavirus 2: a narrative review. Annals of internal medicine. 2020 Jun 2;172(11):726-34.

48. Tang YW, Schmitz JE, Persing DH, Stratton CW. Laboratory diagnosis of COVID-19: current issues and challenges. Journal of clinical microbiology. 2020 May 26;58(6):e00512-20.

49. Graham TG, Dugast-Darzacq C, Dailey GM, Nguyenla XH, Van Dis E, Esbin MN, Abidi A, Stanley SA, Darzacq X, Tjian R. Open-source RNA extraction and RT-qPCR methods for SARS-CoV-2 detection. PloS one. 2021 Feb 3;16(2):e0246647.

50. Wozniak A, Cerda A, Ibarra-Henriquez C, Sebastian V, Armijo G, Lamig L, Miranda C, Lagos M, Solari S, Guzmán AM, Quiroga T. A simple RNA preparation method for SARS-CoV-2 detection by RT-qPCR. Scientific Reports. 2020 Oct 6;10(1):16608.

51. Younes N, Al-Sadeq DW, Al-Jighefee H, Younes S, Al-Jamal O, Daas HI, Yassine H, Nasrallah GK. Challenges in laboratory diagnosis of the novel coronavirus SARS-CoV-2. Viruses. 2020 Jun;12(6):582.

52. Pondaven-Letourmy S, Alvin F, Boumghit Y, Simon F. How to perform a nasopharyngeal swab in adults and children in the COVID-19 era. European annals of otorhinolaryngology, head and neck diseases. 2020 Sep 1;137(4):325-7.

53. Ward S, Lindsley A, Courter J, Assa'ad A. Clinical testing for COVID-19. Journal of allergy and clinical immunology. 2020 Jul 1;146(1):23-34.

54. D'Cruz RJ, Currier AW, Sampson VB. Laboratory testing methods for novel severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2). Frontiers in cell and developmental biology. 2020 Jun 4;8:468.

55. Centers for Disease Control and Prevention [Internet]. [place unknown]: Centers for Disease Control and Prevention; [date unknown]. Interim Guidelines for Collecting and Handling of Clinical Specimens for COVID-19 Testing; 2022 July 15 [cited 2023 Aug 23]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/lab/guidelines-clinical-specimens.html

56. Benoit P, Labbé AC, Lalancette L, Gagnon S, Bonneau E, Lavallée C, Roger M, Fafard J, Dumaresq J, Beauchemin S, Coutlée F. Comparison of SARS-CoV-2 detection with the Cobas® 6800/8800 system on gargle samples using two sample processing methods with

combined oropharyngeal/nasopharyngeal swab. Journal of medical virology. 2021 Dec;93(12):6837-40.

57. Oyewole AO, Barrass L, Robertson EG, Woltmann J, O'Keefe H, Sarpal H, Dangova K, Richmond C, Craig D. COVID-19 Impact on Diagnostic Innovations: Emerging Trends and Implications. Diagnostics. 2021 Feb;11(2):182.

58. Kandel CE, Young M, Serbanescu MA, Powis JE, Bulir D, Callahan J, Katz K, McCready J, Racher H, Sheldrake E, Quon D. Detection of severe acute respiratory coronavirus virus 2 (SARS-CoV-2) in outpatients: A multicenter comparison of self-collected saline gargle, oral swab, and combined oral–anterior nasal swab to a provider collected nasopharyngeal swab. Infection Control & Hospital Epidemiology. 2021 Nov;42(11):1340-4.

59. Kinshella ML, Tilley P, Al-Rawahi GN, Srigley JA, Kayda I, Canes M, McLennan M, Bone JN, Dittrick M, Gadkar VJ, Hoang LM. Evaluation of observed and unobserved self-collection of saline gargle samples for the detection of SARS-CoV-2 in outpatients. Diagnostic microbiology and infectious disease. 2022 Feb 1;102(2):115566.

60. Callahan CJ, Lee R, Zulauf KE, Tamburello L, Smith KP, Previtera J, Cheng A, Green A, Abdul Azim A, Yano A, Doraiswami N. Open development and clinical validation of multiple 3D-printed nasopharyngeal collection swabs: rapid resolution of a critical COVID-19 testing bottleneck. Journal of Clinical Microbiology. 2020 May 11;58(8):e00876-20.

61. McElfish PA, Purvis R, James LP, Willis DE, Andersen JA. Perceived Barriers to COVID-19 Testing. International Journal of Environmental Research and Public Health. 2021 Jan;18(5):2278.

62. Sullivan PS, Sailey C, Guest JL, Guarner J, Kelley C, Siegler AJ, Valentine-Graves M, Gravens L, Del Rio C, Sanchez TH. Detection of SARS-CoV-2 RNA and antibodies in diverse samples: protocol to validate the sufficiency of provider-observed, home-collected blood, saliva, and oropharyngeal samples. JMIR public health and surveillance. 2020 Apr 24;6(2):e19054.

63. Valentine-Graves M, Hall E, Guest JL, Adam E, Valencia R, Shinn K, Hardee I, Sanchez T, Siegler AJ, Sullivan PS. At-home self-collection of saliva, oropharyngeal swabs and dried blood spots for SARS-CoV-2 diagnosis and serology: Post-collection acceptability of specimen collection process and patient confidence in specimens. PloS one. 2020 Aug 5;15(8):e0236775.

64. Global News [Internet]. [place unknown]: Global News; c2020. Coronavirus: Inside the supply chain crisis slowing Canadian labs; 2020 Oct 20 [cited 2023 Aug 25]. Available from: https://globalnews.ca/news/7406308/coronavirus-inside-supply-chain-crisis/

65. CTV News [Internet]. [place unknown]: CTV News; c2020. Long lines for COVID-19 tests reported across Canada; feds announce more funding; 2020 Sept 16 [cited 2023 Aug 25]. Available from: https://www.ctvnews.ca/health/coronavirus/long-lines-for-covid-19-tests-reported-across-canada-feds-announce-more-funding-1.5107864

66. CNN Health [Internet]. [place unknown]: CNN Health; c2020. Lines to get tested for coronavirus are growing long ahead of Thanksgiving and amid rising cases; 2020 Nov 21 [cited 2023 Aug 25]. Available from: https://www.cnn.com/2020/11/20/health/coronavirus-testing-wait-times-wellness/index.html

67. Global News [Internet]. [place unknown]: Global News; c2021. Waiting hours at COVID-19 test clinics, British Columbians hope rapid tests come soon; 2021 Dec 20 [cited 2023 Aug 25]. Available from: https://globalnews.ca/news/8463642/bc-wait-time-covid-testing-sites-rapid-test-kits/

68. CTV News Atlantic [Internet]. [place unknown]: CTV News Atlantic; c2021. Some P.E.I. COVID-19 testing sites experiencing high volumes, long wait times; 2021 Dec 8 [cited 2023 Aug 25]. Available from: https://atlantic.ctvnews.ca/some-p-e-i-covid-19-testing-sites-experiencing-high-volumes-long-wait-times-1.5699159

69. Global News [Internet]. [place unknown]: Global News; c2021. COVID-19: London, Ont. testing capacity feeling the pressure amid case spike, Omicron; 2021 Dec 16 [cited 2023 Aug 25]. Available from: https://globalnews.ca/news/8456055/covid-19-london-omicron-test-booster-capacity/

70. Manabe YC, Sharfstein JS, Armstrong K. The need for more and better testing for COVID-19. Jama. 2020 Dec 1;324(21):2153-4

71. Föh B, Borsche M, Balck A, Taube S, Rupp J, Klein C, Katalinic A. Complications of nasal and pharyngeal swabs: a relevant challenge of the COVID-19 pandemic?. European Respiratory Journal. 2021 Apr 1;57(4).

72. Biber A, Lev D, Mandelboim M, Lustig Y, Harmelin G, Shaham A, Erster O, Schwartz E. The role of mouthwash sampling in SARS-CoV-2 diagnosis. European Journal of Clinical Microbiology & Infectious Diseases. 2021 Oct;40(10):2199-206.

73. Wehrhahn MC, Robson J, Brown S, Bursle E, Byrne S, New D, Chong S, Newcombe JP, Siversten T, Hadlow N. Self-collection: An appropriate alternative during the SARS-CoV-2 pandemic. Journal of Clinical Virology. 2020 Jul 1;128:104417.

74. Kohmer N, Eckermann L, Böddinghaus B, Götsch U, Berger A, Herrmann E, Kortenbusch M, Tinnemann P, Gottschalk R, Hoehl S, Ciesek S. Self-Collected Samples to Detect SARS-CoV-2: Direct Comparison of Saliva, Tongue Swab, Nasal Swab, Chewed Cotton Pads and Gargle Lavage. Journal of Clinical Medicine. 2021 Jan;10(24):5751.

75. Montaño MA, Bemer MJ, Heller KB, Meisner A, Marfatia Z, Rechkina EA, Padgett LR, Ahls CL, Rains D, Hao L, Hsiang TY. Performance of anterior nares and tongue swabs for nucleic acid, nucleocapsid, and spike antigen testing for detecting SARS-CoV-2 against nasopharyngeal PCR and viral culture. International Journal of Infectious Diseases. 2022 Apr 1;117:287-94.

76. Boerger AC, Buckwalter S, Fernholz EC, Jannetto PJ, Binnicker MJ, Reed K, Walchak R, Woodliff E, Johnson M, Pritt BS. Evaluation of self-collected midturbinate nasal swabs and saliva for detection of SARS-CoV-2 RNA. Journal of clinical microbiology. 2021 Aug 18;59(9):10-128.

77. Bergevin MA, Freppel W, Robert G, Ambaraghassi G, Aubry D, Haeck O, Saint-Jean M, Carignan A. Validation of saliva sampling as an alternative to oro-nasopharyngeal swab for detection of SARS-CoV-2 using unextracted rRT-PCR with the Allplex 2019-nCoV assay. Journal of Medical Microbiology. 2021;70(8).

78. Ali F, Sweeney DA. No One Likes a Stick up Their Nose: Making the Case for Saliva-based Testing for Coronavirus Disease 2019 (COVID-19). Clinical Infectious Diseases. 2021 May 1;72(9):e357-8.

79. Poukka E, Mäkelä H, Hagberg L, Vo T, Nohynek H, Ikonen N, Liitsola K, Helve O, Savolainen-Kopra C, Dub T. Detection of SARS-CoV-2 Infection in Gargle, Spit, and Sputum Specimens. Microbiology spectrum. 2021 Sep 1;9(1):e00035-21.

80. Genelhoud G, Adamoski D, Spalanzani RN, Bochnia-Bueno L, de Oliveira JC, Gradia DF, Bonatto AC, Wassem R, Raboni SM, Nogueira MB, de Araujo-Souza PS. Comparison of SARS-CoV-2 molecular detection in nasopharyngeal swab, saliva, and gargle samples. Diagnostic microbiology and infectious disease. 2022 Jun 1;103(2):115678.

81. McLennan K, Barton E, Lang C, Adams IR, McAllister G, Reijns MA, Templeton K, Johannessen I, Leckie A, Gilbert N. User acceptability of saliva and gargle samples for identifying COVID-19 positive high-risk workers and household contacts. Diagnostic Microbiology and Infectious Disease. 2022 May 26:115732.

82. Bastos ML, Perlman-Arrow S, Menzies D, Campbell JR. The sensitivity and costs of testing for SARS-CoV-2 infection with saliva versus nasopharyngeal swabs: a systematic review and meta-analysis. Annals of internal medicine. 2021 Apr;174(4):501-10.

83. Czumbel LM, Kiss S, Farkas N, Mandel I, Hegyi A, Nagy Á, Lohinai Z, Szakács Z, Hegyi P, Steward MC, Varga G. Saliva as a candidate for COVID-19 diagnostic testing: a meta-analysis. Frontiers in medicine. 2020 Aug 4;7:465.

84. Bennett S, Davidson RS, Gunson RN. Comparison of gargle samples and throat swab samples for the detection of respiratory pathogens. Journal of virological methods. 2017 Oct 1;248:83-6.

85. Bennett S, MacLean A, Gunson R. Verification of Cepheid Xpert Xpress Flu/RSV assay for use with gargle samples, sputa and endotracheal secretions. Journal of Hospital Infection. 2019 Jan 1;101(1):114-5.

86. Arbaciauskaite S, Babakhani P, Sandetskaya N, Vitkus D, Jancoriene L, Karosiene D, Karciauskaite D, Zablockiene B, Kuhlmeier D. Self-sampled gargle water direct RT-LAMP as

a screening method for the detection of SARS-CoV-2 infections. Diagnostics. 2022 Mar 22;12(4):775.

87. Dumaresq J, Coutlée F, Dufresne PJ, Longtin J, Fafard J, Bestman-Smith J, Bergevin M, Vallières E, Desforges M, Labbé AC. Natural spring water gargle and direct RT-PCR for the diagnosis of COVID-19 (COVID-SPRING study). Journal of Clinical Virology. 2021 Nov 1;144:104995.

88. Gertler M, Krause E, van Loon W, Krug N, Kausch F, Rohardt C, Rössig H, Michel J, Nitsche A, Mall MA, Nikolai O. Self-collected oral, nasal and saliva samples yield sensitivity comparable to professionally collected oro-nasopharyngeal swabs in SARS-CoV-2 diagnosis among symptomatic outpatients. International Journal of Infectious Diseases. 2021 Sep 1;110:261-6.

89. Gobeille Pare S, Bestman-Smith J, Fafard J, Doualla-Bell F, Jacob-Wagner M, Lavallée C, Charest H, Beauchemin S, Coutlée F, Dumaresq J, Busque L. Natural spring water gargle samples as an alternative to nasopharyngeal swabs for SARS-CoV-2 detection using a laboratory-developed test. Journal of Medical Virology. 2022 Mar;94(3):985-93.

90. Gupta A, Mittal A, Dhakad S, Brijwal M, Soneja M, Srigyan D, Xess A, Kumar A, Ray A, Meena VP, Garg RK. Gargle lavage & saliva: Feasible & cheaper alternatives to nasal & throat swabs for diagnosis of COVID-19. The Indian journal of medical research. 2021 May;153(5-6):665

91. Hitzenbichler F, Bauernfeind S, Salzberger B, Schmidt B, Wenzel JJ. Comparison of throat washings, nasopharyngeal swabs and oropharyngeal swabs for detection of SARS-CoV-2. Viruses. 2021 Apr;13(4):653.

92. Karabay O, Toptan H, Köse E, Güçlü E, Köroğlu M. Comparison of mouthwash and naso-oropharyngeal swab samples in the diagnosis of covid-19. Flora. 2021:378-83.

93. Kocagoz T, Can O, Yurttutan Uyar N, Aksoy E, Polat T, Cankaya D, Karakus B, Mozioglu E, Kocagoz S. Simple concentration method enables the use of gargle and mouthwash instead of nasopharyngeal swab sampling for the diagnosis of COVID-19 by PCR. European Journal of Clinical Microbiology & Infectious Diseases. 2021 Dec;40(12):2617-22.

94. LeBlanc JJ, Pettipas J, Di Quinzio M, Hatchette TF, Patriquin G. Reliable detection of SARS-CoV-2 with patient-collected swabs and saline gargles: a three-headed comparison on multiple molecular platforms. Journal of Virological Methods. 2021 Sep 1;295:114184.

95. Malecki M, Lüsebrink J, Teves S, Wendel AF. Pharynx gargle samples are suitable for SARS-CoV-2 diagnostic use and save personal protective equipment and swabs. Infection Control & Hospital Epidemiology. 2021 Feb;42(2):248-9.

96. Sancak B, Mutlu E, Cirit OS, Kocagöz T, Can Ö, Çicek C, Sayiner AA, Appak Ö, Uyar NY, Külah C, Çiçek AÇ. Gargle and mouthwash can replace nasopharyngeal swab sampling when concentrated with a new method for diagnosis of COVID-19.

97. Zander J, Scholtes S, Ottinger M, Kremer M, Kharazi A, Stadler V, Bickmann J, Zeleny C, Kuiper JW, Hauck CR. Self-Collected Gargle Lavage Allows Reliable Detection of SARS-CoV-2 in an Outpatient Setting. Microbiology spectrum. 2021 Jul 14;9(1):e00361-21.

98. Mittal A, Gupta A, Kumar S, Surjit M, Singh B, Soneja M, Soni KD, Khan AR, Singh K, Naik S, Kumar A. Gargle lavage as a viable alternative to swab for detection of SARS-CoV-2. The Indian journal of medical research. 2020 Jul;152(1-2):77.

99. Olearo F, Nörz D, Hoffman A, Grunwald M, Gatzemeyer K, Christner M, Both A, Campos CE, Braun P, Andersen G, Pfefferle S. Clinical performance and accuracy of a qPCR-based SARS-CoV-2 mass-screening workflow for healthcare-worker surveillance using pooled self-sampled gargling solutions: A cross-sectional study. Journal of Infection. 2021 Nov 1;83(5):589-93.

100. Gruendl M, Kheiroddin P, Althammer M, Schöberl P, Rohrmanstorfer R, Wallerstorfer D, Ambrosch A, Kabesch M. Analysis of COVID-19 Infection Chains in a School Setting: Data From a School-Based rRT-PCR–Gargle Pool Test System. Disaster Medicine and Public Health Preparedness. 2023;17:e312.

101. Vancouver Sun [Internet]. [place unknown]: Vancouver Sun; c2021. COVID-19: B.C. Children's Hospital hands out gargle tests to Vancouver schools; 2021 Apr 12 [cited 2023 Aug 26]. Available from: https://vancouversun.com/news/covid-19-bc-childrens-hospital-hands-out-gargle-tests-to-vancouver-schools

102. CTV News [Internet]. [place unknown]: CTV News; c2020. Goodbye, nasal swabs: B.C. announces non-invasive COVID-19 test for students; 2020 Sept 18 [cited 2023 Aug 26]. Available from: https://bc.ctvnews.ca/goodbye-nasal-swabs-b-c-announces-non-invasive-covid-19-test-for-students-1.5109803

103. The Mayor.eu [Internet]. [place unknown]: The Mayor.eu; c2021. Stockholm to introduce gargle test for Covid; 2021 Oct 26 [cited 2023 Aug 26]. Available from: https://www.themayor.eu/en/a/view/stockholm-to-introduce-gargle-test-for-covid-9203

104. The Australian Financial Review [Internet]. [place unknown]: The Australian Financial Review; c2021. Why citizens in pandemic Vienna keep gargling; 2021 Sept 26 [cited 2023 Aug 26]. Available from: https://www.afr.com/policy/health-and-education/why-citizens-in-pandemic-vienna-keep-gargling-20210922-p58tzr

105. Kheiroddin P, Schöberl P, Althammer M, Cibali E, Würfel T, Wein H, Kulawik B, Buntrock-Döpke H, Weigl E, Gran S, Gründl M. Results of WICOVIR Gargle Pool PCR testing in German schools based on the first 100,000 tests. Frontiers in pediatrics. 2021 Oct 28;9:721518.

106. Campbell JR, Dion C, Uppal A, Yansouni CP, Menzies D. Systematic on-site testing for SARS-CoV-2 infection among asymptomatic essential workers in Montréal, Canada: a prospective observational and cost-assessment study. Canadian Medical Association Open Access Journal. 2022 Apr 1;10(2):E409-19.

107. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. Continuing education in anaesthesia, critical care & pain. 2008 Dec 1;8(6):221-3.

108. Wajnberg A, Mansour M, Leven E, Bouvier NM, Patel G, Firpo-Betancourt A, Mendu R, Jhang J, Arinsburg S, Gitman M, Houldsworth J. Humoral response and PCR positivity in patients with COVID-19 in the New York City region, USA: an observational study. The Lancet Microbe. 2020 Nov 1;1(7):e283-9.

109. Dendukuri N, Schiller I, De Groot J, Libman M, Moons K, Reitsma J, Van Smeden M. Concerns about composite reference standards in diagnostic research. bmj. 2018 Jan 18;360.

110. Singh S, Singh A, Prajapati S, Kabra SK, Lodha R, Mukherjee A, Singh V, Hesseling AC, Grewal H. Xpert MTB/RIF assay can be used on archived gastric aspirate and induced sputum samples for sensitive diagnosis of paediatric tuberculosis. BMC microbiology. 2015 Dec;15(1):1-0.

111. Tsang NN, So HC, Cowling BJ, Leung GM, Ip DK. Performance of saline and water gargling for SARS-CoV-2 reverse transcriptase PCR testing: a systematic review and meta-analysis. European Respiratory Review. 2022 Sep 30;31(165).

112. Walker E, Hernandez AV, Kattan MW. Meta-analysis: Its strengths and limitations. Cleveland Clinic journal of medicine. 2008 Jun 1;75(6):431.

113. Haidich AB. Meta-analysis in medical research. Hippokratia. 2010 Dec;14(Suppl 1):29.

114. Luijendijk HJ. How to create PICO questions about diagnostic tests. BMJ Evidence-Based Medicine. 2021 Aug 1;26(4):155-7.

115. Nishikawa-Pacher A. Research Questions with PICO: A Universal Mnemonic. Publications. 2022 Jun 22;10(3):21.

116. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. Bmj. 2015 Jan 2;349.

117. PROSPERO [Internet]. [place unknown]: PROSPERO; c2022. Sensitivity of gargle samples compared to swabs for SARS-CoV-2 detection with nucleic acid amplification testing; 2023 June 5 [cited 2023 Aug 26]. Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022312628

118. Page M J, Moher D, Bossuyt P M, Boutron I, Hoffmann T C, Mulrow C D et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews BMJ 2021; 372 :n160 doi:10.1136/bmj.n160

119. McInnes MDF, Moher D, Thombs BD, et al, and the PRISMA-DTA Group. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: The PRISMA DTA statement. JAMA 2018;319:388-96. doi:10.1001/jama.2017.19163

120. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Jama. 2000 Apr 19;283(15):2008-12.

121. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, Koffel JB. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Systematic reviews. 2021 Dec;10(1):1-9.

122. Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. Systematic reviews. 2017 Dec;6(1):1-2.

123. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM, QUADAS-2 Group*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine. 2011 Oct 18;155(8):529-36.

124. Moreira VM, Mascarenhas P, Machado V, Botelho J, Mendes JJ, Taveira N, Almeida MG. Diagnosis of SARS-Cov-2 Infection by RT-PCR using specimens other than naso-and oropharyngeal swabs: A systematic review and meta-analysis. Diagnostics. 2021 Feb;11(2):363.

125. Doing Meta-Analysis in R: A Hands-on Guide [Internet]. [place unknown]: Doing Meta-Analysis in R: A Hands-on Guide; [date unknown]. Pooling Effect Sizes; [date unknown] [cited 2023 Aug 26]. Available from: https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/pooling-es.html

126. Newcombe RG. Improved confidence intervals for the difference between binomial proportions based on paired data. Statistics in medicine. 1998 Nov 30;17(22):2635-50.

127. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Bmj. 2003 Sep 4;327(7414):557-60.

128. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC medical research methodology. 2014 Dec;14(1):1-2.

129. Veroniki AA, Jackson D, Viechtbauer W, Bender R, Bowden J, Knapp G, Kuss O, Higgins JP, Langan D, Salanti G. Methods to estimate the between-study variance and its uncertainty in meta-analysis. Research synthesis methods. 2016 Mar;7(1):55-79.

130. Veroniki AA, Jackson D, Bender R, Kuss O, Langan D, Higgins JP, Knapp G, Salanti G. Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis. Research synthesis methods. 2019 Mar;10(1):23-43.

131. Hamza TH, van Houwelingen HC, Stijnen T. The binomial distribution of meta-analysis was preferred to model within-study variability. Journal of clinical epidemiology. 2008 Jan 1;61(1):41-51.

132. Doing Meta-Analysis in R: A Hands-on Guide [Internet]. [place unknown]: Doing Meta-Analysis in R: A Hands-on Guide; [date unknown]. Between-Study Heterogeneity; [date unknown] [cited 2023 Aug 26]. Available from: https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/heterogeneity.html

133. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. Journal of clinical epidemiology. 2005 Sep 1;58(9):882-93.

134. Centers for Disease Control and Prevention [Internet]. [place unknown]: Centers for Disease Control and Prevention; [date unknown]. Considerations for SARS-CoV-2 Antigen Testing for Healthcare Providers Testing Individuals in the Community; 2023 May 11 [cited 2023 Sept 17]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/antigen-tests-guidelines.html

135. Farland LV, Correia KF, Wise LA, Williams PL, Ginsburg ES, Missmer SA. P-values and reproductive health: what can clinical researchers learn from the American Statistical Association?. Human Reproduction. 2016 Nov 21;31(11):2406-10.

136. Ek P, Böttiger B, Dahlman D, Hansen KB, Nyman M, Nilsson AC. A combination of naso-and oropharyngeal swabs improves the diagnostic yield of respiratory viruses in adult emergency department patients. Infectious Diseases. 2019 Apr 3;51(4):241-8.

Appendices

Appendix A: Search strategy for all platforms searched

Ovid EMBASE	
	Embase <1974 to 2022 May 05>
1	exp coronavirus disease 2019/
2	exp Severe acute respiratory syndrome coronavirus 2/
3	(covid-19 or covid19 or covid 19 or coronavirus disease 2019 or coronavirus disease-19 or disease 2019, coronavirus or 2019 novel coronavirus disease* or 2019 novel coronavirus infection* or 2019 ncov disease* or 2019 ncov infection* or 2019-ncov disease* or 2019-ncov infection* or covid 19 virus disease* or covid 19 virus infection* or covid-19 virus disease* or covid-19 virus infection* or coronavirus disease 19 or disease, 2019-ncov or disease, covid-19 virus or infection, 2019-ncov or infection, covid-19 virus or infection, sars-cov-2 or sars cov 2 infection* or sars coronavirus 2 infection* or sars-cov-2 infection*).tw,kw.
4	(2019 novel coronavirus or 2019 novel coronaviruses or 2019-ncov or covid 19 virus or covid-19 virus or covid-19 viruses or coronavirus 2, sars or coronavirus, 2019 novel or coronavirus disease 2019 virus or novel coronavirus, 2019 or sars cov 2 virus or sars coronavirus 2 or sars-cov-2 virus or sars-cov-2 viruses or severe acute respiratory syndrome coronavirus 2 or virus, covid-19 or virus, sars-cov-2).tw,kw.
5	1 or 2 or 3 or 4
6	exp gargle/
7	exp mouthwash/
8	(gargl* or (saline adj5 gargl*) or (gargle adj5 lavag*) or (gargle adj5 wash*) or (gargle adj5 rins*) or (throat adj5 lavag*) or (throat adj5 wash*) or (throat adj5 rins*) or mouthwash* or (mouth adj5 wash*) or (mouth adj5 lavag*) or (mouth adj5 rins*) or (oral adj5 rins*) or (oral adj5 wash*) or (oral adj5 lavag*)).tw,kw.
9	6 or 7 or 8
10	molecular diagnosis/ or diagnosis/ or virus diagnosis/ or laboratory diagnosis/
11	exp nucleic acid amplification/
12	covid-19 nucleic acid testing/
13	(diagnos* or detect* or test* or nucleic acid amplification or nucleic acid or naat or polymerase chain reaction or pcr or rt-pcr or rt pcr).tw,kw.
14	"sensitivity and specificity"/
15	10 or 11 or 12 or 13 or 14
16	"swab (sampler)"/
17	exp flocked swab/ or exp oral swab/ or exp nasopharyngeal swab/ or exp throat swab/ or exp mid-turbinate swab/ or exp oropharyngeal swab/
18	(swab* or nasopharyngeal or (nasopharyngeal adj5 swab*) or np or (np adj5 swab*) or oropharyngeal or (oropharyngeal adj5 swab*) or nasal or (nasal adj5 swab) or mid-turbinate or (mid-turbinate adj5 swab*) or throat or (throat adj5 swab) or oral or (oral adj5 swab)).tw,kw.
19	16 or 17 or 18
20	5 and 9 and 15 and 19

Web of Science	
1	TI= ("covid-19" OR "covid19" OR "covid 19" OR "coronavirus disease 2019" OR "coronavirus disease-19" OR "2019 novel coronavirus disease*" OR "2019 novel coronavirus infection*" OR "2019 ncov disease*" OR "2019 ncov infection*" OR "2019-ncov disease*" OR "2019-ncov infection*" OR "covid 19 virus disease*" OR "covid 19 virus infection*" OR "covid-19 virus disease*" OR "covid-19 virus infection*" OR "coronavirus disease 19" OR "sars-cov-2" OR "sars cov 2" OR "sars coronavirus 2" OR "2019 novel coronavirus" OR "2019 novel coronaviruses" OR "2019-ncov" OR "sars-cov-2 virus" or "severe acute respiratory syndrome coronavirus 2") OR AB=("covid-19" OR "covid19" OR "covid 19" OR "coronavirus disease 2019" OR "coronavirus disease-19" OR "2019 novel coronavirus disease*" OR "2019 novel coronavirus infection*" OR "2019 ncov disease*" OR "2019 ncov infection*" OR "2019-ncov disease*" OR "2019-ncov infection*" OR "covid 19 virus disease*" OR "covid 19 virus infection*" OR "covid-19 virus disease*" OR "covid-19 virus infection*" OR "coronavirus disease 19" OR "sars-cov-2" OR "sars cov 2" OR "sars coronavirus 2" OR "2019 novel coronavirus" OR "2019 novel coronaviruses" OR "2019-ncov" OR "sars-cov-2 virus" or "severe acute respiratory syndrome coronavirus 2") OR AK=("covid-19" OR "covid19" OR "covid 19" OR "coronavirus disease 2019" OR "coronavirus disease-19" OR "2019 novel coronavirus disease*" OR "2019 novel coronavirus infection*" OR "2019 ncov disease*" OR "2019 ncov infection*" OR "2019-ncov disease*" OR "2019-ncov infection*" OR "covid 19 virus disease*" OR "covid 19 virus infection*" OR "covid-19 virus disease*" OR "covid-19 virus infection*" OR "coronavirus disease 19" OR "sars-cov-2" OR "sars cov 2" OR "sars coronavirus 2" OR "2019 novel coronavirus" OR "2019 novel coronaviruses" OR "2019-ncov" OR "sars-cov-2 virus" or "severe acute respiratory syndrome coronavirus 2")
2	TI=(gargl* OR "saline gargl*" OR "gargle lavag*" OR "gargle wash*" OR "gargle rins*" OR "throat lavag*" OR "throat wash*" OR "throat rins*" OR mouthwash* OR "mouth wash*" OR "mouth lavag*" OR "mouth rins*" OR "oral rins*" OR "oral wash*" or "oral lavag*") OR AB=(gargl* OR "saline gargl*" OR "gargle lavag*" OR "gargle wash*" OR "gargle rins*" OR "throat lavag*" OR "throat wash*" OR "throat rins*" OR mouthwash* OR "mouth wash*" OR "mouth lavag*" OR "mouth rins*" OR "oral rins*" OR "oral wash*" or "oral lavag*") OR AK=(gargl* OR "saline gargl*" OR "gargle lavag*" OR "gargle wash*" OR "gargle rins*" OR "throat lavag*" OR "throat wash*" OR "throat rins*" OR mouthwash* OR "mouth wash*" OR "mouth lavag*" OR "mouth rins*" OR "oral rins*" OR "oral wash*" or "oral lavag*")
3	TI=(diagnos* OR "molecular diagnos*" or "molecular detect*" OR detect* OR test* OR "nucleic acid amplification" OR "nucleic acid amplification test*" OR "nucleic acid test*" OR "polymerase chain reaction" OR "reverse transcription polymerase chain reaction" OR "real time polymerase chain reaction" OR "PCR" OR "RT-PCR" OR "RT PCR" OR "covid-19 test*" OR "covid 19 test*" OR "covid19 test*") OR AB=(diagnos* OR "molecular diagnos*" or "molecular detect*" OR detect* OR test* OR "nucleic acid amplification" OR "nucleic acid amplification test*" OR "nucleic acid test*" OR "polymerase chain reaction" OR "reverse transcription polymerase chain reaction" OR "real time polymerase chain reaction" OR "PCR" OR "RT-PCR" OR "RT PCR" OR "covid-19 test*" OR "covid 19 test*" OR "covid19 test*") OR AK=(diagnos* OR "molecular diagnos*" or "molecular detect*" OR detect* OR test* OR "nucleic acid amplification" OR "nucleic acid amplification test*" OR "nucleic acid test*" OR "polymerase chain reaction" OR "reverse transcription polymerase chain reaction" OR "real time polymerase chain reaction" OR "PCR" OR "RT-PCR" OR "RT PCR" OR "covid-19 test*" OR "covid 19 test*" OR "covid19 test*")
4	TI=(swab* OR "nasopharyngeal swab*" OR "oropharyngeal swab*" OR "nasal swab*" OR "throat swab*" OR "oral swab*" OR "flocked swab*" OR "mid-turbinate swab*" OR "mid turbinate swab*") OR AB=(swab* OR "nasopharyngeal swab*" OR "oropharyngeal swab*" OR "nasal swab*" OR "throat swab*" OR "oral swab*" OR "flocked swab*" OR "mid-turbinate swab*" OR "mid turbinate swab*") OR AK=(swab* OR "nasopharyngeal swab*" OR "oropharyngeal swab*" OR "nasal swab*" OR "throat swab*" OR "oral swab*" OR "flocked swab*" OR "mid-turbinate swab*" OR "mid turbinate swab*")
5	1 AND 2 AND 3 AND 4

Global Index Medicus	medrxiv and biorxiv	Research Square
Searched "*All indexes*" and "*All information sources*". Searched title, abstract, subject using field codes, TW:=Text Words (words) Title + Abstract + DeCS/MeSH terms.	Searched medrxiv and biorxiv at the same time using the advanced search option. Searched "*search terms & keywords*" across full text, abstract, and title.	Searched using basic search bar in Research Square.
(tw:(gargl*))	gargl*	gargl*
(tw:(throat lavage))	throat lavage	throat lavage
(tw:(throat wash))	throat wash	throat wash
(tw:(throat rinse))	throat rinse	throat rinse
(tw:(mouthwash))	covid-19 AND test*	covid-19 AND test*
(tw:(covid-19))	covid-19 AND diagnos*	covid-19 AND diagnos*
(tw:(covid-19)) AND (tw:(test*))	covid-19 AND gargl*	covid-19 AND gargl*
(tw:(covid-19)) AND (tw:(detect*))	covid-19 AND throat lavage	covid-19 AND throat lavage
(tw:(covid-19)) AND (tw:(diagnos*))	sars-cov-2 AND test*	sars-cov-2 AND test*
(tw:(covid-19)) AND (tw:(gargl*))	sars-cov-2 AND gargl*	sars-cov-2 AND gargl*
(tw:(covid19)) AND (tw:(gargl*))	covid-19 AND swab*	covid-19 AND swab*
(tw:(covid 19)) AND (tw:(gargl*))	sars-cov-2 AND swab*	sars-cov-2 AND swab*
(tw:(covid-19)) AND (tw:(throat lavage))		
(tw:(covid19)) AND (tw:(throat lavage))		
(tw:(covid 19)) AND (tw:(throat lavage))		
(tw:(sars-cov-2))		
(tw:(sars-cov-2)) AND (tw:(test*))		
(tw:(sars-cov-2)) AND (tw:(detect*))		
(tw:(sars-cov-2)) AND (tw:(gargl*))		
(tw:(covid-19)) AND (tw:(swab*))		
(tw:(sars-cov-2)) AND (tw:(swab*))		

Appendix B: Excluded full-text articles with reason for exclusion

	Study	Reference	Reason for exclusion
1	Caza 2021	Caza M, Hogan CA, Jassem A, Prystajecky N, Hadzic A, Wilmer A. Evaluation of the clinical and analytical performance of the Seegene allplex™ SARS-CoV-2 variants I assay for the detection of variants of concern (VOC) and variants of interests (VOI). Journal of Clinical Virology. 2021 Nov 1;144:104996.	Samples not matched
2	Mora-Aguilera 2022	Mora-Aguilera G, Martínez-Bustamante V, Acevedo-Sánchez G, Coria-Contreras JJ, Guzmán-Hernández E, Flores-Colorado OE, Mendoza-Ramos C, Hernández-Nava G, Álvarez-Maya I, Gutiérrez-Espinosa MA, Gómez-Linton R. Surveillance web system and mouthwash-saliva qPCR for labor ambulatory SARS-CoV-2 detection and prevention. International Journal of Environmental Research and Public Health. 2022 Jan 24;19(3):1271.	No gargling during collection
3	Jin 2020	Jin XD, Li Y, Song YS, Yang ZZ, Wang P, Wei TT, Fan TL. Progress in research on the detection of the novel coronavirus in human samples of different groups. Eur Rev Med Pharmacol Sci. 2020 Jan 1;24(20):10879-84.	Wrong type of article
4	Campbell 2021	Campbell JR, Dion C, Uppal A, Yansouni CP, Menzies D. Systematic testing for SARS-CoV-2 infection among essential workers in Montréal, Canada: a prospective observational and cost assessment study. medRxiv. 2021 Jan 1.	Wrong study objective
5	Maricic 2020	Maricic T, Nickel O, Aximu-Petri A, Essel E, Gansauge M, Kanis P, Macak D, Richter J, Riesenberg S, Bokelmann L, Zeberg H. A direct RT-qPCR approach to test large numbers of individuals for SARS-CoV-2. PLoS One. 2020 Dec 31;15(12):e0244824.	Wrong study objective
6	Guo 2020	Guo WL, Jiang Q, Ye F, Li SQ, Hong C, Chen LY, Li SY. Effect of throat washings on detection of 2019 novel coronavirus. Clinical Infectious Diseases. 2020 Oct 15;71(8):1980-1.	Tested after 10 days since diagnosis (for re-sampling)
7	Laferl 2022	Laferl H, Seitz T, Baier-Grabner S, Kelani H, Scholz E, Heger F, Götzinger F, Frischer T, Wenisch C, Allerberger F. Evaluation of RT-qPCR of mouthwash and buccal swabs for detection of SARS-CoV-2 in children and adults. American Journal of Infection Control. 2022 Feb 1;50(2):176-81.	No gargling during collection
8	Kang 2022	Kang H, Allison S, Spangenberg A, Carr T, Sprissler R, Halonen M, Cusanovich DA. Evaluation of Swab-Seq as a scalable, sensitive assay for community surveillance of SARS-CoV-2 infection. Scientific Reports. 2022 Feb 23;12(1):1-2.	Wrong study objective
9	Defeche 2021	Defêche J, Azarzar S, Mesdagh A, Dellot P, Tytgat A, Bureau F, Gillet L, Belhadj Y, Bontems S, Hayette MP, Schils R. In-depth longitudinal comparison of clinical specimens to detect SARS-CoV-2. Pathogens. 2021 Oct 21;10(11):1362.	Tested after 10 days since diagnosis (for re-sampling)
10	Clementino 2022	Clementino M, Cavalcante KF, Viana VA, Silva DD, Damasceno CR, Fernandes de Souza J, Gondim RN, Jorge DM, Magalhães LM, Arruda ÉA, Neto RD. Detection of SARS-CoV-2 in different human biofluids using the loop-mediated isothermal amplification assay: A prospective diagnostic study in Fortaleza, Brazil. Journal of Medical Virology. 2022 Sep;94(9):4170-80.	Wrong study objective
11	Turriziani 2021	Turriziani O, Sciandra I, Mazzuti L, Di Carlo D, Bitossi C, Calabretto M, Guerrizio G, Oliveto G, Riveros Cabral RJ, Viscido A, Falasca F. SARS-CoV-2 diagnostics in the virology laboratory of a University Hospital in Rome during the lockdown period. Journal of medical virology. 2021 Feb;93(2):886-91.	Wrong study objective
12	Jadhav 2022	JADHAV RB, PATIL SS, DEOLEKAR P, YADAV P, DONGERKERY K. A comparative study to evaluate the use of saline nasal lavage and gargling in patients with Covid-19 infection. International Journal of Pharmaceutical Research. 2022 Jan;14(1).	Wrong study objective
13	McMillen 2020	McMillen T, Jani K, Viale A, Robilotti E, Aslam A, Sokoli D, Mason G, Shah M, Korenstein D, Kamboj M, Babady E. Comparison of Oral Rinses and Nasopharyngeal Swabs for the Detection of SARS-CoV-2 RNA. Journal of Molecular Diagnostics. 2020:S37-8.	No gargling during collection
14	Saito 2020	Saito M, Adachi E, Yamayoshi S, Koga M, Iwatsuki-Horimoto K, Kawaoka Y, Yotsuyanagi H. Gargle lavage as a safe and sensitive alternative to swab samples to diagnose COVID-19: a case report in Japan. Clinical Infectious Diseases. 2020 Jul 28;71(15):893-4.	Wrong type of article

	Study	Reference	Reason for exclusion
15	Kheiroddin 2021	Kheiroddin P, Schöberl P, Althammer M, Cibali E, Würfel T, Wein H, Kulawik B, Buntrock-Döpke H, Weigl E, Gran S, Gründl M. Results of WICOVIR Gargle Pool PCR Testing in German Schools Based on the First 100,000 Tests. Frontiers in pediatrics. 2021:1162.	Wrong study objective
16	Michel 2021	Michel W, Färber J, Dilas M, Heuft HG, Tammer I, Baar J, Kaasch AJ. A combined oro-nasopharyngeal swab is more sensitive than mouthwash in detecting SARS-CoV-2 by a high-throughput PCR assay. Infection. 2021 Jun;49(3):527-31.	Tested after 10 days since diagnosis (for re-sampling)
17	Lopez-Lopes 2020	Lopez-Lopes GI, Ahagon C, Benega MA, da Silva DB, Silva VO, de Oliveira Santos KC, do Prado LS, dos Santos FP, Cilli A, Saraceni C, da Cruz NB. Throat wash as a source of SARS-CoV-2 RNA to monitor community spread of COVID-19. medRxiv. 2020 Jan 1.	Samples not matched
18	Ali 2020	Ali F, Sweeney DA. Throat-Wash Testing and Coronavirus Disease 2019: Should We Put Our Money Where Our Mouth Is?. Clinical Infectious Diseases. 2020 Oct 15;71(8):1982-3.	Wrong type of article
19	Meuris 2021	Meuris C, Kremer C, Geerinck A, Locquet M, Bruyère O, Defêche J, Meex C, Hayette MP, Duchene L, Dellot P, Azarzar S. Transmission of SARS-CoV-2 after COVID-19 screening and mitigation measures for primary school children attending school in Liège, Belgium. JAMA network open. 2021 Oct 1;4(10):e2128757-.	No comparison conducted between swabs and gargling
20	Babady 2021	Babady NE, McMillen T, Jani K, Viale A, Robilotti EV, Aslam A, Diver M, Sokoli D, Mason G, Shah MK, Korenstein D. Performance of severe acute respiratory syndrome coronavirus 2 real-time RT-PCR tests on oral rinses and saliva samples. The Journal of Molecular Diagnostics. 2021 Jan 1;23(1):3-9.	No gargling during collection
21	Lévesque 2022	Lévesque S, Beauchemin S, Vallée M, Longtin J, Jacob-Wagner M, Dumaresq J, Dulcey C, Labbé AC. Evaluation of water gargle samples for SARS-CoV-2 detection using Abbott ID NOW COVID-19 assay. Journal of Medical Virology. 2022 May 9.	No comparison conducted between swabs and gargling
22	Lai 2021	Lai CK, Lui GC, Chen Z, Cheung YY, Cheng KC, Leung AS, Ng RW, Cheung JL, Yeung AC, Ho WC, Chan KC. Comparison of self-collected mouth gargle with deep-throat saliva samples for the diagnosis of COVID-19: Mouth gargle for diagnosis of COVID-19. Journal of Infection. 2021 Oct 1;83(4):496-522.	Wrong comparator

Appendix C: Data extraction form

Data field	Extracted data
Extraction information	Date form completed, name of data extractor
Study information	Study title, study ID, reference, author contact, country, language and translation information, publication type, funding source, conflicts of interest, study duration, study design
Participants and setting	Age group, median age, sex distribution, presence of symptoms, setting and participant group
Sample collection and testing methods	Timing of collections and testing
Reference swab sample	Reference swab type used, site sampled, transport media used
Gargle sample	Type and volume of liquid, gargle collection method used, length of time required to wait prior to gargle
Lab testing	Nucleic acid amplification method used, testing kit and device, gene targets and controls used, criteria for positive result
Results and outcomes	Main study aims and conclusions
Testing results	Number of paired samples tested, number of swabs tested, number of gargles tested, numbers of reported positives on gargle, swab, both, or either sample type
Other information	Correspondence needed for further information, further study information requested, correspondence details

Appendix D: Adapted QUADAS-2 quality assessment form

Quality assessment information	Date form completed, name of assessor
Study information	Study title, Study ID
I. PATIENT SELECTION	
Risk of bias	
Was a consecutive or random sample of patients enrolled?	yes/no/unclear
Was a case-control design avoided?	yes/no/unclear
Did the study avoid inappropriate exclusions?	yes/no/unclear
Could the selection of patients have introduced bias?	**RISK: LOW/HIGH/UNCLEAR**
Concerns regarding applicability	
Is there concern that the included patients do not match the review question?	**CONCERN: LOW/HIGH/UNCLEAR**
II. INDEX TEST	
Risk of bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	yes/no/unclear
If a Ct value threshold was used, was it pre-specified?	yes/no/unclear
Could the conduct or interpretation of the index test have introduced bias?	**RISK: LOW/HIGH/UNCLEAR**
Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation, differ from the review question?	**CONCERN: LOW/HIGH/UNCLEAR**
III. REFERENCE STANDARD	
Risk of bias	
Is the reference standard likely to correctly classify the target condition?	yes/no/unclear
Were the reference standard results interpreted without knowledge of the results of the index test?	yes/no/unclear
Could the reference standard, its conduct, or interpretation have introduced bias?	**RISK: LOW/HIGH/UNCLEAR**
Concerns regarding applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN: LOW/HIGH/UNCLEAR**
IV. FLOW AND TIMING	
Risk of bias	
Was there an appropriate interval between index test and reference standard?	yes/no/unclear
Did patients receive the same reference standard?	yes/no/unclear
Were all patients included in the analysis?	yes/no/unclear
Could the patient flow have introduced bias?	**RISK: LOW/HIGH/UNCLEAR**

Appendix E: Adapted QUADAS-2 quality assessment form guidance

Review question: What is the sensitivity of self-collected gargle samples when compared to healthcare worker collected swab samples for the detection of SARS-CoV-2 using nucleic acid amplification testing in individuals with confirmed or potential COVID-19?

Population: Individuals (children and/or adults) providing matched gargle and swab samples for COVID-19 testing
Presentation: asymptomatic or symptomatic; any setting (outpatient, inpatient etc); may have a previously NAAT lab-confirmed COVID-19 diagnosis and being resampled for study
Target condition: COVID-19 diagnosis via nucleic acid amplification test (NAAT)
Index test: self-collected gargle sample (including water or saline) on a NAAT
Reference standard: healthcare worker collected swab sample (including nasopharyngeal, oropharyngeal, oral, nasal, and/or throat swabs or combinations of these) on a NAAT

- Go step-by-step through each of the 4 domains in this quality assessment form
    - Review the study report and questions asked in the QA form - make judgments and record in the QA form.
    - Make sure to describe relevant information to support/justify the judgments made (ie. summarize ideas, direct quotes, explanation of flaws in study methods etc.) in the description box provided.
    - The “unclear” category should be used only when insufficient data are reported to permit a judgment. Indicate “NR” or “not reported” in the description box provided.

I. PATIENT SELECTION	
Risk of bias	
Was a consecutive or random sample of patients enrolled?	ie. if prospectively enrolled all eligible patients with suspected COVID-19 (no participants with previous, lab-confirmed COVID diagnosis) -> yes ie. if enrolled only (or some) patients with a lab-confirmed, known COVID-19 diagnosis and patients resampled for study -> no ie. if not enough information reported on study enrollment methods to make judgment -> unclear
Was a case-control design avoided?	ie. if prospectively enrolled all eligible patients with suspected COVID-19 (no participants with previous, lab-confirmed COVID diagnosis) -> yes ie. if enrolled only (or some) patients with a lab-confirmed, known COVID-19 diagnosis and patients resampled for study -> no ie. if not enough information reported on study enrollment methods to make judgment -> unclear
Did the study avoid inappropriate exclusions?	ie. if didn't exclude any eligible patients; or if appropriately excluded patients (ie. based on who couldn't gargle, couldn't provide both matched samples, no consent provided) -> yes ie. if excluded eligible patients for inappropriate reasons (ie. based on race, SES, gender) -> no ie. if not enough information reported on study enrollment methods to make judgment -> unclear
Could the selection of patients have introduced bias?	**RISK: LOW/HIGH/UNCLEAR** Summarize answers to the 3 signaling questions to judge overall risk. If all "yes" -> **LOW** If all "no" -> **HIGH** If any two "yes" and one "no" -> **HIGH** If any one "yes" and two "no" -> **HIGH** If any one "unclear" and two "no" -> **HIGH** If all "unclear" -> **UNCLEAR** If any two "yes" and one "unclear" -> **UNCLEAR** If any one "yes" and two "unclear" -> **UNCLEAR** If any one "no" and two "unclear" -> **UNCLEAR** If any one "yes", one "no", one "unclear -> **UNCLEAR**
Concerns regarding applicability	
Is there concern that the included patients do not match the review question?	**CONCERN: LOW/HIGH/UNCLEAR** See review question details and compare study's applicability to review question.
II. INDEX TEST	
Risk of bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	ie. if gargle NAAT results interpreted without knowledge of swab NAAT results -> yes ie. if gargle NAAT results interpreted after knowledge of swab NAAT results -> no ie. if not enough information reported on testing and interpretation order to make a judgment -> unclear
If a Ct value threshold was used, was it pre-specified?	ie. if used lab-developed test and prespecified Ct value cutoff values for result interpretation before testing all samples on an assay; or if used pre-specified Ct value cutoff values from a manufacturer for result interpretation before testing all samples on assay -> yes

	ie. if did not use pre-specified Ct cutoff values from LDT or manufacturer before testing all samples on assay -> no ie. if not enough information reported on testing and test interpretation to make a judgment -> unclear
Could the conduct or interpretation of the index test have introduced bias?	**RISK: LOW/HIGH/UNCLEAR** Summarize answers to the 2 signaling questions to judge overall risk. If all "yes" -> **LOW** If all "no" -> **HIGH** If any one "yes" and one "no" -> **HIGH** If all "unclear" -> **UNCLEAR** If any one "yes" and one "unclear" -> **UNCLEAR** If any one "no" and one "unclear" -> **UNCLEAR**
Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation, differ from the review question?	**CONCERN: LOW/HIGH/UNCLEAR** See review question details and compare study's applicability to review question.
III. REFERENCE STANDARD	
Risk of bias	
Is the reference standard likely to correctly classify the target condition?	ie. if used healthcare worker collected swab on nucleic acid amplification test as reference standard -> yes ie. if did not use a healthcare worker collected swab on nucleic acid amplification test as reference standard -> no ie. if not enough information reported on collection and testing methods to make a judgment -> unclear
Were the reference standard results interpreted without knowledge of the results of the index test?	ie. if swab NAAT results interpreted without knowledge of gargle NAAT results -> yes ie. if swab NAAT results interpreted after knowledge of gargle NAAT results -> no ie. if not enough information reported on testing and interpretation order to make a judgment -> unclear
Could the reference standard, its conduct, or interpretation have introduced bias?	**RISK: LOW/HIGH/UNCLEAR** Summarize answers to the 2 signaling questions to judge overall risk. If all "yes" -> **LOW** If all "no" -> **HIGH** If any one "yes" and one "no" -> **HIGH** If all "unclear" -> **UNCLEAR** If any one "yes" and one "unclear" -> **UNCLEAR** If any one "no" and one "unclear" -> **UNCLEAR**
Concerns regarding applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN: LOW/HIGH/UNCLEAR** See review question details and compare study's applicability to review question.

IV. FLOW AND TIMING	
Risk of bias	
Was there an appropriate interval between index test and reference standard?	ie. if matched samples collected at same time; or if matched samples collected within 24 hours of each other -> yes ie. if matched samples collected at different times more than 24 hours apart -> no ie. if not enough information reported on sample collection and testing to make judgment -> unclear
Did patients receive the same reference standard?	ie. if all patients enrolled in study consistently had same healthcare worker swab (ie. all NP swabs) collected, which were all tested with the same nucleic acid amplification method (ie. same RT-PCR assay and reagents) -> yes ie. if all (or some) patients had different swab samples (ie. some NP, some OP swabs) collected and/or tested with different nucleic acid amplification methods (ie. some with one RT-PCR assay and reagents, others with different assay or reagents); collection and testing inconsistent within study -> no ie. if not enough information reported on collection and testing methods to make judgment -> unclear
Were all patients included in the analysis?	ie. if all patients in study were included in data analysis; or patients appropriately excluded (ie. failed collection, indeterminate result) but reasons for exclusion justified and reported -> yes ie. if some patients in study were inappropriately excluded (ie. based on race, SES, gender) from data analysis; or reasons for exclusion not justified or reported -> no ie. if not enough information reported on data analysis methods to make judgment -> unclear
Could the patient flow have introduced bias?	**RISK: LOW/HIGH/UNCLEAR** Summarize answers to the 3 signaling questions to judge overall risk. If all "yes" -> **LOW** If all "no" -> **HIGH** If any two "yes" and one "no" -> **HIGH** If any one "yes" and two "no" -> **HIGH** If any one "unclear" and two "no" -> **HIGH** If all "unclear" -> **UNCLEAR** If any two "yes" and one "unclear" -> **UNCLEAR** If any one "yes" and two "unclear" -> **UNCLEAR** If any one "no" and two "unclear" -> **UNCLEAR** If any one "yes", one "no", one "unclear -> **UNCLEAR**

Overall assessment of study

Overall risk of bias?	**LOW RISK OF BIAS / AT RISK OF BIAS** Summarize overall risk of bias from all 4 domains. If a study is "LOW" on all 4 bias domains -> **LOW RISK OF BIAS** If a study is "HIGH" or "UNCLEAR" in even 1 of 4 bias domains -> **AT RISK OF BIAS**
Overall concerns for applicability	**LOW CONCERN REGARDING APPLICABILITY / CONCERNS REGARDING APPLICABILITY** Summarize overall concern regarding applicability from all 3 domains. If a study is "LOW" on all 3 applicability domains -> **LOW CONCERN REGARDING APPLICABILITY** If a study is "HIGH" or "UNCLEAR" in even 1 of 3 applicability domains -> **CONCERNS REGARDING APPLICABILITY**

Appendix F: Characteristics of all included studies/comparison groups

	Study ID [Reference]	Country	Study duration	Study design	Age group	Presence of symptoms	Setting and participants	Reference swab sample type	Gargle sample type	Gargle collection method	Wait time requirement prior to gargle collection	Testing method	Testing kit and device	Testing targets and controls
1	Arbaciauskaite 2022 [86]	Lithuania	March - April 2021	Prospective cohort	Adults	Symptomatic	Inpatients and outpatients	Nasopharyngeal and oropharyngeal swab Synthetic fiber swab with universal transport media (UTM; Copan, Brescia, Italy)	Saline gargle 10 mL of 0.9% saline	Gargle for around 5–10 seconds	None	RT-PCR	**Kit:** TaqPath COVID-19 Combo qPCR kit (ThermoFischer Scientific, Waltham, MA, USA) **Device:** CFX96 C1000 thermal cycler (Bio-Rad Laboratories, Philadelphia, PA, USA)	**Target(s):** E, N, and S **Control(s):** Manufacturer internal control
2	Benoit 2021a [56]	Canada	November - December 2020	Prospective cohort	Adults and children (> 7 years old)	Symptomatic	Outpatients only HCWs included	Nasopharyngeal and oropharyngeal swab Flocked swab with 4.3 mL of PCR media	Water gargle 5 mL of natural spring water (ESKA®, St-Mathieu-d'Harricana	Gargle for 5 seconds in the mouth, 5 seconds in the throat, to repeat this process once	Participants were asked not to eat, drink, or smoke for 15 min before.	RT-PCR	**Kit:** Cobas® SARS-CoV-2 test **Device:** Cobas® 6800 and Cobas® 8800 systems *Gargle sample in Cobas® PCR Media	**Target(s):** E and ORF1a/b **Control(s):** Manufacturer internal control
3	Benoit 2021b [56]	Canada	November - December 2020	Prospective cohort	Adults and children (> 7 years old)	Symptomatic	Outpatients only HCWs included	Nasopharyngeal and oropharyngeal swab Flocked swab with 4.3 mL of PCR media	Water gargle 5 mL of natural spring water (ESKA®, St-Mathieu-d'Harricana)	Gargle for 5 seconds in the mouth, 5 seconds in the throat, to repeat this process once	Participants were asked not to eat, drink, or smoke for 15 min before.	RT-PCR	**Kit:** Cobas® SARS-CoV-2 test **Device:** Cobas® 6800 and Cobas® 8800 systems *Gargle sample in Cobas® Omni Lysis Reagent	**Target(s):** E and ORF1a/b **Control(s):** Manufacturer internal control
4	Biber 2021* [72]	Israel	July - September 2020	Prospective cohort	Adults	Symptomatic and asymptomatic	Outpatients only Lab-confirmed COVID-19 patients included	Nasopharyngeal and oropharyngeal swab	Saline gargle 10 mL of 0.9% saline	Rinse and gargle for about 10–20 seconds	Not reported	RT-PCR	**Kit:** Seegene Allplex CoV19 detection kit **Device:** Bio-Rad CFX96 thermal cycler	**Target(s):** E, N, and RdRP **Control(s):** Manufacturer internal control
5	Dumaresq 2021 [87]	Canada	October 2020	Prospective cohort	Adults and children (≥ 6 years old)	Symptomatic and asymptomatic	Outpatients only	Nasopharyngeal and oropharyngeal swab Flocked swab with 3 mL of molecular water (PCR grade water)	Water gargle 5 mL of natural spring water (Eska water, St-Mathieu-d'Harricana, Québec, Canada or Naya water, Mirabel, Québec, Canada)	Rinse their mouth for 5 seconds, tilt their head back and gargle for 5 seconds, repeat this cycle once	None	Direct RT-PCR	**Kit:** Allplex™ 2019-nCoV Assay kit **Device:** CFX96 Touch Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA)	**Target(s):** E, N, and RdRP **Control(s):** Manufacturer internal control
6	Genelhoud 2022 [80]	Brazil	August - November 2020	Prospective cohort	Adults	Symptomatic and asymptomatic	Outpatients only HCWs included	Nasopharyngeal swab Rayon swab with viral transport media	Saline gargle 5 mL of 0.9% saline	Instructed to avoid mouth washing and perform actual gargling for at least ten seconds	None	RT-PCR	**Kit:** BIOMOL OneStep/COVID-19 kit (Instituto de Biologia Molecular do Paraná - IBMP, Brazil) **Device:** ViiA7™ instrument (Thermo Fisher Scientific Inc., USA)	**Target(s):** N and ORF1a/b **Control(s):** Human RNase P
7	Gertler 2021 [88]	Germany	December 2020 - January 2021	Prospective cohort	Adults and teens (≥ 17 years old)	Symptomatic	Outpatients only Lab-confirmed COVID-19 patients included	Nasopharyngeal and oropharyngeal swab Nylon flocked swab (ESwab®Copan, Italy) with 1 mL of Amies preservation medium	Water gargle 10 mL of tap water	Gargle with the water for at least 10-15 seconds	None	RT-PCR	**Kit:** AgPath-ID™ One-Step RT-PCR Reagents kit (Applied Biosystems, Foster City, CA USA) **Device:** Bio-Rad CFX96 device	**Target(s):** E and ORF1a/b **Control(s):** Human *c-myc* gene and an artificial *KoMa* sequence

	Study ID [Reference]	Country	Study duration	Study design	Age group	Presence of symptoms	Setting and participants	Reference swab sample type	Gargle sample type	Gargle collection method	Wait time requirement prior to gargle collection	Testing method	Testing kit and device	Testing targets and controls
8	Gobeille Pare 2022 [89]	Canada	November - December 2020	Prospective cohort	Adults	Symptomatic and asymptomatic	Outpatients only HCWs included	Nasopharyngeal and oropharyngeal swab Nylon flocked swab with 3 mL of molecular water (RNase/DNase free)	Water gargle 5 mL of natural spring water (ESKA®)	Rinse their mouth and their throat for a total of 20 seconds (5 seconds in the mouth, 5 seconds in the throat, 5 seconds in the mouth, 5 seconds in the throat)	Participants were asked not to eat, drink, or smoke for 15 min before.	RT-PCR	**Kit:** Laboratory developed test (LDT) with TaqPath 1-Step Multiplex NO ROX; ThermoFisher Scientific catno. A28523) **Device:** Roche LightCycler 480 II instrument (Roche) and QuantStudio™6 Real-TimePCR System (Thermo Fisher)	**Target(s):** E **Control(s):** Human RNase P or beta-actin
9	Goldfarb 2021a [41]	Canada	May - September 2020	Prospective cohort	Adults and children (≥ 4 years old)	Symptomatic	Outpatients only Lab-confirmed COVID-19 patients included	Nasopharyngeal swab Nylon flocked swab with 3 mL of universal viral transport system medium (Beckon Dickinson, Sparks, MD)	Saline gargle 5 mL of 0.9% saline (AddiPak; Teleflex Medical, Research Triangle Park, NC)	Swish the contents for 5 seconds followed by tilting their heads back and gargling for 5 seconds. This swish/gargle cycle was repeated 2 more times	Participants were asked to not eat, drink, smoke, brush their teeth, or chew gum 1 hour prior to collection.	RT-PCR	**Kit:** Laboratory developed test (LDT) **Device:** Applied Biosystems 7500 fast real-time PCR system (Life Technologies, Carlsbad, CA)	**Target(s):** E and RdRP **Control(s):** Human RNase P
10	Goldfarb 2021b [41]	Canada	May - September 2020	Prospective cohort	Adults and children (≥ 4 years old)	Symptomatic	Outpatients only Lab-confirmed COVID-19 patients included	Nasopharyngeal swab Nylon flocked swab with 3 mL of universal viral transport system medium (Beckon Dickinson, Sparks, MD)	Saline gargle 5 mL of 0.9% saline (AddiPak; Teleflex Medical, Research Triangle Park, NC)	Swish the contents for 5 seconds followed by tilting their heads back and gargling for 5 seconds. This swish/gargle cycle was repeated 2 more times	Participants were asked to not eat, drink, smoke, brush their teeth, or chew gum 1 hour prior to collection.	RT-PCR	**Kit:** Cepheid Xpert Xpress SARS-CoV-2 assay **Device:** GeneXpert system (Cepheid, Sunnyvale, CA)	**Target(s):** E and N2 **Control(s):** Manufacturer internal control
11	Gupta 2021 [90]	India	Not reported	Prospective cohort	Adults	Symptomatic and asymptomatic	Inpatients only Lab-confirmed COVID-19 patients included	Nasal (mid turbinate) and oropharyngeal swab Nylon flocked swab with 2-3 mL of viral transport medium	Saline gargle 5 mL of normal saline	Gargle for 15-20 seconds	Not reported	RT-PCR	**Kit:** Fosun COVID-19 RT-PCR detection kit (Shanghai, China) **Device:** Agilent AriaMx real-time PCR system (Agilent Technologies Inc., USA)	**Target(s):** E, N, and ORF1a/b **Control(s):** Manufacturer internal control
12	Hitzenbichler 2021a [91]	Germany	April - December 2020	Prospective cohort	Adults	Symptomatic and asymptomatic	Inpatients only Lab-confirmed COVID-19 patients included	Nasopharyngeal swab Nylon flocked swab (FLOQSwabs 519C, Copan) with no liquid transport media (dry)	Saline or water gargle 10 mL of medical grade saline or water	Gargle for 5–10 seconds	None	RT-PCR	**Kit:** TaqPathTM 1-Step RT-qPCR Master Mix, CG (Thermo Fisher Scientific, Waltham, MA, USA) **Device:** StepOnePlus Real-Time PCR System (Thermo Fisher)	**Target(s):** E **Control(s):** MS2 bacteriophages
13	Hitzenbichler 2021b [91]	Germany	April – December 2020	Prospective cohort	Adults	Symptomatic and asymptomatic	Inpatients only Lab-confirmed COVID-19 patients included	Oropharyngeal swab Nylon flocked swab (FLOQSwabs 519C, Copan) with no liquid transport media (dry)	Saline or water gargle 10 mL of medical grade saline or water	Gargle for 5–10 seconds	None	RT-PCR	**Kit:** TaqPathTM 1-Step RT-qPCR Master Mix, CG (Thermo Fisher Scientific, Waltham, MA, USA) **Device:** StepOnePlus Real-Time PCR System (Thermo Fisher)	**Target(s):** E **Control(s):** MS2 bacteriophages

	Study ID [Reference]	Country	Study duration	Study design	Age group	Presence of symptoms	Setting and participants	Reference swab sample type	Gargle sample type	Gargle collection method	Wait time requirement prior to gargle collection	Testing method	Testing kit and device	Testing targets and controls
14	Kandel 2021 [58]	Canada	August - September 2020	Prospective cohort	Adults	Symptomatic and asymptomatic	Outpatients only	Nasopharyngeal swab Nylon flocked swab (Iclean) with guanidine thiocyanate-based transport medium (McMaster Molecular Medium, Bay Area Health Trustee Corporation, Hamilton, Ontario)	Saline gargle 3 mL of 0.9% saline	Rinse your mouth by swishing the liquid back and forth for 5 seconds then gargle like you are using mouthwash for 5 seconds. Repeat this two additional times	None	RT-PCR	**Swab** **Kit:** Luna Universal Probe One-Step RT-qPCR kit (New England Biolabs, Whitby, Ontario) **Device:** CFX96 Touch Real-time PCR detection system (BioRad, Mississauga, Ontario) **Gargle** **Kit:** ThermoFisher TaqPath COVID19 Combo Kit (ThermoFisher Scientific, Waltham, MA) **Device:** Real-time PCR system 7500 Fast or QuantStudio 6 (Applied Biosystems, Waltham, MA) *Paired samples tested on different platforms	**Swab** **Target(s):** E and 5'-UTR **Control(s):** Human RNase P **Gargle** **Target(s):** E, S, and ORF1a/b **Control(s):** Manufacturer internal control Paired samples tested on different platforms
15	Karabay 2021* [92]	Turkey	March - April 2020	Prospective cohort	Not reported	Symptomatic	Outpatients only	Nasopharyngeal and oropharyngeal swab Dacron swab with 2 mL of viral transport medium	Pre-made solution gargle 5 mL of pre-made solution	Gargle with the head in full flexion position for at least 5 seconds	Not reported	RT-PCR	**Kit:** genesis Real-Time PCR COVID-19 (Primer Design, UK) kit **Device:** Not reported	**Target(s):** ORF1a/b **Control(s):** Not reported
16	Kinshella 2022 [59]	Canada	September - October 2020	Prospective cohort	Adults and children (≥ 4 years old)	Symptomatic	Outpatients only	Nasopharyngeal swab Nylon flocked swab with 3 mL of universal viral transport system media (Beckon Dickinson, Sparks, MD)	Saline gargle 5 mL of 0.9% saline (Addipak, Teleflex Medical, Research Triangle Park, NC, USA)	Swish the contents for 5 seconds followed by tilting their heads back and gargling for 5 seconds. Instructions indicated that users should repeat this swish/gargle cycle 2 more times	Participants were eligible if they had not eaten, drank, or brushed their teeth within the hour prior to sample collection.	RT-PCR	**Kit:** Laboratory developed test (LDT) **Device:** Applied Biosystems 7500 Fast Real-Time PCR System (Life Technologies, Carlsbad, CA)	**Target(s):** E and RdRP **Control(s):** Human RNase P
17	Kocagoz 2021a [93]	Turkey	Not reported	Prospective cohort	Adults	Symptomatic	Inpatients only	Nasopharyngeal swab	Water gargle Few sips of regular drinking water	Gargle and rigorously rinse their mouth forcefully for at least 10 seconds	Not reported	Direct RT-PCR	**Kit:** Commercial PCR kits (Bioeksen and A1 Lifesciences, Istanbul, Turkey) **Device:** Not reported *Gargle sample unconcentrated	**Target(s):** Not reported **Control(s):** Not reported
18	Kocagoz 2021b [93]	Turkey	Not reported	Prospective cohort	Adults	Symptomatic	Inpatients only	Nasopharyngeal swab	Water gargle Few sips of regular drinking water	Gargle and rigorously rinse their mouth forcefully for at least 10 seconds	Not reported	Direct RT-PCR	**Kit:** Commercial PCR kits (Bioeksen and A1 Lifesciences, Istanbul, Turkey) **Device:** Not reported *Gargle sample concentrated with MyMagiCon-RW100®	**Target(s):** Not reported **Control(s):** Not reported

	Study ID [Reference]	Country	Study duration	Study design	Age group	Presence of symptoms	Setting and participants	Reference swab sample type	Gargle sample type	Gargle collection method	Wait time requirement prior to gargle collection	Testing method	Testing kit and device	Testing targets and controls
19	Kohmer 2021 [74]	Germany	November 2020 - April 2021	Prospective cohort	Adults	Symptomatic and asymptomatic (mostly symptomatic)	Outpatients only Lab-confirmed COVID-19 patients included	Nasopharyngeal swab	Water gargle 10 mL of tap water	Gargled for 15 seconds in the throat	Not reported	RT-PCR	**Kit:** RealStar® SARS-CoV-2 RT-PCR Kit 1.0 (Altona Diagnostics GmbH, Hamburg, Germany) **Device:** ABI PRISM® 7500 Analyser (Applied Biosystems, Waltham, MA, USA)	**Target(s):** E and S **Control(s):** Manufacturer internal control
20	LeBlanc 2021a [94]	Canada	August - November 2020	Prospective cohort	Adults	Symptomatic	Lab-confirmed COVID-19 patients included	Nasal (anterior nares) and oropharyngeal swab Foam swab (BD ProbeTec Qx Collection Kit for Endocervical or Lesion Specimens, Becton, Dickinson and Company, Sparks MD USA) with 3 mL of 1× phosphate-buffered saline with pH 7.4 (Gibco, ThermoFisher Scientific)	Saline gargle 5 mL of 0.9% saline (Addipak, Teleflex, Morrisville, North Carolina, USA)	Performed three alternating cycles of swishing the saline in their cheeks (5 seconds each cycle) and gargling in the posterior oropharynx (5 seconds each cycle)	Patients were asked to abstain from eating, drinking, smoking, chewing gum, and brushing their teeth for at least 1 hour prior to collection	RT-PCR	**Kit:** Taqman virus-1 FAST kit **Device:** ABI 7500 Fast	**Target(s):** E and RdRP **Control(s):** None
21	LeBlanc 2021b [94]	Canada	August – November 2020	Prospective cohort	Adults	Symptomatic	Lab-confirmed COVID-19 patients included	Nasal (anterior nares) and oropharyngeal swab Foam swab (BD ProbeTec Qx Collection Kit for Endocervical or Lesion Specimens, Becton, Dickinson and Company, Sparks MD USA) with 3 mL of 1× phosphate-buffered saline with pH 7.4 (Gibco, ThermoFisher Scientific)	Saline gargle 5 mL of 0.9% saline (Addipak, Teleflex, Morrisville, North Carolina, USA)	Performed three alternating cycles of swishing the saline in their cheeks (5 seconds each cycle) and gargling in the posterior oropharynx (5 seconds each cycle)	Patients were asked to abstain from eating, drinking, smoking, chewing gum, and brushing their teeth for at least 1 hour prior to collection	RT-PCR	**Kit:** SARS-CoV-2 Test (Roche Diagnostics, Rotkreuz, Switzerland) **Device:** Cobas 6800 System (Roche Diagnostics, Rotkreuz, Switzerland)	**Target(s):** E and ORF1a/b **Control(s):** Manufacturer internal control
22	LeBlanc 2021c [94]	Canada	August - November 2020	Prospective cohort	Adults	Symptomatic	Lab-confirmed COVID-19 patients included	Nasal (anterior nares) and oropharyngeal swab Foam swab (BD ProbeTec Qx Collection Kit for Endocervical or Lesion Specimens, Becton, Dickinson and Company, Sparks MD USA) with 3 mL of 1× phosphate-buffered saline with pH 7.4 (Gibco, ThermoFisher Scientific)	Saline gargle 5 mL of 0.9% saline (Addipak, Teleflex, Morrisville, North Carolina, USA)	Performed three alternating cycles of swishing the saline in their cheeks (5 seconds each cycle) and gargling in the posterior oropharynx (5 seconds each cycle)	Patients were asked to abstain from eating, drinking, smoking, chewing gum, and brushing their teeth for at least 1 hour prior to collection	RT-PCR	**Kit:** Aptima SARS-CoV-2 Assay (Hologic Inc., San Diego, CA) **Device:** Panther System (Hologic Inc., San Diego, CA)	**Target(s):** ORF1a/b **Control(s):** Manufacturer internal control
23	Malecki 2021 [95]	Germany	March - April 2020	Prospective cohort	Adults	Symptomatic and asymptomatic	Outpatients only HCWs included	Oropharyngeal swab	Saline gargle 10 mL of normal saline	Gargling time 10–30 seconds	Not reported	RT-PCR	**Kit:** RealStar SARS-CoV-2 RT-PCR Kit (Altona Diagnostics, Germany) **Device:** Not reported	**Target(s):** E and S **Control(s):** Manufacturer internal control

	Study ID [Reference]	Country	Study duration	Study design	Age group	Presence of symptoms	Setting and participants	Reference swab sample type	Gargle sample type	Gargle collection method	Wait time requirement prior to gargle collection	Testing method	Testing kit and device	Testing targets and controls
24	McLennan 2022 [81]	United Kingdom	November 2020	Prospective cohort	Adults and children (> 5 years old)	Symptomatic	Outpatients only HCWs included	Nasopharyngeal and oropharyngeal swab With viral transport media	Saline gargle 10 mL of 0.9% saline	Gargle for 20 seconds	Not eaten, had a drink, smoked, chewed gum, or brushed their teeth within the 30 minute period preceding the test.	RT-PCR	**Kit:** Real-Star SARS-CoV-2 RT-PCR Kit (Altona-Diagnostics) **Device:** ABI 7500FAST Dx instrument	**Target(s):** E and S **Control(s):** Manufacturer internal control
25	Mittal 2020 [98]	India	May - June 2020	Prospective cohort	Adults	Symptomatic and asymptomatic	Inpatients only Lab-confirmed COVID-19 patients included	Nasopharyngeal and oropharyngeal swab Nylon flocked swab with 2 mL of normal saline	Saline gargle 8-10 mL of normal saline	Gargle for 15-20 seconds	Not reported	RT-PCR	**Kit:** SOLIScript 1-step Probe Kit (Solis BioDyne, Newmarket Scientific, UK) with TaqMan reagents **Device:** Not reported	**Target(s):** N1 and N2 **Control(s):** Manufacturer internal control
26	Olearo 2021 [99]	Germany	Not reported	Prospective cohort	Adults	Symptomatic and asymptomatic	Outpatients only Lab-confirmed COVID-19 patients included HCWs included	Nasopharyngeal swab Flocked swab with universal transport media (Copan, Italy, Brescia)	Water gargle 5-7 mL of tap water	Gargle for 30 seconds	None	RT-PCR	**Kit:** Cobas SARS-CoV-2 IVD test **Device:** Cobas 6800 system (Roche, Mannheim, Germany)	**Target(s):** E and ORF1a/b **Control(s):** Manufacturer internal control
27	Poukka 2021 [79]	Finland	June – December 2020	Prospective cohort	Adults and children (≥ 8 years old)	Symptomatic and asymptomatic	Outpatients only Lab-confirmed COVID-19 patients included	Nasopharyngeal swab Nylon swab with 3 mL of Copan UTM	Water gargle Sip of water	Gargle for 5-20 seconds	None	RT-PCR	**Kit:** qScript XLT one-step RT-quantitative PCR (qPCR) ToughMix (Quantabio), primers and probes were based on the Corman E gene primer/probe set **Device:** CFX thermal cycler (Bio-Rad)	**Target(s):** E **Control(s):** Not reported
28	Sancak 2022a* [96]	Turkey	Not reported	Prospective cohort	Not reported	Symptomatic	Lab-confirmed COVID-19 patients included	Nasopharyngeal swab	Water gargle 20 mL of regular drinking water	Gargle and rigorously rinse their mouth forcefully for at least 10 seconds	Not reported	RT-PCR	**Kit:** Commercial PCR kits (Bioeksen and A1 Lifesciences, İstanbul, Turkey) **Device:** Not reported *Gargle sample unconcentrated	**Target(s):** Not reported **Control(s):** Not reported
29	Sancak 2022b* [96]	Turkey	Not reported	Prospective cohort	Not reported	Symptomatic	Lab-confirmed COVID-19 patients included	Nasopharyngeal swab	Water gargle 20 mL of regular drinking water	Gargle and rigorously rinse their mouth forcefully for at least 10 seconds	Not reported	RT-PCR	**Kit:** Commercial PCR kits (Bioeksen and A1 Lifesciences, İstanbul, Turkey) **Device:** Not reported *Gargle sample concentrated with MyMagiCon-RW100® (Bio-T, Istanbul, Turkey)	**Target(s):** Not reported **Control(s):** Not reported
30	Utama 2022a [45]	Indonesia	March - July 2021	Prospective cohort	Adults	Symptomatic and asymptomatic	Inpatients only Lab-confirmed COVID-19 patients included	Nasopharyngeal and oropharyngeal swab With viral transport media	Pre-made solution gargle 2.5 mL of "Gargle Solution"	Patients were required to deeply inhale 5 to 6 times, throat cough 5 to 6 times while having their masks on, and gargle the provided solution	Satisfy a 45-min fasting period during which they were not allowed to eat, drink, smoke, brush their teeth, and use mouthwash; No dental procedures 24 hours prior to sample collection.	RT-PCR	**Kit:** mBioCoV19 Multiplex qRT-PCR Diagnostic Kit (PT Biofarma, Bandung, Indonesia) **Device:** LightCycler® 480 Instrument (Roche Life Science, Penzberg, Germany)	**Target(s):** Helicase and RdRP **Control(s):** Human RNaseP

	Study ID [Reference]	Country	Study duration	Study design	Age group	Presence of symptoms	Setting and participants	Reference swab sample type	Gargle sample type	Gargle collection method	Wait time requirement prior to gargle collection	Testing method	Testing kit and device	Testing targets and controls
31	Utama 2022b [45]	Indonesia	March - July 2021	Prospective cohort	Adults and children (child age range not reported)	Symptomatic and asymptomatic	Outpatients only	Nasopharyngeal and oropharyngeal swab With viral transport media	Pre-made solution gargle 2.5 mL of "Gargle Solution"	Patients were required to deeply inhale 5 to 6 times, throat cough 5 to 6 times while having their masks on, and gargle the provided solution	Satisfy a 45-min fasting period during which they were not allowed to eat, drink, smoke, brush their teeth, and use mouthwash; No dental procedures 24 hours prior to sample collection.	RT-PCR	**Kit:** mBioCoV19 Multiplex qRT-PCR Diagnostic Kit (PT Biofarma, Bandung, Indonesia) **Device:** LightCycler® 480 Instrument (Roche Life Science, Penzberg, Germany)	**Target(s):** Helicase and RdRP **Control(s):** Human RNaseP
32	Zander 2021 [97]	Germany	October - December 2020	Prospective cohort	Adults and children (≥ 13 years old)	Symptomatic and asymptomatic (mostly symptomatic)	Outpatients only	Nasopharyngeal swab Nylon flocked swab (Copan eSwab, MAST Group) with 1 mL of Amies preservation medium	Saline gargle 5 mL of 0.9% saline	30 seconds gargling	None	RT-PCR	**Kit:** Rida Gene SARS-CoV-2 assay (R-Biopharm) **Device:** CFX96-Dx Cycler (BioRad)	**Target(s):** E **Control(s):** Manufacturer internal control

* Data needed to calculate a study estimate unavailable, excluded from meta-analysis