Cochrane Library

Cochrane Database of Systematic Reviews

Thoracic imaging tests for the diagnosis of COVID-19 (Review)

Islam N, Ebrahimzadeh S, Salameh JP, Kazi S, Fabiano N, Treanor L, Absi M, Hallgrimson Z, Leeflang MMG, Hooft L, van der Pol CB, Prager R, Hare SS, Dennie C, Spijker R, Deeks JJ, Dinnes J, Jenniskens K, Korevaar DA, Cohen JF, Van den Bruel A, Takwoingi Y, van de Wijgert J, Damen JAAG, Wang J, McInnes MDF, Cochrane COVID-19 Diagnostic Test Accuracy Group

Islam N, Ebrahimzadeh S, Salameh J-P, Kazi S, Fabiano N, Treanor L, Absi M, Hallgrimson Z, Leeflang MMG, Hooft L, van der Pol CB, Prager R, Hare SS, Dennie C, Spijker R, Deeks JJ, Dinnes J, Jenniskens K, Korevaar DA, Cohen JF, Van den Bruel A, Takwoingi Y, van de Wijgert J, Damen JAAG, Wang J, McInnes MDF, Cochrane COVID-19 Diagnostic Test Accuracy Group. Thoracic imaging tests for the diagnosis of COVID-19.
Cochrane Database of Systematic Reviews 2021, Issue 3. Art. No.: CD013639.
DOI: 10.1002/14651858.CD013639.pub4.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Review]

Thoracic imaging tests for the diagnosis of COVID-19

Nayaar Islam[1,2], Sanam Ebrahimzadeh[1], Jean-Paul Salameh[1], Sakib Kazi[1], Nicholas Fabiano[3], Lee Treanor[3], Marissa Absi[3], Zachary Hallgrimson[3], Mariska MG Leeflang[4], Lotty Hooft[5], Christian B van der Pol[6], Ross Prager[7], Samanjit S Hare[8], Carole Dennie[1,9], René Spijker[5,10], Jonathan J Deeks[11,12], Jacqueline Dinnes[11,13], Kevin Jenniskens[14], Daniël A Korevaar[15], Jérémie F Cohen[16], Ann Van den Bruel[17], Yemisi Takwoingi[11,12], Janneke van de Wijgert[18,19], Johanna AAG Damen[14], Junfeng Wang[20], Matthew DF McInnes[2,3], Cochrane COVID-19 Diagnostic Test Accuracy Group[11]

[1]Department of Radiology , University of Ottawa, Ottawa, Canada. [2]Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada. [3]Department of Radiology, University of Ottawa, Ottawa, Canada. [4]Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands. [5]Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht , Netherlands. [6]Department of Radiology , McMaster University, Hamilton, Canada. [7]Department of Medicine, University of Ottawa , Ottawa, Canada. [8]Department of Radiology , Royal Free London NHS Trust, London , UK. [9]Department of Medical Imaging, The Ottawa Hospital, Ottawa, Canada. [10]Medical Library, Amsterdam UMC, University of Amsterdam, Amsterdam Public Health, Amsterdam, Netherlands. [11]NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham, Birmingham, UK. [12]Test Evaluation Research Group, Institute of Applied Health Research, University of Birmingham, Birmingham, UK. [13]Test Evaluation Research Group, Institute of Applied Health Research, University of Birmingham, Birmingham , UK. [14]Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands. [15]Department of Respiratory Medicine, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands. [16]Obstetrical, Perinatal and Pediatric Epidemiology Research Team (EPOPé), Centre of Research in Epidemiology and Statistics (CRESS), UMR1153, Université de Paris, Paris, France. [17]Academic of Primary Care , KU Leuven, Leuven, Belgium. [18]Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands. [19]Institute of Infection, Veterinary, and Ecological Sciences, University of Liverpool, Liverpool, UK. [20]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Netherlands

Contact address: Matthew DF McInnes, mmcinnes@toh.ca.

Editorial group: Cochrane Infectious Diseases Group.
Publication status and date: New search for studies and content updated (conclusions changed), published in Issue 3, 2021.

Citation: Islam N, Ebrahimzadeh S, Salameh J-P, Kazi S, Fabiano N, Treanor L, Absi M, Hallgrimson Z, Leeflang MMG, Hooft L, van der Pol CB, Prager R, Hare SS, Dennie C, Spijker R, Deeks JJ, Dinnes J, Jenniskens K, Korevaar DA, Cohen JF, Van den Bruel A, Takwoingi Y, van de Wijgert J, Damen JAAG, Wang J, McInnes MDF, Cochrane COVID-19 Diagnostic Test Accuracy Group. Thoracic imaging tests for the diagnosis of COVID-19. *Cochrane Database of Systematic Reviews* 2021, Issue 3. Art. No.: CD013639. DOI: 10.1002/14651858.CD013639.pub4.

ABSTRACT

Background

The respiratory illness caused by SARS-CoV-2 infection continues to present diagnostic challenges. Our 2020 edition of this review showed thoracic (chest) imaging to be sensitive and moderately specific in the diagnosis of coronavirus disease 2019 (COVID-19). In this update, we include new relevant studies, and have removed studies with case-control designs, and those not intended to be diagnostic test accuracy studies.

Objectives

To evaluate the diagnostic accuracy of thoracic imaging (computed tomography (CT), X-ray and ultrasound) in people with suspected COVID-19.

Search methods

We searched the COVID-19 Living Evidence Database from the University of Bern, the Cochrane COVID-19 Study Register, The Stephen B. Thacker CDC Library, and repositories of COVID-19 publications through to 30 September 2020. We did not apply any language restrictions.

Selection criteria

We included studies of all designs, except for case-control, that recruited participants of any age group suspected to have COVID-19 and that reported estimates of test accuracy or provided data from which we could compute estimates.

Data collection and analysis

The review authors independently and in duplicate screened articles, extracted data and assessed risk of bias and applicability concerns using the QUADAS-2 domain-list. We presented the results of estimated sensitivity and specificity using paired forest plots, and we summarised pooled estimates in tables. We used a bivariate meta-analysis model where appropriate. We presented the uncertainty of accuracy estimates using 95% confidence intervals (CIs).

Main results

We included 51 studies with 19,775 participants suspected of having COVID-19, of whom 10,155 (51%) had a final diagnosis of COVID-19. Forty-seven studies evaluated one imaging modality each, and four studies evaluated two imaging modalities each. All studies used RT-PCR as the reference standard for the diagnosis of COVID-19, with 47 studies using only RT-PCR and four studies using a combination of RT-PCR and other criteria (such as clinical signs, imaging tests, positive contacts, and follow-up phone calls) as the reference standard.

Studies were conducted in Europe (33), Asia (13), North America (3) and South America (2); including only adults (26), all ages (21), children only (1), adults over 70 years (1), and unclear (2); in inpatients (2), outpatients (32), and setting unclear (17).

Risk of bias was high or unclear in thirty-two (63%) studies with respect to participant selection, 40 (78%) studies with respect to reference standard, 30 (59%) studies with respect to index test, and 24 (47%) studies with respect to participant flow.

For chest CT (41 studies, 16,133 participants, 8110 (50%) cases), the sensitivity ranged from 56.3% to 100%, and specificity ranged from 25.4% to 97.4%. The pooled sensitivity of chest CT was 87.9% (95% CI 84.6 to 90.6) and the pooled specificity was 80.0% (95% CI 74.9 to 84.3). There was no statistical evidence indicating that reference standard conduct and definition for index test positivity were sources of heterogeneity for CT studies.

Nine chest CT studies (2807 participants, 1139 (41%) cases) used the COVID-19 Reporting and Data System (CO-RADS) scoring system, which has five thresholds to define index test positivity. At a CO-RADS threshold of 5 (7 studies), the sensitivity ranged from 41.5% to 77.9% and the pooled sensitivity was 67.0% (95% CI 56.4 to 76.2); the specificity ranged from 83.5% to 96.2%; and the pooled specificity was 91.3% (95% CI 87.6 to 94.0). At a CO-RADS threshold of 4 (7 studies), the sensitivity ranged from 56.3% to 92.9% and the pooled sensitivity was 83.5% (95% CI 74.4 to 89.7); the specificity ranged from 77.2% to 90.4% and the pooled specificity was 83.6% (95% CI 80.5 to 86.4).

For chest X-ray (9 studies, 3694 participants, 2111 (57%) cases) the sensitivity ranged from 51.9% to 94.4% and specificity ranged from 40.4% to 88.9%. The pooled sensitivity of chest X-ray was 80.6% (95% CI 69.1 to 88.6) and the pooled specificity was 71.5% (95% CI 59.8 to 80.8).

For ultrasound of the lungs (5 studies, 446 participants, 211 (47%) cases) the sensitivity ranged from 68.2% to 96.8% and specificity ranged from 21.3% to 78.9%. The pooled sensitivity of ultrasound was 86.4% (95% CI 72.7 to 93.9) and the pooled specificity was 54.6% (95% CI 35.3 to 72.6).

Based on an indirect comparison using all included studies, chest CT had a higher specificity than ultrasound. For indirect comparisons of chest CT and chest X-ray, or chest X-ray and ultrasound, the data did not show differences in specificity or sensitivity.

Authors' conclusions

Our findings indicate that chest CT is sensitive and moderately specific for the diagnosis of COVID-19. Chest X-ray is moderately sensitive and moderately specific for the diagnosis of COVID-19. Ultrasound is sensitive but not specific for the diagnosis of COVID-19. Thus, chest CT and ultrasound may have more utility for excluding COVID-19 than for differentiating SARS-CoV-2 infection from other causes of respiratory illness.

Future diagnostic accuracy studies should pre-define positive imaging findings, include direct comparisons of the various modalities of interest in the same participant population, and implement improved reporting practices.

PLAIN LANGUAGE SUMMARY

How accurate is chest imaging for diagnosing COVID-19?

Why is this question important?

People with suspected COVID-19 need to know quickly whether they are infected, so they can receive appropriate treatment, self-isolate, and inform close contacts.

Currently, a formal diagnosis of COVID-19 requires a laboratory test (RT-PCR) of nose and throat samples. RT-PCR requires specialist equipment and takes at least 24 hours to produce a result. It is not completely accurate, and may require a second RT-PCR or a different test to confirm diagnosis.

COVID-19 is a respiratory disease. Clinicians may use chest imaging to diagnose people who have COVID-19 symptoms, while awaiting RT-PCR results or when RT-PCR results are negative, and the person has COVID-19 symptoms.

What did we want to find out?

We wanted to know whether chest imaging is accurate enough to diagnose COVID-19 in people with suspected infection. This is the second update of this review; in it, we included studies in people with suspected COVID-19 only; we excluded studies in people with confirmed COVID-19.

The evidence is up to date to 30 September 2020.

What are chest imaging tests?

X-rays or scans produce an image of the organs and structures in the chest.

- X-rays (radiography) use radiation to produce a 2-D image. Usually done in hospitals, using fixed equipment by a radiographer, they can also be done on portable machines.

- Computed tomography (CT) scans use a computer to merge 2-D X-ray images and convert them to a 3-D image. They require highly specialised equipment and are done in hospital by a specialist radiographer.

- Ultrasound scans use high-frequency sound waves to produce an image. They can be done in hospitals or other healthcare settings, such as a doctor's office.

What did we do?

We searched for studies that assessed the accuracy of chest imaging to diagnose COVID-19 in people of any age with suspected COVID-19. Studies could be of any design, except for case control studies, and could take place anywhere.

What did we find?

We found 51 studies with 19,775 participants. Of these people, 10,155 (51%) had a final diagnosis of COVID-19. Forty-seven studies confirmed COVID-19 infection using RT-PCR alone. Four studies used RT-PCR with another test.

Forty-seven studies evaluated one imaging technique each, and four studies evaluated two imaging techniques each.

Chest CT was evaluated by 41 studies (16,133 participants, 8110 (50%) confirmed COVID-19 cases), chest X-ray by nine studies (3694 participants, 2111 (57%) confirmed COVID-19 cases), and ultrasound by five studies (446 participants, 211 (47%) confirmed COVID-19 cases). Thirty-three studies were conducted in Europe, 13 in Asia, three in North America and two in South America. Twenty-six studies included only adults, 21 included both adults and children, one included only children, one included participants aged 70 years and older, and two studies did not report participants' ages. Two studies included hospital inpatients and 32 included hospital outpatients. The setting was unclear in the remaining 17 studies.

Where four or more studies evaluated a particular type of chest imaging, we pooled their results and analysed them together.

Chest CT

Pooled results showed that chest CT correctly diagnosed COVID-19 in 87.9% of people who had COVID-19. However, it incorrectly identified COVID-19 in 20% of people who did not have COVID-19.

Chest X-ray

Pooled results showed that chest X-ray correctly diagnosed COVID-19 in 80.6% of people who had COVID-19. However, it incorrectly identified COVID-19 in 28.5% of people who did not have COVID-19.

Lung ultrasound

Pooled results showed that lung ultrasound correctly diagnosed COVID-19 in 86.4% of people with COVID-19. However, it incorrectly diagnosed COVID-19 in 45% of people who did not have COVID-19.

How reliable are the results?

The studies differed from each other and used different methods to report their results. Few studies evaluated chest X-ray and chest ultrasound; and very few studies directly compared one type of imaging test with another. Therefore, we cannot draw confident conclusions based on results from studies in this review.

What does this mean?

The evidence suggests that chest CT is better at ruling out COVID-19 infection than distinguishing it from other respiratory problems. So, its usefulness may be limited to excluding COVID-19 infection rather than distinguishing it from other causes of lung infection.

We plan to update this review as more evidence becomes available. Future studies should predefine what a positive test is, and compare different types of imaging tests on similar groups of people.

SUMMARY OF FINDINGS

Summary of findings 1. Thoracic imaging tests for the diagnosis of COVID-19

Thoracic imaging tests for the diagnosis of COVID-19	
Question	What is the diagnostic accuracy of chest imaging (computed tomography (CT), chest X-ray and ultrasound) in the evaluation of people suspected of having COVID-19?
Population	Children or adults suspected of having COVID-19
Index test	Chest imaging tests used for the diagnosis of COVID-19, including: • chest CT • chest X-rays • ultrasound of the lungs
Target condition	COVID-19, the illness following acute infection with SARS-CoV-2
Reference standard	A positive diagnosis for COVID-19 by one or a combination of the following. • A positive RT-PCR test for SARS-CoV-2 infection, from any manufacturer in any country, from any source, including nasopharyngeal swabs or aspirates, oropharyngeal swabs, bronchoalveolar lavage fluid (BALF), sputum, saliva, serum, urine, rectal or faecal samples • Positive on WHO criteria for COVID-19, which includes some testing RT-PCR-negative • Positive on China CDC criteria for COVID-19, which includes some testing RT-PCR negative • Positive serology in addition to consistent symptomatology • Positive on study-specific list of criteria for COVID-19, which includes some testing RT-PCR-negative • Other criteria (symptoms, imaging findings, other tests, infected contacts) A negative diagnosis for COVID-19 by one or a combination of the following. • People with suspected COVID-19 with negative RT-PCR test results, whether tested once or more than once. • Currently healthy or with another disease (no RT-PCR test)
Limitations in the evidence	
Risk of bias	• Participant selection: high in 10 (20%) studies and unclear in 22 (43%) studies • Application of index tests – chest CT: high in 5/41 (12%) studies and unclear in 15/41 (37%) studies • Application of index tests – chest X-ray: unclear in 6/9 (67%) studies • Application of index tests – ultrasound of the lungs: unclear in 4/5 (80%) studies • Reference standard: high in 20 (39%) studies and unclear in 20 (39%) studies • Flow and timing: high in 2 (3.9%) studies and unclear in 22 (43%) studies
Concerns about applicability of the evidence	• Participants: high in 1 (2%) study • Index test – chest CT: low in all 41 studies • Index test – chest X-ray: high in 1/9 (11%) study and unclear in 1/9 (11%) study • Index test – ultrasound of the lungs: unclear in 1/5 (20%) study • Reference standard: low in all 51 studies
Findings	
• We included 51 studies (19,775 participants suspected of having COVID-19, 10,155 (51%) cases) • Studies evaluated chest CT scans (41 studies), chest X-ray (9 studies) and ultrasound of the lungs (5 studies)	

- Chest CT was sensitive and moderately specific in the diagnosis of COVID-19 in suspected cases.
- Chest X-ray was moderately sensitive and moderately specific in the diagnosis of COVID-19 in suspected cases.
- Ultrasound of the lungs was sensitive, but not specific in the diagnosis of COVID-19 in suspected cases.
- Sensitivity analysis in chest CT studies showed that publication status had a minimal effect on our findings.
- The 'threshold' effect in chest CT studies that used the CO-RADS scoring system demonstrated a trade-off between sensitivity and specificity; as the threshold for index test positivity increased from 2 to 5, sensitivity decreased, and specificity increased.
- There was no statistical evidence indicating that reference standard conduct and definition of index test positivity were sources of heterogeneity for chest CT studies.
- Indirect test comparisons showed that chest CT has a higher specificity than ultrasound. Chest CT and ultrasound have similar sensitivities, chest CT and chest X-ray have similar sensitivities and specificities, and chest X-ray and ultrasound have similar sensitivities and specificities.

Evidence for participants suspected of having COVID-19

Imaging modality	Sensitivity (95% CI)	Specificity (95% CI)	Number of participants (cases)
Chest CT	87.9% (84.6 to 90.6)	80.0% (74.9 to 84.3)	16,133 (8110)
Chest X-ray	80.6% (69.1 to 88.6)	71.5% (59.8 to 80.8)	3694 (2111)
Ultrasound of the lungs	86.4% (72.7 to 93.9)	54.6% (35.3 to 72.6)	446 (211)

Predicted outcomes

Given various prevalence settings, predicted outcomes for the number of individuals receiving a false positive result or a false negative (missed) result per 1000 people undergoing chest CT, chest X-ray, and ultrasound of the lungs are outlined as follows.

Predicted outcomes per 1000 people undergoing chest CT

Prevalence of COVID-19	Positive CT result n (95% CI)	False positive CT result n (95% CI)	Negative CT result n (95% CI)	False negative CT result n (95% CI)
50%	440 (423 to 453)	100 (79 to 126)	400 (374 to 421)	60 (47 to 77)
20%	176 (169 to 181)	160 (126 to 200)	640 (599 to 674)	24 (19 to 31)
5%	44 (42 to 45)	190 (149 to 238)	760 (712 to 801)	6 (5 to 8)

Predicted outcomes per 1000 people undergoing chest X-ray

Prevalence of COVID-19	Positive CT result n (95% CI)	False positive CT result n (95% CI)	Negative CT result n (95% CI)	False negative CT result n (95% CI)
50%	403 (346 to 443)	143 (96 to 201)	357 (299 to 404)	97 (57 to 154)
20%	161 (138 to 177)	228 (154 to 322)	572 (478 to 646)	39 (23 to 62)
5%	40 (35 to 44)	271 (182 to 382)	679 (568 to 768)	10 (6 to 15)

Predicted outcomes per 1000 people undergoing ultrasound of the lungs

Prevalence of COVID-19	Positive CT result	False positive CT result n (95% CI)	Negative CT result n (95% CI)	False negative CT result n (95% CI)

	n (95% CI)			
50%	432 (364 to 470)	227 (137 to 234)	273 (176 to 363)	68 (30 to 136)
20%	173 (145 to 188)	363 (219 to 518)	437 (282 to 581)	27 (12 to 55)
5%	43 (36 to 47)	431 (260 to 615)	519 (335 to 690)	7 (3 to 14)

CI: confidence interval; **CT:** computed tomography; **n:** number; **RT-PCR:** reverse transcription polymerase chain reaction

BACKGROUND

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection and resulting coronavirus disease 2019 (COVID-19) pandemic continue to present diagnostic evaluation challenges. While the World Health Organization (WHO) reports laboratory confirmation of COVID-19 infection, such as a positive reverse transcriptase polymerase chain reaction (RT-PCR) result as the standard for diagnosing COVID-19, the value of imaging tests in the diagnostic pathway remains undefined (WHO 2020). Research on the role of imaging in COVID-19 patients is evolving and more refined assessment methods for imaging tests, such as the COVID-19 Reporting and Data System (CO-RADS), are being investigated (Prokop 2020).

Decisions about patient and isolation pathways for COVID-19 vary according to health services and settings, available resources, and outbreaks in different settings. They will change over time, as accurate tests, effective treatments, and vaccines are identified. The decision points between these pathways vary, but all include points at which knowledge of the accuracy of diagnostic information is needed to inform medical decisions. Therefore, it is essential to understand the accuracy of tests and diagnostic features to develop effective diagnostic and management pathways for different settings. This supports strategies aiming to identify those who are infected, and consequently the management of patients either through isolation precautions, contact tracing, quarantine, hospital admission or admission to a specialised facility, admission to the intensive care unit, or initiation of specific therapies, and implementation of mitigation strategies to limit the spread of the disease.

This review from the suite of Cochrane 'living systematic reviews' summarises evidence on the accuracy of different imaging tests and diagnostic features in participants regardless of their symptoms. Estimates of accuracy from this review will help inform diagnostic, screening, isolation, and patient-management decisions. We have included an explanation of terminology and acronyms in Appendix 1.

Target condition being diagnosed

The target condition being evaluated is COVID-19, the illness following acute infection with SARS-CoV-2 (Datta 2020). People infected with SARS-CoV-2 can be asymptomatic; these people are not considered to have COVID-19 and thus not within the scope of this review. People with COVID-19 can have a wide variety of symptoms, including fever, sore throat, diarrhea, dyspnoea, headache, chest pain, stomach ache, nausea, loss of taste, loss of smell, myalgia, fatigue, runny nose, cough, aches, and lethargy (either without difficulty breathing at rest or with shortness of breath and increased respiratory rate potentially requiring supplemental oxygen or mechanical ventilation). Furthermore, in people diagnosed with a pulmonary condition (e.g. pulmonary embolism), symptoms could be indicative of COVID-19, or could be a manifestation of the pre-existing condition. In this review, we focused on people suspected of having COVID-19 who had thoracic imaging as part of their evaluation or care.

Index test(s)

Chest computed tomography (CT)

Chest CT refers to the acquisition of images of the chest using computed tomography. Typical imaging protocols would not use intravenous (IV) contrast; however, in this review we considered all variations of imaging protocols with the exception of studies specifically targeted at evaluating the coronary arteries or the heart, which did not include the entire lungs in the field of view. This includes, but is not limited to, non-contrast chest CT, low-dose chest CT (with or without contrast), high-resolution chest CT, and chest CT with IV contrast (routine or pulmonary angiogram).

Chest radiographs/chest X-rays

Chest radiography refers to the evaluation of the lungs using X-rays. This often involves two orthogonal views, posterior-anterior (PA) and lateral, but may be done by a portable machine and only acquire an anterior-posterior (AP) view. In this review, we considered any and all variations of chest radiography protocols that evaluated the lungs. We did not include protocols that did not include the entire thorax and were done for reasons other than for assessment of pulmonary status (e.g. assessment of feeding tube position, which typically only includes the lower thorax, or dedicated evaluation of the ribs).

Ultrasound of the lungs

Ultrasound of the lungs refers to any ultrasound of the thorax done with the intention of evaluating the status of the lungs. This includes, but is not limited to, point-of-care ultrasound, done at the bedside by a physician, as well as what is often termed 'consultative' ultrasound, which is done by a technologist and subsequently interpreted by a physician (typically a radiologist).

We considered all possible technical parameters (e.g. type of probe, transducer frequency, use of contrast). This did not include ultrasound done with the intended purpose of evaluating only the heart or vessels of the chest.

Clinical pathway

At present, the optimal diagnostic pathway and the role of thoracic imaging for identifying people with COVID-19 is unclear. Compared to RT-PCR testing, a potential major advantage of thoracic imaging is that results are available faster and that it provides a better insight into the status of the lungs. However, chest CT imaging is typically only available in secondary and tertiary healthcare settings, and availability varies across these settings.

Role of index test(s)

1. Thoracic imaging may play an integral role in 'ruling out' COVID-19 pneumonia when RT-PCR is unavailable, pending or negative, or when clinical suspicion is 'low' based on other signs, symptoms and routine laboratory tests. Role of test: triage for RT-PCR, to make decisions about performing or not performing RT-PCR or other diagnostic tests.
2. Rapid testing - thoracic imaging is used to rule in or rule out COVID-19 when results from other tests (e.g. RT-PCR) are not available in a timely manner.
3. Concurrent/combination testing with other diagnostic tests (as part of a pair or group of tests) to improve the accuracy of diagnosis. For example, thoracic imaging could be used to

identify false negatives of other tests (e.g. RT-PCR), and to improve the overall accuracy of the testing strategy.

Several diagnostic pathways have been proposed that provide guidance for physicians to identify people with COVID-19. The order and components of these pathways differ with varying dependence on pre-test probability, physical examination, laboratory tests and findings based on RT-PCR results and availability. However, some professional organisations recommend imaging for patients with moderate or severe features of COVID-19 (Rubin 2020). In some hospitals, the results of low-dose chest CT are one of the many parameters (among molecular test results, routine laboratory results and clinical signs and symptoms) used to categorise patients as low risk, moderate to high risk, and proven COVID-19 cases (China National Health Comission 2020).

Given the rapid progression of COVID-19 and the constantly evolving evidence base, the diagnostic accuracy to inform the utility of thoracic imaging in these pathways is difficult to estimate. This 'living systematic review' aims to identify data regarding the diagnostic accuracy of thoracic imaging in people with suspected COVID-19. This represents our second update of this 'living systematic review' (Islam 2020).

Alternative test(s)

Other Cochrane diagnostic test accuracy (DTA) reviews in the suite of reviews address the following tests.

1. Signs and symptoms, which will be mainly used in primary care, including when presenting at the emergency department (Struyf 2020)
2. Routine laboratory testing, such as for C-reactive protein (CRP) and procalcitonin (PCT) (Stegeman 2020)
3. Antibody tests (Deeks 2020)
4. Laboratory-independent point-of-care and near-patient molecular and antigen tests (Dinnes 2020)
5. Molecular laboratory tests

Summary of previous versions of the review

In our initial review, studies that only included confirmed cases of COVID-19 reported high pooled sensitivities for chest CT and X-ray: 93.1% (95% CI 90.2 to 95.0) and 82.1% (95% CI 62.5 to 92.7), respectively (Salameh 2020a). Thirteen studies that assessed chest CT in participants with suspected COVID-19 demonstrated a sensitivity of 86.2% (95% CI 71.9 to 93.8) but a low specificity of 18.1% (95% CI 3.71 to 55.8). This indicated a lack of discrimination, as the chances of getting a positive chest CT result are 86% in patients with a SARS-CoV-2 infection and 82% in patients without. We did not evaluate accuracy estimates for chest X-ray and ultrasound of the lungs in participants with suspected COVID-19 in the initial review as these data were not available.

The first update of this review focused on people suspected of having COVID-19 and excluded studies evaluating only confirmed cases of COVID-19 (Islam 2020). Thirty-one studies that evaluated chest CT in suspected participants demonstrated a pooled sensitivity of 89.9% (95% CI 85.7 to 92.9) and a pooled specificity of 61.1% (95% CI 42.3 to 77.1). This indicated that chest CT performs well in identifying COVID-19, but may have limited capability in differentiating SARS-CoV-2 infection from other causes of respiratory illness. We did not identify publication status as a source of variability for accuracy estimates of chest CT, and further investigations of additional variables were not possible due to limited data. We were not able to evaluate pooled accuracy estimates for chest X-ray and ultrasound of the lungs in participants with suspected COVID-19 in the first update of this review due to limited data. We did explore the value of formal scoring systems for the evaluation of index tests, and 'threshold' effects of index test positivity, however, we could not perform formal analyses due to the limited number of included studies.

Compared to the first update, this second update has stricter inclusion criteria, excluding studies of case-control design and those that report an overview of index test findings without explicitly classifying the imaging test as either COVID-19 positive or negative. We included more studies in this update and we evaluated both chest X-ray and ultrasound of the lungs in addition to chest CT. Furthermore, this update formally assesses the value of formal scoring systems, 'threshold' effects, and time trends of accuracy estimates of chest CT, as well as indirect comparisons with respect to accuracy of imaging modalities (i.e. chest CT, X-ray and ultrasound).

Changes in the evidence base since previous versions

Evolving research on imaging tests in COVID-19 patients includes the use of formal scoring systems to evaluate imaging tests, which offer the potential for improved specificity. Formal scoring systems include CO-RADS (Prokop 2020), the British Society of Thoracic Imaging (BSTI) COVID-19 Reporting Templates (BSTI 2020), and the Radiological Society of North America (RSNA) Expert Consensus on Reporting Chest CT Findings for COVID-19 (Simpson 2020). In the initial version of this review, most studies either did not specify what criteria were used for index test positivity, or used 'any abnormality' to define index test positive. In the first update of this review, we explored the value of formal scoring systems but we could not formally analyse them due to a limited number of studies that used these systems. In this update, as well as in future updates of this review, we will evaluate the value of formal scoring systems and the impact of 'threshold' effects of index test positivity on accuracy estimates of imaging tests (Irwig 1995).

OBJECTIVES

To evaluate the diagnostic accuracy of thoracic imaging (computed tomography (CT), chest X-ray and ultrasound) in the evaluation of people with suspected COVID-19.

Secondary objectives

To evaluate 'threshold' effects of index test positivity on accuracy.

METHODS

Criteria for considering studies for this review

Types of studies

We kept the eligibility criteria broad to be able to include all settings and all variations of a test. We included studies of all designs, with the exception of case-control studies. Studies had to include participants suspected of having the target condition and produce estimates of test accuracy or provide 2x2 data (true positive (TP), true negative (TN), false positive (FP), false negative (FN)) from which we could compute estimates for the primary objective. If data were not available, we contacted study authors for additional

data if the study met the primary objective only. Studies with fewer than 10 participants who underwent the index test and reference standard were excluded.

Participants

Our focus was on studies that recruited participants suspected of having COVID-19 as outlined in the Target condition being diagnosed section. We included studies with 'symptomatic populations' or 'mixed populations' (asymptomatic and symptomatic participants). There were no age or gender restrictions.

Index tests

The index tests were chest CT, chest X-ray, or ultrasound of the lungs, meeting the criteria described in the Index test(s) section. The roles of the test could have been a replacement of RT-PCR, an add-on test, a triage test, rapid testing, or used concurrently with other diagnostic tests.

We included only index tests interpreted by humans, and not an algorithm (machine learning/artificial intelligence (AI)). We included studies involving interpretation by an algorithm only if they provided data pertaining to diagnostic accuracy of human interpretation.

Definitions of imaging test positivity

Inclusion was limited to 'diagnostic test accuracy studies' in which the study authors explicitly indicated that the index test aims to distinguish between patients with and without COVID-19 were included. Specifically, studies with index test readers either (1) using a radiological scoring system (e.g. CO-RADS), or (2) explicitly classifying patients as having a positive or negative imaging test were included. Studies that reported an overview of index test findings without explicitly classifying the imaging test as either COVID-19 positive or negative were excluded.

Since COVID-19 is such a new disease, and the imaging findings were unknown until recently, there is considerable heterogeneity and change in the definitions used for positivity. Some groups have used constellations of specific findings (such as multiple peripheral ground-glass opacities on CT), some have used an approach in which they consider the combined effect of specific findings (a 'gestalt' approach), and some have used formal scoring systems, such as CO-RADS (5 categories; Prokop 2020), the BSTI COVID-19 Reporting Templates (4 categories; BSTI 2020), and the RSNA Expert Consensus on Reporting Chest CT Findings for COVID-19 (4 categories; Simpson 2020). As such, we did not limit ourselves to a predefined definition or threshold for positivity. Instead, we extracted the definition for positivity used in each study, and the constellation of imaging features used to inform this definition. This offers an opportunity to determine if the definition of positivity contributes to variability in accuracy.

Target conditions

As explained above, our target condition is COVID-19. However, we included all studies reporting data on COVID-19 or COVID-19 pneumonia that might provide data relevant to our objective.

Reference standards

A positive diagnosis for COVID-19 by one or a combination of the following:

1. a positive RT-PCR test for SARS-CoV-2 infection, from any manufacturer in any country, and from any sample type, including nasopharyngeal swabs or aspirates, oropharyngeal swabs, bronchoalveolar lavage fluid, sputum, saliva, serum, urine, rectal or faecal samples;
2. positive on WHO criteria for COVID-19;
3. positive on China CDC criteria for COVID-19;
4. positive serology for SARS-CoV-2 antibodies in addition to consistent symptomatology;
5. positive on study-specific list of criteria for COVID-19 which includes:
    a. other criteria (symptoms, imaging findings, other tests, infected contacts).

A negative diagnosis for COVID-19 by one or a combination of the following:

1. suspected COVID-19 with negative RT-PCR test results, whether tested once or more than once;
2. currently healthy or with another disease (no RT-PCR test).

We assessed methodological quality based on our judgement of how likely it was that the reference standard definition used in each study would correctly classify individuals as positive or negative for COVID-19. All reference standards are likely to be imperfect in some way; details of reference standard evaluation are provided in Appendix 2. We used a consensus process to agree on the classification of the reference standard as to what we regarded as good, moderate and poor. 'Good' reference standards need to have very little chance of misclassification; 'moderate', a small but acceptable risk; and 'poor', a larger and probably unacceptable risk.

Search methods for identification of studies

Electronic searches

We used three different sources for our electronic searches through 30 September 2020, which were devised with the help of an experienced Cochrane Information Specialist with DTA expertise (RSp). These searches aimed to identify all articles related to COVID-19 and SARS-CoV-2 and were not restricted to those evaluating imaging tests. Thus, the searches used no terms that specifically focused on an index test, diagnostic accuracy or study methodology.

Due to the increased volume of published and preprint articles, we used artificial intelligence text analysis from 25 May 2020 and onwards to conduct an initial classification of documents, based on their title and abstract information, for relevant and irrelevant documents. Appendix 3.

1. Living search from the University of Bern

We used the COVID-19 living search results of the Institute of Social and Preventive Medicine (ISPM) at the University of Bern. This search includes PubMed, Embase and preprints indexed in bioRxiv and medRxiv databases. The strategies as described on the ISPM website (ispmbern.github.io/covid-19), are shown in Appendix 4.

2. Cochrane COVID-19 Study Register searches

We also included searches undertaken by Cochrane to develop the Cochrane COVID-19 Study Register. These include searches of trials registers at ClinicalTrials.gov and the World Health Organization

International Clinical Trials Registry Platform (WHO ICTRP), as well as PubMed (see Appendix 4 for details). Search strategies were designed for maximum sensitivity, to retrieve all human studies on COVID-19. We did not apply any language limits.

3. The Stephen B. Thacker CDC Library, COVID-19 Research Articles Downloadable Database

We included Embase records within the CDC library on COVID-19 research articles database (see Appendix 4 for details) and deduplicated these against the Cochrane COVID-19 Study Register.

Searching other resources

We checked repositories of COVID-19 publications against these search results including the following.

1. EPPI centre eppi.ioe.ac.uk/COVID19_MAP/covid_map_v4.html
2. The Norwegian Institute of Public Health 'NIPH systematic and living map on COVID-19 evidence www.nornesk.no/forskningskart/NIPH_diagnosisMap.html
3. From these websites we searched company and product websites for studies about test accuracy.
4. We contacted companies to ask for further information about studies.
5. We also contacted research groups that we were made aware of who are completing test evaluations (e.g. UK Public Health England-funded studies, Foundation for Innovative New Diagnostics (FIND) studies).

Data collection and analysis

Selection of studies

The review authors screened studies independently, in duplicate. A third, experienced review author resolved disagreements about initial title and abstract screening. We resolved disagreements about eligibility assessments through discussion between three review authors.

Data extraction and management

The review authors performed data extraction independently, in duplicate. Three review authors discussed any disagreements to resolve them.

For each study, we extracted 2x2 contingency tables of the number of true positives, false positives, false negatives and true negatives. If a study reported accuracy data for more than one index test reader, we took the average of the data from all readers to compute the average 2x2 contingency table (McGrath 2017). If a study reported accuracy data for both an AI algorithm and one or more radiologists, we extracted only the 2x2 contingency table corresponding to the radiologist accuracy data. If a study used multiple reference standards, but we could determine 2x2 contingency tables that included only RT-PCR as the reference standard, we extracted and analysed these data. If a study reported accuracy data for multiple thresholds of index test positivity (e.g. studies that used the CO-RADS scoring system), we extracted the 2x2 contingency table for all available thresholds.

Three of the nine studies that used the CO-RADS scoring system did not report the 2x2 data for all five CO-RADS thresholds. We contacted the corresponding authors and successfully received the complete data for one of the three studies. For the two remaining studies, we were only able to extract data for a CO-RADS threshold of 3.

In addition, we extracted the following items.

1. Study setting (including country), age of study participants, study dates, disease prevalence at the time of acquisition (as reported in the study), number of participants, participant symptoms, number of imaging studies (and if more than one study was done per participant), participant outcomes and other relevant participant demographic parameters.
2. Study design
3. Imaging timing relative to disease course
4. CT, chest X-ray and ultrasound findings
5. Criteria for 'positive' diagnosis of COVID-19 on imaging
6. Index test technical parameters
7. Reference standard results and details. If RT-PCR was performed, timing of test, number of tests and method of acquisition (or similar details regarding other reference standards used).
8. Details regarding interpretation of the index test (level of training, number of readers, the inter-observer variability)
9. The number of true positives, false positives, false negatives and true negatives or summary statistics from which they can be computed
10. Participant co-morbidities as described in the studies

Assessment of methodological quality

The review authors assessed the risk of bias and applicability concerns independently, in duplicate, using the QUADAS-2 domain-list. Three review authors resolved any disagreements through discussion. See Appendix 2 for an explanation of the operationalisation of the four QUADAS-2 domains – participant selection, index test(s), reference standard(s), flow and timing.

Statistical analysis and data synthesis

We presented estimates of sensitivity and specificity using paired forest plots and we summarised pooled estimates in tables. We analysed the data on a participant level, not a lesion or lung-segment level, since this is what determines care.

We used a bivariate model for meta-analyses, taking into account the within- and between-study variance, and the correlation between sensitivity and specificity across studies (Chu 2006; Reitsma 2005). We performed meta-analyses when four or more studies evaluated a given modality. We also performed sensitivity analyses by limiting inclusion in the meta-analysis to studies published in peer-reviewed journals. We undertook meta-analyses using metandi in STATA (Harbord 2009; StataCorp 2019).

If a study reported accuracy data at multiple thresholds of index test positivity, we used the 2x2 contingency table corresponding to the threshold producing the highest Youden's Index (YI) (YI = sensitivity + specificity – 1) for inclusion in the meta-analysis.

Investigations of heterogeneity

We investigated heterogeneity by visual inspection of paired forest plots and summary receiver operating characteristics (SROC) plots. For chest CT studies, we evaluated the impact of definition for reference standard conduct (RT-PCR performed at least twice in all participants with initial negative results versus RT-PCR performed

only once in all participants with initial negative results or RT-PCR performed twice in some but not all participants with initial negative results), and definition for index test positivity (radiologist impression versus formal scoring system) on accuracy estimates using meta-regression with the variable of interest added as a covariate to a bivariate model. Using the model parameters, we used a postestimation command to compute absolute differences in pooled sensitivity and specificity and we obtained their 95% CIs using the delta method. We obtained P values using the Wald test. We performed meta-regression when variables of interest consisted of subgroups with five or more studies in each subgroup, an arbitrary threshold chosen to facilitate convergence of the analyses using the bivariate model. We undertook meta-regression using meqrlogit in STATA (StataCorp 2019).

If a study within a subgroup reported accuracy data at multiple thresholds of index test positivity, we used the 2x2 contingency table corresponding to the threshold producing the highest YI for inclusion in the meta-regression.

Threshold effects

We performed meta-analyses using a bivariate model for studies that used common thresholds for test positivity (i.e. chest CT studies at CO-RADS thresholds 2, 3, 4 and 5).

We used ggplot2 and ggforce in R to generate a plot displaying pooled accuracy estimates at varying CO-RADS thresholds (Wickham 2016; Pedersen 2020; R Core Team 2021).

Direct and indirect test comparisons

We undertook comparisons of test accuracy between imaging modalities, regardless of whether or not studies compared imaging modalities head-to-head in the same study population (i.e. indirect comparison). We performed this using meta-regression with modality type (i.e. chest CT, chest X-ray, and ultrasound of the lungs) added as a covariate to a bivariate model. We obtained P values using the Wald test. In future updates, as more data become available, we will also perform test comparisons that are restricted to only comparative studies (i.e. direct comparisons).

Cumulative meta-analyses and time trends

For chest CT studies, we performed univariate, cumulative, random-effects meta-analyses of sensitivity and specificity with logit-transforms using STATA (StataCorp 2019). We incorporated primary studies sequentially in the meta-analysis, according to their rank with respect to publication date, to iteratively recalculate summary estimates of logit sensitivity, logit specificity and their variances (Cohen 2016; Lau 1992). Then we assessed time trends by fitting a weighted linear regression model in which the summary estimate up to and including a given primary study is modeled as a linear function of rank of publication, with a first-order autoregressive process to account for the correlation between successive estimates. We generated plots displaying changes in cumulative logit sensitivity and cumulative logit specificity over time. The above cumulative meta-analyses were restricted to the chest CT studies included in the current review.

We also generated a plot displaying meta-analysis results across all versions of this review (i.e. pooled sensitivity and specificity estimates from the initial version published in September 2020 (Salameh 2020a), the first update published in November 2020 (Islam 2020), and this current update) using ggplot2 and ggforce in R (Wickham 2016; Pedersen 2020; R Core Team 2021).

Assessment of reporting bias

For this review, we did not undertake tests for publication bias and made no formal assessment of reporting bias.

Summary of findings

We provided a summary of the key findings of this review in Summary of findings 1, indicating the strength of evidence for each finding and emphasising the main gaps in our current level of available evidence.

Updating

The prior version of this review contained studies up to 22 June 2020 (Islam 2020). This updated review contains the results of an updated search performed on 30 September 2020. With the substantial number of studies published since 30 September 2020, we plan to update this review including studies up to February 2021.

RESULTS

Results of the search

We identified 4734 search results and imported 782 studies for screening. Subsequently, we removed nine duplicates. We then screened a total of 773 unique references (published or preprint studies) for inclusion; this is inclusive of the 668 references we screened in our previous reviews. Of the 358 records selected for full-text assessment, we included 51 studies in this review; of these 51 included studies, four have been included since our initial review (Salameh 2020a), and 12 have been included since the first update of this review (Islam 2020). Refer to Figure 1 for the PRISMA flow diagram of search and inclusion results (Salameh 2020b; Moher 2009). Exclusions were mainly due to ineligible study design (17 studies), ineligible study outcomes (9 studies), or ineligible patient populations (6 studies); see Figure 1.

Figure 1. Study flow diagram.

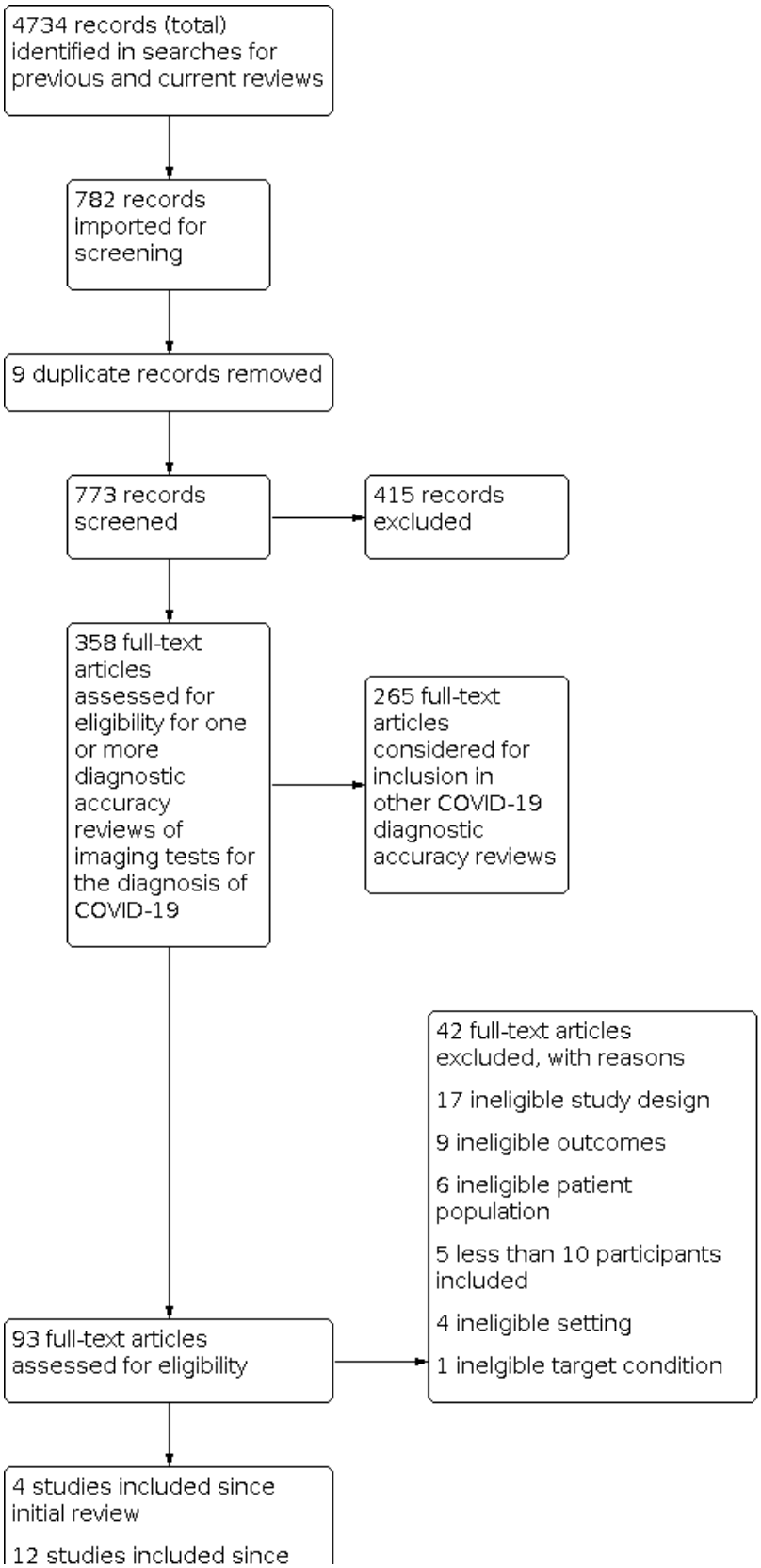
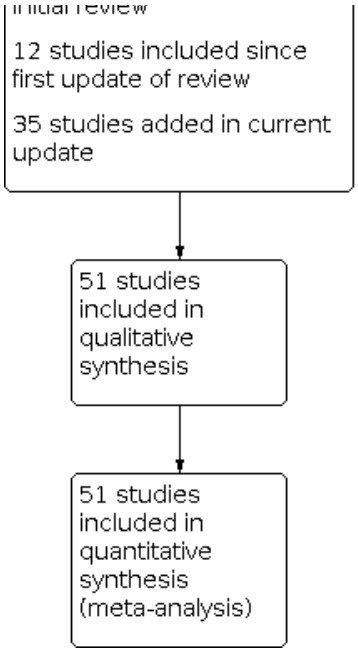


Figure 1. (Continued)



Description of included studies

We included 51 studies (38 CT, seven X-ray, two ultrasound, one both CT and X-ray, two both CT and ultrasound, and one both X-ray and ultrasound) with a total of 19,775 participants suspected of having COVID-19, of whom 10,155 (51%) had a final diagnosis of COVID-19.

The median sample size was 211 (interquartile range 94.5 to 486). Thirty-three studies were conducted in Europe (Italy 11, the Netherlands 7, France 5, Belgium 3, Turkey 3, Germany 2, UK 2), 13 were conducted in Asia (China 10, Korea 1, India 1, Japan 1) and the remaining studies were conducted in North America (USA 3) and South America (Brazil 2).

Index test readings were performed by radiologists in 39 studies (76%), radiology residents in two studies (4%), both radiologists and residents in one (2%) study, and radiographers in one study (2%); eight studies (16%) did not clearly report the level of training of readers. Technical parameters regarding the protocol of chest CT used in 41 studies were not clearly reported in 18 (44%) studies, while non-contrast CT was used in 10 (24%) studies, high-resolution chest CT was used in three (7%) studies, low-dose CT with or without contrast was used in seven (17%) studies and CT with IV contrast was used in three (7%) studies.

Manuscripts of three (6%) of the studies were published as preprints at the time of the search. We updated the publication status of all four of the preprint studies previously included in the first update of our review as of 1 November 2020: two studies were published since then, though there were no changes to the data between the preprint and published versions; one study remained as a preprint and had an updated version with updated data, which we re-extracted for analysis; and one study remained as a preprint without any updated versions.

Characteristics of the included studies are summarised in Table 1, and outlined in detail in the Characteristics of included studies.

Participant characteristics

Twenty-six studies included only adult participants (aged 16 years and over), 21 studies included both children and adults (although in most cases, only a minority of included patients were children), one study included only children, one study included participants aged 70 years and older, and the remaining two studies did not clearly report the age range of participants. All participants were suspected of having COVID-19. Thirty-three (65%) studies involved only symptomatic participants, 11 (22%) studies involved symptomatic and asymptomatic participants, and seven (14%) studies did not clearly report participants' symptom status.

All the studies used RT-PCR as the reference standard for the diagnosis of COVID-19, with 47 studies using only RT-PCR as the reference standard and four studies using a combination of RT-PCR and other criteria (clinical signs 1, clinical signs and imaging tests 1, positive contacts 1, and follow-up phone calls 1) as the reference standard. With respect to RT-PCR testing, two studies tested each participant once, 18 studies tested some participants at least twice, if necessary, 11 studies tested all participants at least twice, if necessary, and 20 studies did not report on the frequency of testing per participant. Two studies included inpatients, 32 studies included outpatients, while the remaining 17 studies were conducted in unclear settings. Seventeen (33%) studies described the co-morbidities of the study population, which commonly included hypertension, cardiovascular disease, and diabetes; however, the overall presence of co-morbidities in the participant groups of these studies was unclear.

Index tests

Forty-seven studies evaluated a single imaging modality and four studies evaluated two imaging modalities. In total, the 51 studies reported a total of 55 imaging modality evaluations. Chest CT was evaluated in 41 studies, chest X-ray was evaluated in nine studies, and ultrasound of the lungs was evaluated in five studies.

Methodological quality of included studies

Figure 2 provides a summary of the overall methodological quality assessment using the QUADAS-2 tool for all 51 included studies. Figure 3 displays a study-level quality assessment.

Figure 2. Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies.

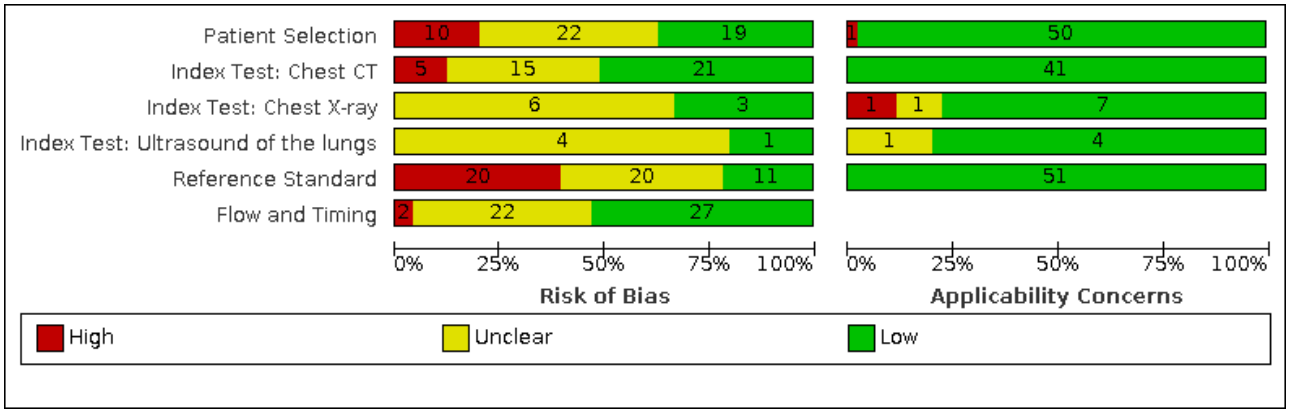
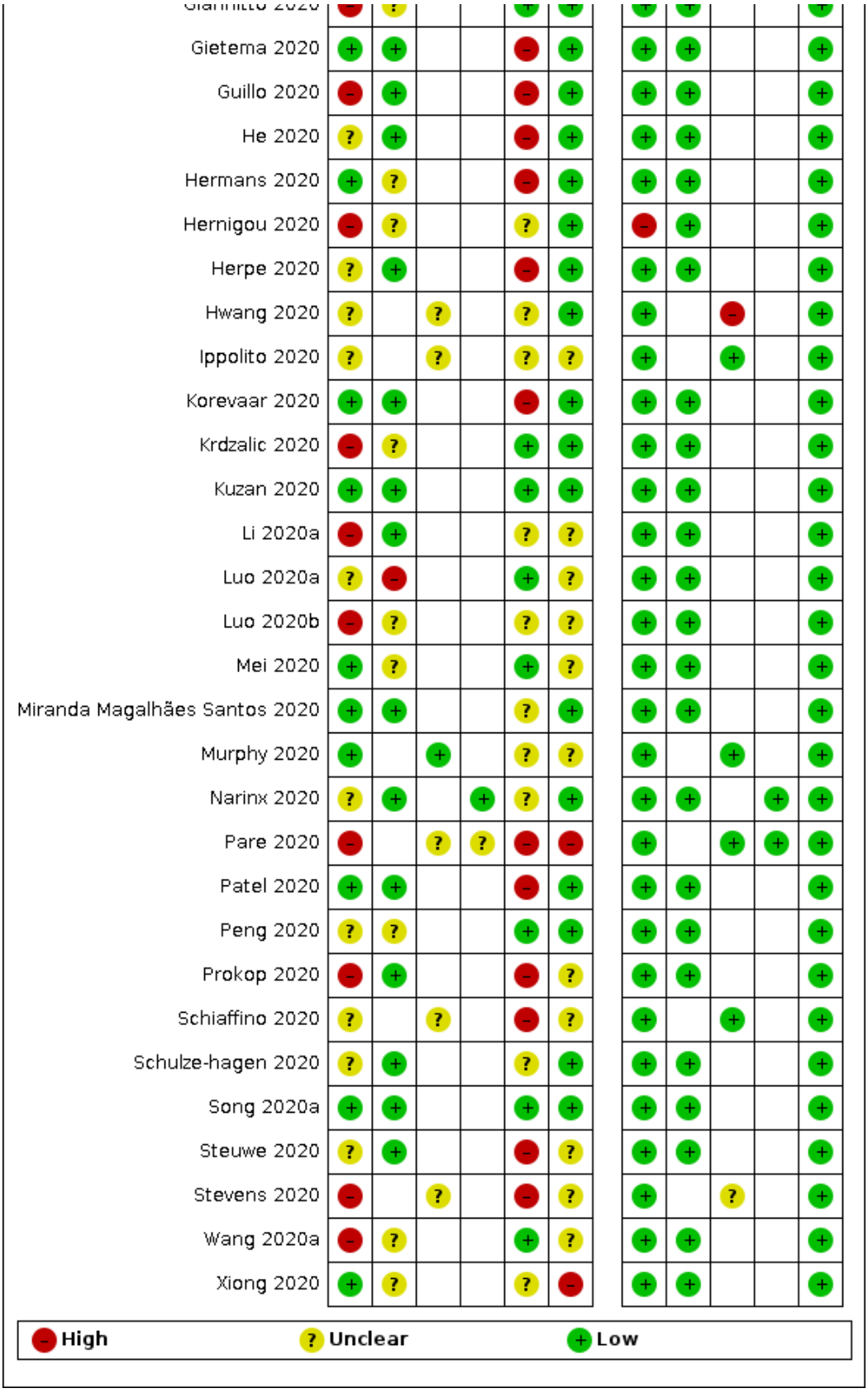


Figure 3. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study.

Figure 3. (Continued)



Overall, we found risk of bias based on concerns about the selection of participants to be high in 10 (20%) and unclear in 22 (43%) studies; the main concern in this domain was high risk of bias due to inappropriate exclusions (n = 10).

Risk of bias because of concerns regarding application of chest CT (41 studies) was high in five (12%) and unclear in 15 (37%) studies; risk of bias because of concerns regarding application of chest X-ray (9 studies) was unclear in six (67%) studies, and risk of bias because of concerns regarding application of ultrasound (5 studies) was unclear in four (80%) studies. The five CT studies with a high risk of bias did not predefine the positivity criteria for index tests (n = 3) or did not blind index test readers to reference standard results (n = 2).

Risk of bias based on concerns about the reference standard was high in 20 (39%) and unclear in 20 (39%) studies; the 20 studies with a high risk of bias used an RT-PCR protocol that was not likely to correctly classify the target condition.

Risk of bias based on concerns related to participant flow and timing was high in two (3.9%) and unclear in 22 (43%) studies; the two studies with a high risk of bias did not provide the same reference standard to all participants (n = 1), or did not have an appropriate time interval between the reference standard and index test (n = 1).

Concerns about the applicability of the evidence to participants were high in one study (2%). Concerns about the applicability of the evidence to the index test were low in all 41 chest CT studies, high in one (11%) and unclear in one (11%) chest X-ray study, (9 studies), and unclear in one (20%) ultrasound study (5 studies). Concerns about the applicability of the evidence to the reference standard were low in all 51 studies. Additional details about risk of bias and applicability assessment are presented in Figure 3.

Findings

Pooled estimates

The forest plot for chest CT is presented in Figure 4. The sensitivity of CT in 41 studies (involving 8110 (50%) cases in 16,133 participants) ranged from 56.3% to 100%, and the specificity ranged from 25.4% to 97.4%. The pooled sensitivity for chest CT was 87.9% (95% CI 84.6 to 90.6) and the pooled specificity was 80.0% (95% CI 74.9 to 84.3). The scatter of the study points in ROC space on the SROC plot (Figure 5) shows substantial variability in sensitivity and specificity.

Figure 4. Forest plot of chest CT in suspected cases.

Study	TP	FP	FN	TN	CO-RADS threshold	Sensitivity (95% CI)	Specificity (95% CI)
Steuwe 2020	19	19	0	67	-	1.00 [0.82, 1.00]	0.78 [0.68, 0.86]
Deng 2020	423	71	10	83	-	0.98 [0.96, 0.99]	0.54 [0.46, 0.62]
Li 2020a	42	19	1	30	-	0.98 [0.88, 1.00]	0.61 [0.46, 0.75]
Song 2020a	108	55	3	45	-	0.97 [0.92, 0.99]	0.45 [0.35, 0.55]
Caruso 2020	60	42	2	54	-	0.97 [0.89, 1.00]	0.56 [0.46, 0.66]
Wang 2020a	28	33	1	128	-	0.97 [0.82, 1.00]	0.80 [0.72, 0.85]
Ai 2020a	580	308	21	105	-	0.97 [0.95, 0.98]	0.25 [0.21, 0.30]
Xiong 2020	19	8	1	19	-	0.95 [0.75, 1.00]	0.70 [0.50, 0.86]
Schulze-hagen 2020	65	16	4	106	3	0.94 [0.86, 0.98]	0.87 [0.80, 0.92]
Gezer 2020	92	5	6	119	-	0.94 [0.87, 0.98]	0.96 [0.91, 0.99]
Fonsi 2020	41	2	3	17	-	0.93 [0.81, 0.99]	0.89 [0.67, 0.99]
Korevaar 2020	117	25	9	88	4	0.93 [0.87, 0.97]	0.78 [0.69, 0.85]
Barbosa 2020	23	25	2	41	-	0.92 [0.74, 0.99]	0.62 [0.49, 0.74]
Cartocci 2020	147	24	15	128	-	0.91 [0.85, 0.95]	0.84 [0.77, 0.90]
Falaschi 2020	419	66	43	245	-	0.91 [0.88, 0.93]	0.79 [0.74, 0.83]
Aslan 2020	226	20	24	36	-	0.90 [0.86, 0.94]	0.64 [0.50, 0.77]
Ducray 2020	259	49	28	358	-	0.90 [0.86, 0.93]	0.88 [0.84, 0.91]
Hermans 2020	120	22	13	164	4	0.90 [0.84, 0.95]	0.88 [0.83, 0.92]
Luo 2020b	70	7	8	55	-	0.90 [0.81, 0.95]	0.89 [0.78, 0.95]
Krdzalic 2020	25	7	3	21	3	0.89 [0.72, 0.98]	0.75 [0.55, 0.89]
Gietema 2020	74	35	9	75	-	0.89 [0.80, 0.95]	0.68 [0.59, 0.77]
Herpe 2020	1999	525	250	2050	-	0.89 [0.88, 0.90]	0.80 [0.78, 0.81]
Dofferhoff 2020	136	36	18	122	4	0.88 [0.82, 0.93]	0.77 [0.70, 0.84]
Luo 2020a	26	14	4	29	-	0.87 [0.69, 0.96]	0.67 [0.51, 0.81]
Borakati 2020	162	55	29	56	-	0.85 [0.79, 0.90]	0.50 [0.41, 0.60]
Miranda Magalhães Santos 2020	30	1	6	38	-	0.83 [0.67, 0.94]	0.97 [0.87, 1.00]
Hernigou 2020	13	2	3	29	-	0.81 [0.54, 0.96]	0.94 [0.79, 0.99]
Prokop 2020	43	5	10	47	4	0.81 [0.68, 0.91]	0.90 [0.79, 0.97]
Narinx 2020	12	10	3	65	-	0.80 [0.52, 0.96]	0.87 [0.77, 0.93]
Guillo 2020	103	11	26	74	-	0.80 [0.72, 0.86]	0.87 [0.78, 0.93]
De Smet 2020	279	33	79	468	5	0.78 [0.73, 0.82]	0.93 [0.91, 0.95]
Patel 2020	125	41	36	115	-	0.78 [0.70, 0.84]	0.74 [0.66, 0.80]
He 2020	26	2	8	46	-	0.76 [0.59, 0.89]	0.96 [0.86, 0.99]
Debray 2020	120	4	40	77	-	0.75 [0.68, 0.81]	0.95 [0.88, 0.99]
Fujioka 2020	57	10	19	68	4	0.75 [0.64, 0.84]	0.87 [0.78, 0.94]
Besutti 2020	438	16	158	84	-	0.73 [0.70, 0.77]	0.84 [0.75, 0.91]
Peng 2020	28	13	11	20	-	0.72 [0.55, 0.85]	0.61 [0.42, 0.77]
Giannitto 2020	14	10	6	38	-	0.70 [0.46, 0.88]	0.79 [0.65, 0.90]
Kuzan 2020	48	21	21	30	-	0.70 [0.57, 0.80]	0.59 [0.44, 0.72]
Mei 2020	274	39	145	447	-	0.65 [0.61, 0.70]	0.92 [0.89, 0.94]
Bellini 2020	80	76	62	354	4	0.56 [0.48, 0.65]	0.82 [0.78, 0.86]

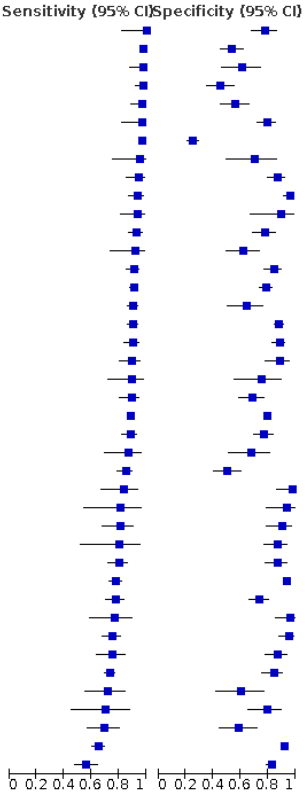
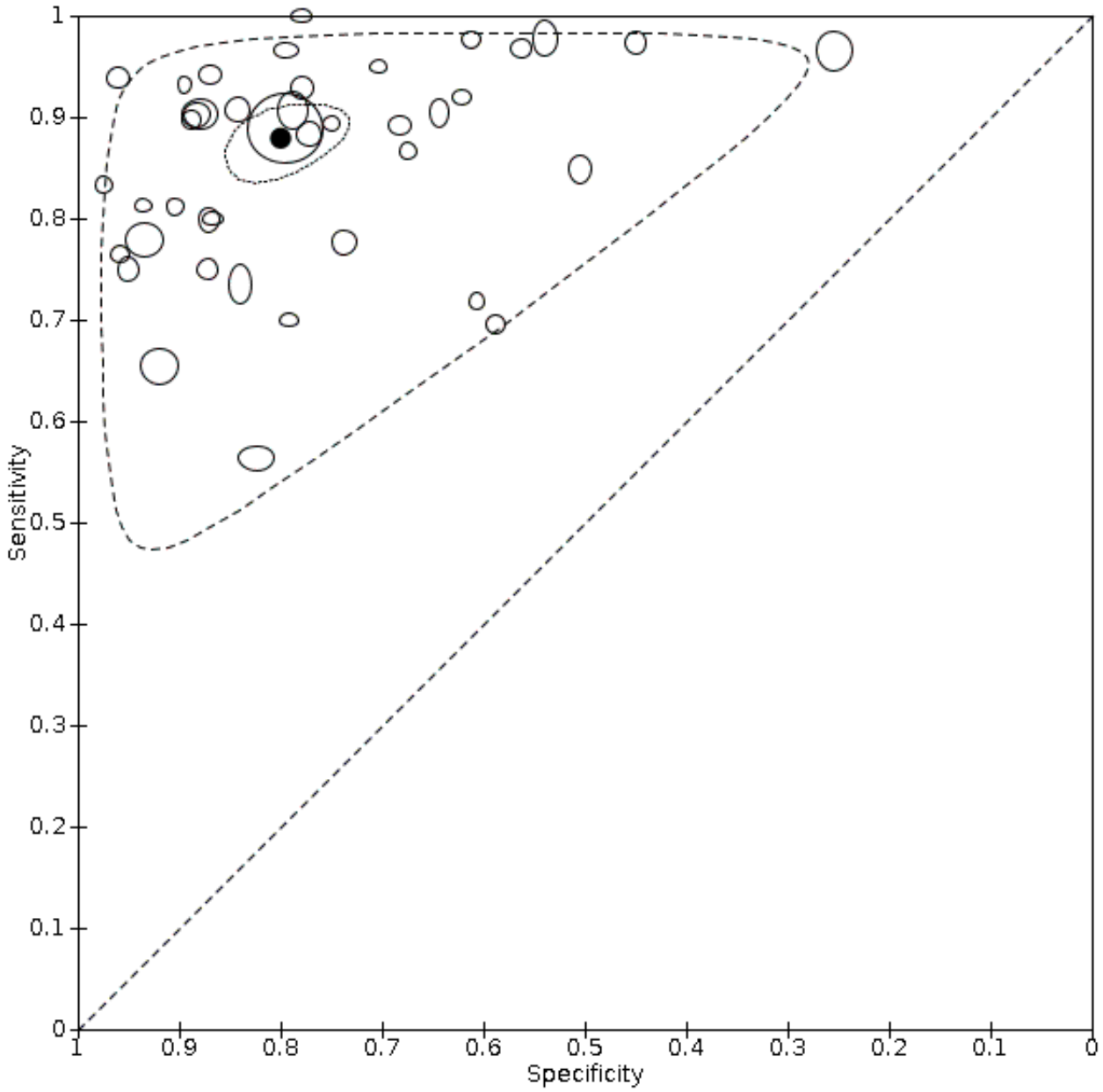


Figure 5. Summary ROC plot of chest CT in suspected cases. The summary point is indicated by the solid black circle, individual studies are indicated by outlined circles (scale=study sample size). The dotted border and the dashed border represent 95% confidence regions and 95% prediction regions, respectively.



The forest plots for chest X-ray and ultrasound of the lungs are presented in Figure 6. The sensitivity of chest X-ray in nine studies (including 2111 (57%) cases in 3694 participants) ranged from 51.9% to 94.4% and the specificity ranged from 40.4% to 88.9%. The pooled sensitivity for chest X-ray was 80.6% (95% CI 69.1 to 88.6) and the pooled specificity was 71.5% (95% CI 59.8 to 80.8). The scatter of the study points in ROC space on the SROC plot (Figure 7) shows substantial variability in sensitivity and specificity for chest X-ray.

Figure 6. Forest plot of A) chest X-ray and B) ultrasound in suspected cases.

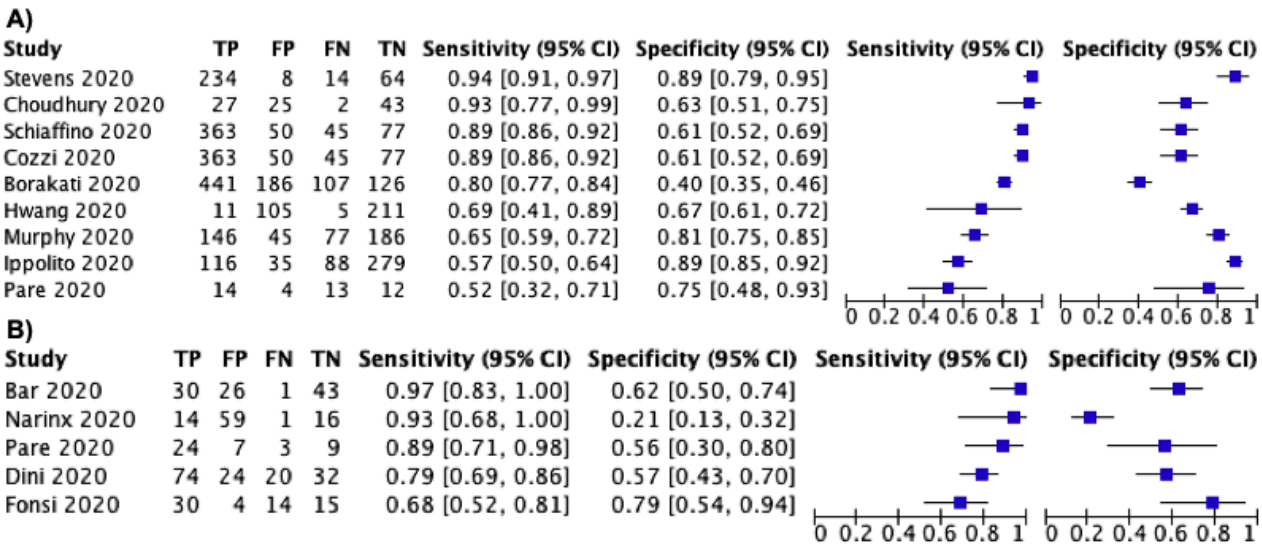
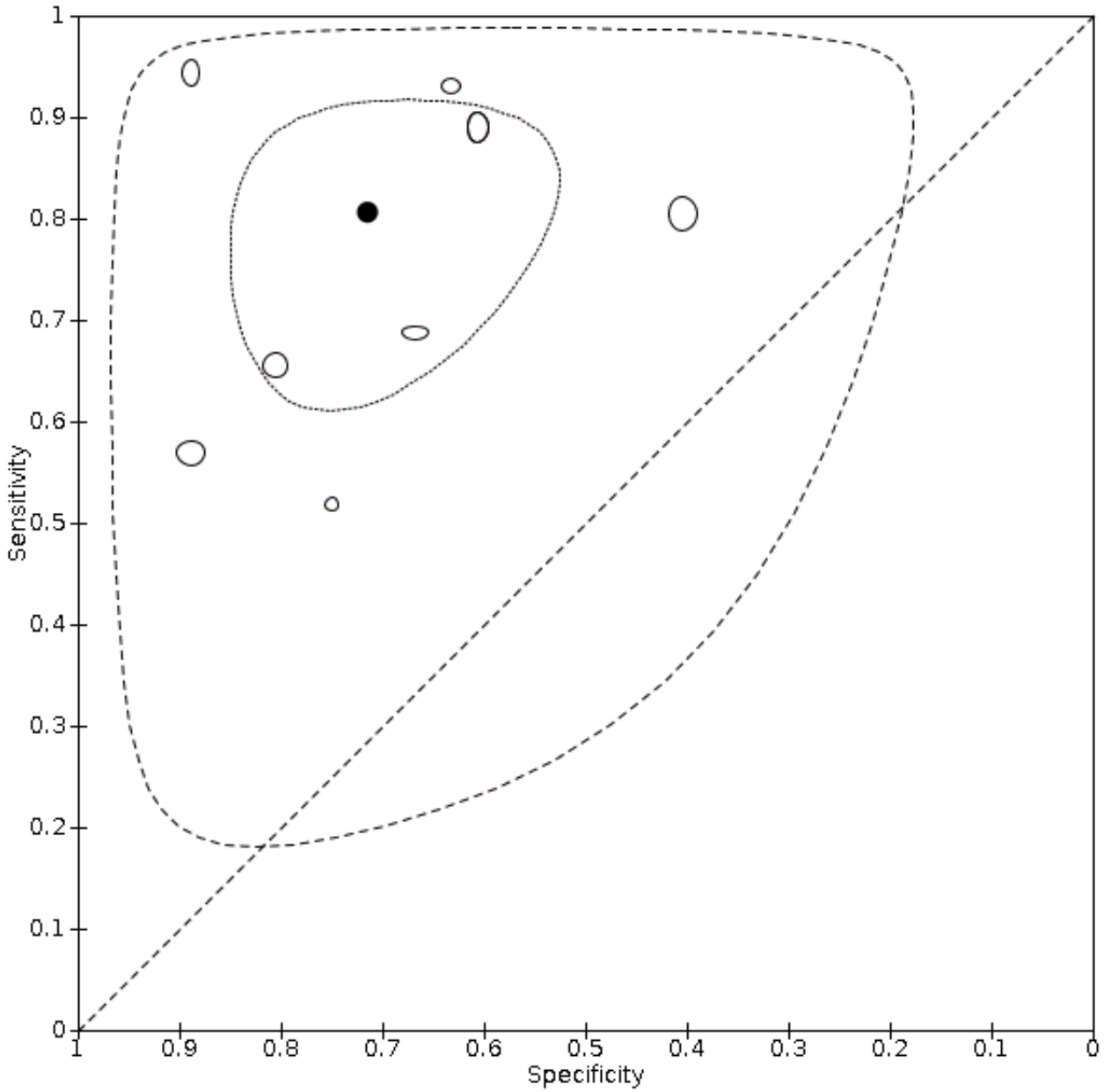
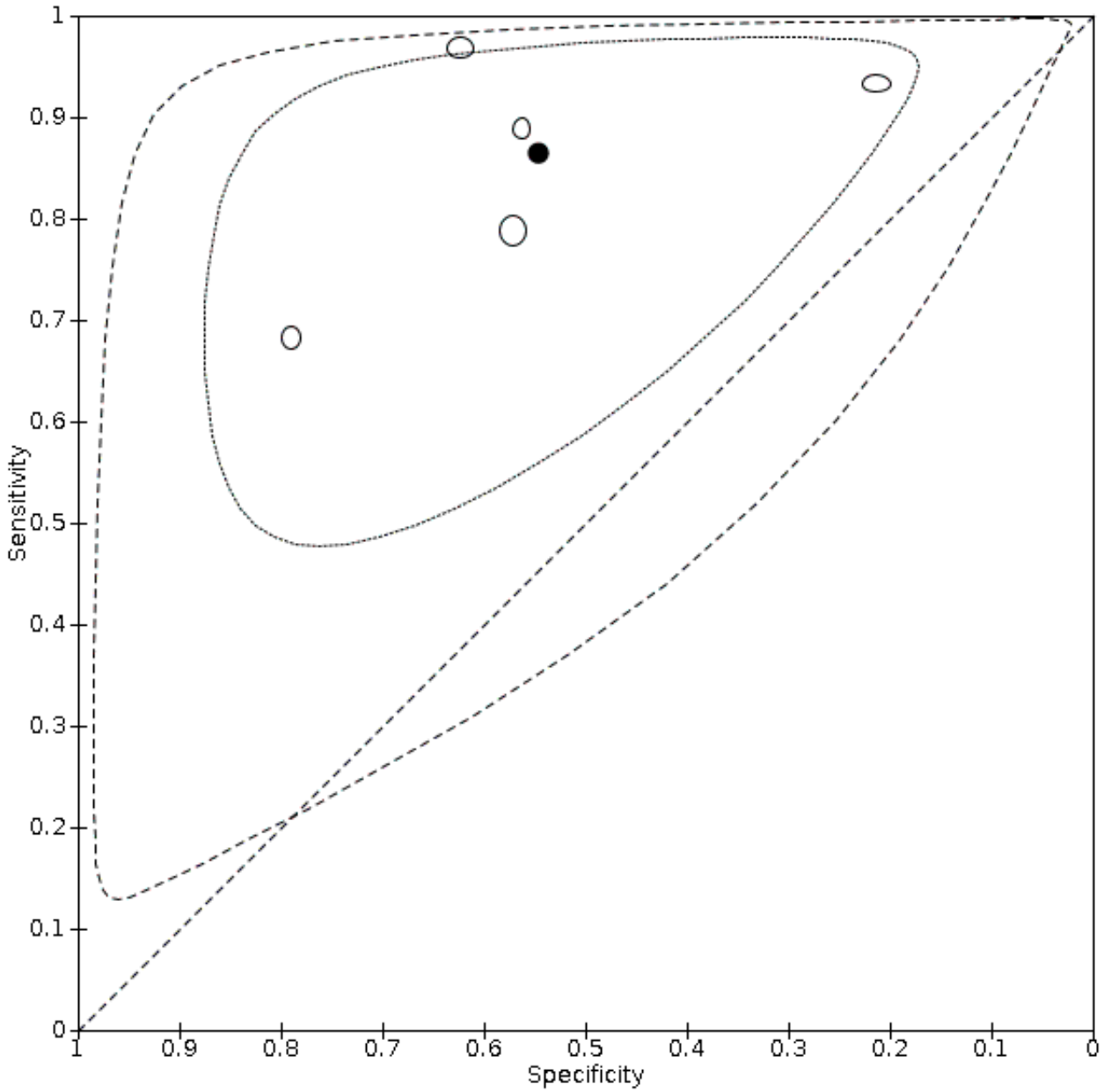


Figure 7. Summary ROC plot of chest X-ray in suspected cases. The summary point is indicated by the solid black circle, individual studies are indicated by outlined circles (scale=study sample size). The dotted border and the dashed border represent 95% confidence regions and 95% prediction regions, respectively.



The sensitivity of ultrasound of the lungs in five studies (including 211 (47%) cases in 446 participants) ranged from 68.2% to 96.8% and the specificity ranged from 21.3% to 78.9%. The pooled sensitivity for ultrasound was 86.4% (95% CI 72.7 to 93.9) and the pooled specificity was 54.6% (95% CI 35.3 to 72.6). The scatter of the study points in ROC space on the SROC plot (Figure 8) shows substantial variability in sensitivity and specificity for ultrasound of the lungs.

Figure 8. Summary ROC plot of ultrasound of the lungs in suspected cases. The summary point is indicated by the solid black circle, individual studies are indicated by outlined circles (scale=study sample size). The dotted border and the dashed border represent 95% confidence regions and 95% prediction regions, respectively.



Sensitivity analyses

Sensitivity analysis for CT studies limiting inclusion to studies published in peer-reviewed journals gave accuracy estimates similar to those of the overall included studies. When we excluded the three studies published as preprints, studies published in peer-reviewed journals (n = 38) had a pooled sensitivity of 88.5% (95% CI 85.2 to 91.2) and a pooled specificity of 81.2% (95% CI 76.0 to 85.3). These results are outlined in Table 2.

Investigations of heterogeneity

Investigations of heterogeneity for chest CT studies found that reference standard conduct, as well as definition for index test positivity, did not have an effect on accuracy estimates. The results of the investigations of heterogeneity are outlined in Table 3.

Stratification by reference standard conduct gave pooled sensitivity estimates of 88.1% (95% CI 78.5 to 93.8) for studies that performed RT-PCR testing at least twice for all participants with initial negative results versus 86.7% (95% CI 82.3 to 90.1) for studies that did not perform repeat RT-PCR testing for all participants with initial negative results (P = 0.74). Pooled specificity estimates were 71.3%

(95% CI 59.9 to 80.6) for studies that performed RT-PCR testing at least twice for all participants with initial negative results versus 82.6% (95% CI 77.8 to 86.6) for studies that did not perform repeat RT-PCR testing for all participants with initial negative results (P = 0.05). For the subgroup of CT studies that did not perform repeat RT-PCR testing in all participants with initial negative results (n = 22), the proportion of participants that underwent repeat RT-PCR testing ranged from 0% to 61% amongst the nine studies that reported this information; the remaining thirteen studies did not clearly report this information.

Stratification by definition used for index test positivity gave pooled sensitivity estimates of 90.3% (95% CI 84.5 to 94.1) for studies that defined index test positivity based on radiologist's impressions versus 85.9% (95% CI 81.2 to 89.2) for studies that used a formal scoring system to define index test positivity (P = 0.15). Pooled specificity estimates were 77.2% (95% CI 67.0 to 84.9) for studies that used radiologist's impressions versus 80.0% (95% CI 75.0 to 84.2) for studies that used a formal scoring system (P = 0.58). For studies that used a formal scoring system, we used the threshold demonstrating the highest Youden's index in each study (or as in the cases of two studies that did not report data at all thresholds, the only threshold that was available (i.e. CO-RADS 3)) in the analysis.

Threshold effects

Nine studies that evaluated CT (involving 1139 (41%) cases amongst 2807 participants) used the CO-RADS scoring system to define index test positivity. We obtained the 2x2 data at all five CO-RADS thresholds for seven studies; two studies only reported 2x2 data at a CO-RADS threshold of 3, and the authors could not provide any additional data. The forest plots of chest CT studies that used CO-RADS and reported 2x2 data for CO-RADS thresholds 2, 3, 4 and 5 are presented in Figure 9.

Figure 9. Forest plot of chest CT studies in suspected cases that used the CO-RADS scoring system. Threshold (i.e. 2, 3 4, or 5) indicated in parentheses ahead of Study ID. Grouped by threshold.

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
(2) Bellini 2020	107	216	35	214	0.75 [0.67, 0.82]	0.50 [0.45, 0.55]
(2) De Smet 2020	331	215	27	286	0.92 [0.89, 0.95]	0.57 [0.53, 0.61]
(2) Dofferhoff 2020	145	79	9	79	0.94 [0.89, 0.97]	0.50 [0.42, 0.58]
(2) Fujioka 2020	70	50	6	28	0.92 [0.84, 0.97]	0.36 [0.25, 0.48]
(2) Hermans 2020	127	93	6	93	0.95 [0.90, 0.98]	0.50 [0.43, 0.57]
(2) Korevaar 2020	124	83	2	30	0.98 [0.94, 1.00]	0.27 [0.19, 0.36]
(2) Prokop 2020	334	33	1	19	1.00 [0.98, 1.00]	0.37 [0.24, 0.51]
(3) Bellini 2020	93	134	49	296	0.65 [0.57, 0.73]	0.69 [0.64, 0.73]
(3) De Smet 2020	319	138	39	363	0.89 [0.85, 0.92]	0.72 [0.68, 0.76]
(3) Dofferhoff 2020	142	61	12	97	0.92 [0.87, 0.96]	0.61 [0.53, 0.69]
(3) Fujioka 2020	67	26	9	52	0.88 [0.79, 0.94]	0.67 [0.55, 0.77]
(3) Hermans 2020	124	64	9	122	0.93 [0.88, 0.97]	0.66 [0.58, 0.72]
(3) Korevaar 2020	122	49	4	64	0.97 [0.92, 0.99]	0.57 [0.47, 0.66]
(3) Krdzalic 2020	25	7	3	21	0.89 [0.72, 0.98]	0.75 [0.55, 0.89]
(3) Prokop 2020	325	16	4	36	0.99 [0.97, 1.00]	0.69 [0.55, 0.81]
(3) Schulze-Hagen 2020	65	16	4	106	0.94 [0.86, 0.98]	0.87 [0.80, 0.92]
(4) Bellini 2020	80	76	62	354	0.56 [0.48, 0.65]	0.82 [0.78, 0.86]
(4) De Smet 2020	304	76	54	425	0.85 [0.81, 0.88]	0.85 [0.81, 0.88]
(4) Dofferhoff 2020	136	36	18	122	0.88 [0.82, 0.93]	0.77 [0.70, 0.84]
(4) Fujioka 2020	57	10	19	68	0.75 [0.64, 0.84]	0.87 [0.78, 0.94]
(4) Hermans 2020	120	22	13	164	0.90 [0.84, 0.95]	0.88 [0.83, 0.92]
(4) Korevaar 2020	117	25	9	88	0.93 [0.87, 0.97]	0.78 [0.69, 0.85]
(4) Prokop 2020	43	5	10	47	0.81 [0.68, 0.91]	0.90 [0.79, 0.97]
(5) Bellini 2020	59	41	83	389	0.42 [0.33, 0.50]	0.90 [0.87, 0.93]
(5) De Smet 2020	279	33	79	468	0.78 [0.73, 0.82]	0.93 [0.91, 0.95]
(5) Dofferhoff 2020	119	26	35	132	0.77 [0.70, 0.84]	0.84 [0.77, 0.89]
(5) Fujioka 2020	39	3	37	75	0.51 [0.40, 0.63]	0.96 [0.89, 0.99]
(5) Hermans 2020	100	11	33	175	0.75 [0.67, 0.82]	0.94 [0.90, 0.97]
(5) Korevaar 2020	94	18	32	95	0.75 [0.66, 0.82]	0.84 [0.76, 0.90]
(5) Prokop 2020	33	3	20	49	0.62 [0.48, 0.75]	0.94 [0.84, 0.99]

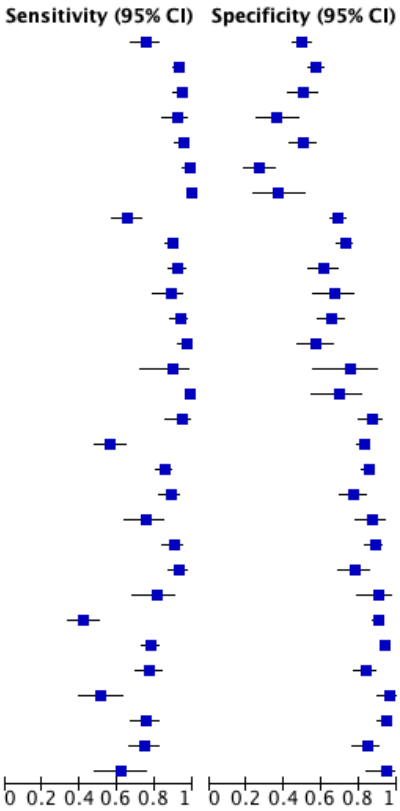
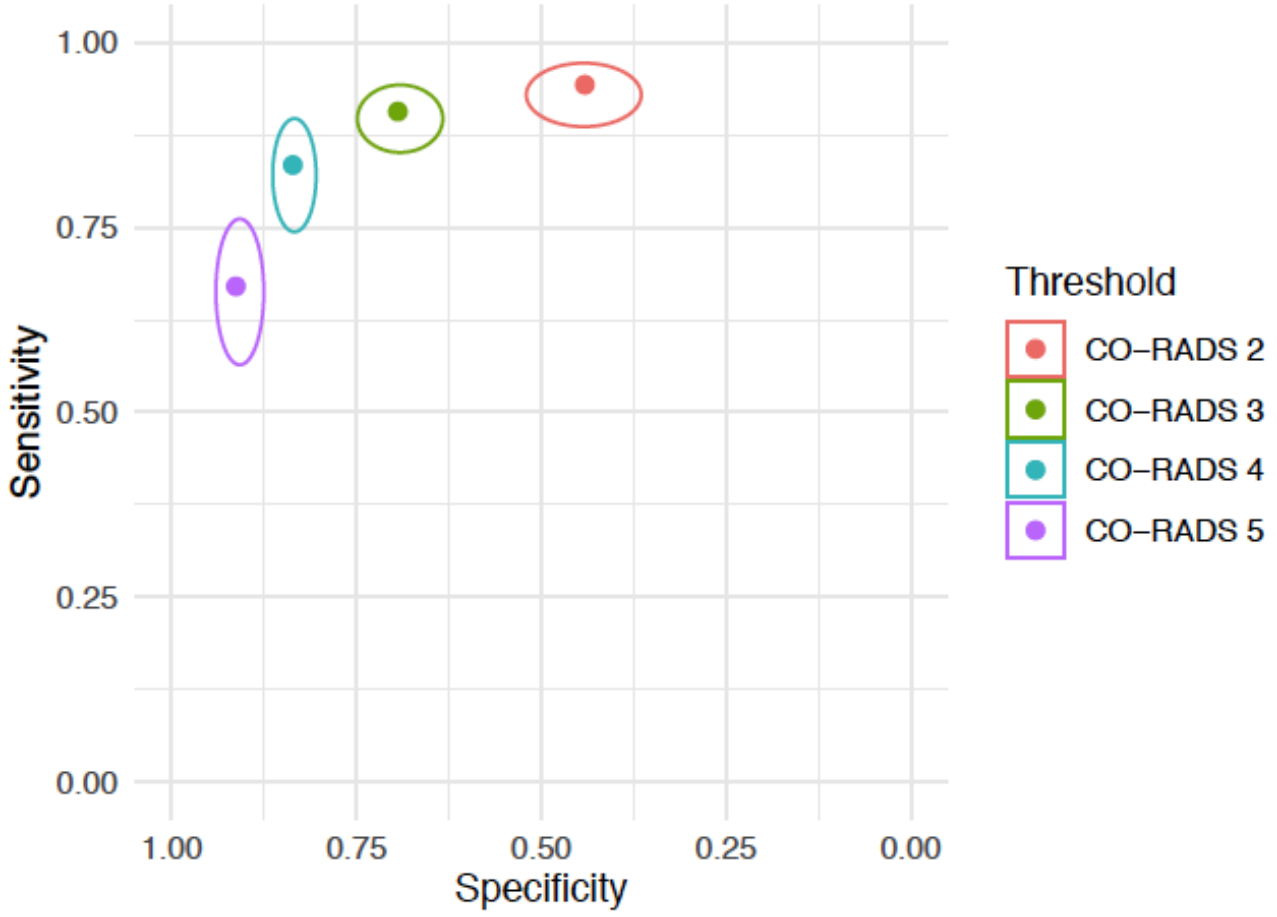


Figure 10 displays pooled sensitivity and specificity estimates with 95% CIs at each CO-RADS thresholds; Table 4 summarises the results.

Figure 10. Pooled sensitivity and specificity estimates and 95% confidence intervals at varying CO-RADS thresholds: CO-RADS 2 (n = 7), CO-RADS 3 (n = 9), CO-RADS 4 (n = 7), and CO-RADS 5 (n = 7).



- At a CO-RADS threshold of 5 (7 studies), the sensitivity ranged from 41.5% to 77.9% and the specificity ranged from 83.5% to 96.2%; the pooled sensitivity was 67.0% (95% CI 56.4 to 76.2) and the pooled specificity was 91.3% (95% CI 87.6 to 94.0).
- At a CO-RADS threshold of 4 (7 studies), the sensitivity ranged from 56.3% to 92.9% and the specificity ranged from 77.2% to 90.4%; the pooled sensitivity was 83.5% (95% CI 74.4 to 89.7) and the pooled specificity was 83.6% (95% CI 80.5 to 86.4).
- At a CO-RADS threshold of 3 (9 studies), the sensitivity ranged from 65.5% to 98.8% and the specificity ranged from 56.6% to 86.9%; the pooled sensitivity was 90.7% (95% CI 85.2 to 94.3) and the pooled specificity was 69.4% (95% CI 63.3 to 74.9).
- At a CO-RADS threshold of 2 (7 studies), the sensitivity ranged from 75.4% to 99.7% and the specificity ranged from 26.5% to 57.1%; the pooled sensitivity was 94.3% (95% CI 88.6 to 97.2) and the pooled specificity was 44.1% (95% CI 36.5 to 52.0).
- We did not perform meta-analysis for a CO-RADS threshold of 1, since at this threshold, all sensitivity values are equal to 1, and all specificity values are equal to 0.

Direct and indirect test comparisons

Four included studies evaluated two modalities each (one study on chest CT and chest X-ray, two studies on chest CT and ultrasound, and one study on chest X-ray and ultrasound). Both modalities in each study were evaluated in the same population, or the second modality was evaluated in a subset of the population in which the first modality was assessed (i.e. direct comparisons). Paired SROC plots for these four comparative studies are presented in Figure 11, Figure 12, and Figure 13. In the one study that evaluated chest CT and chest X-ray, both modalities had similar sensitivities and specificities. In the two studies that evaluated chest CT and ultrasound, chest CT had similar sensitivities and higher specificities compared to ultrasound. In the one study that evaluated chest X-ray and ultrasound, chest X-ray had a lower sensitivity and a higher specificity compared to ultrasound. We could not perform formal analyses to compare accuracy estimates of modalities assessed in the same population directly due to the limited number of studies.

Figure 11. Summary ROC plot of comparative studies: chest CT versus chest X-ray. The dotted line connects index tests pairs evaluated in the same study (i.e. a comparative study).

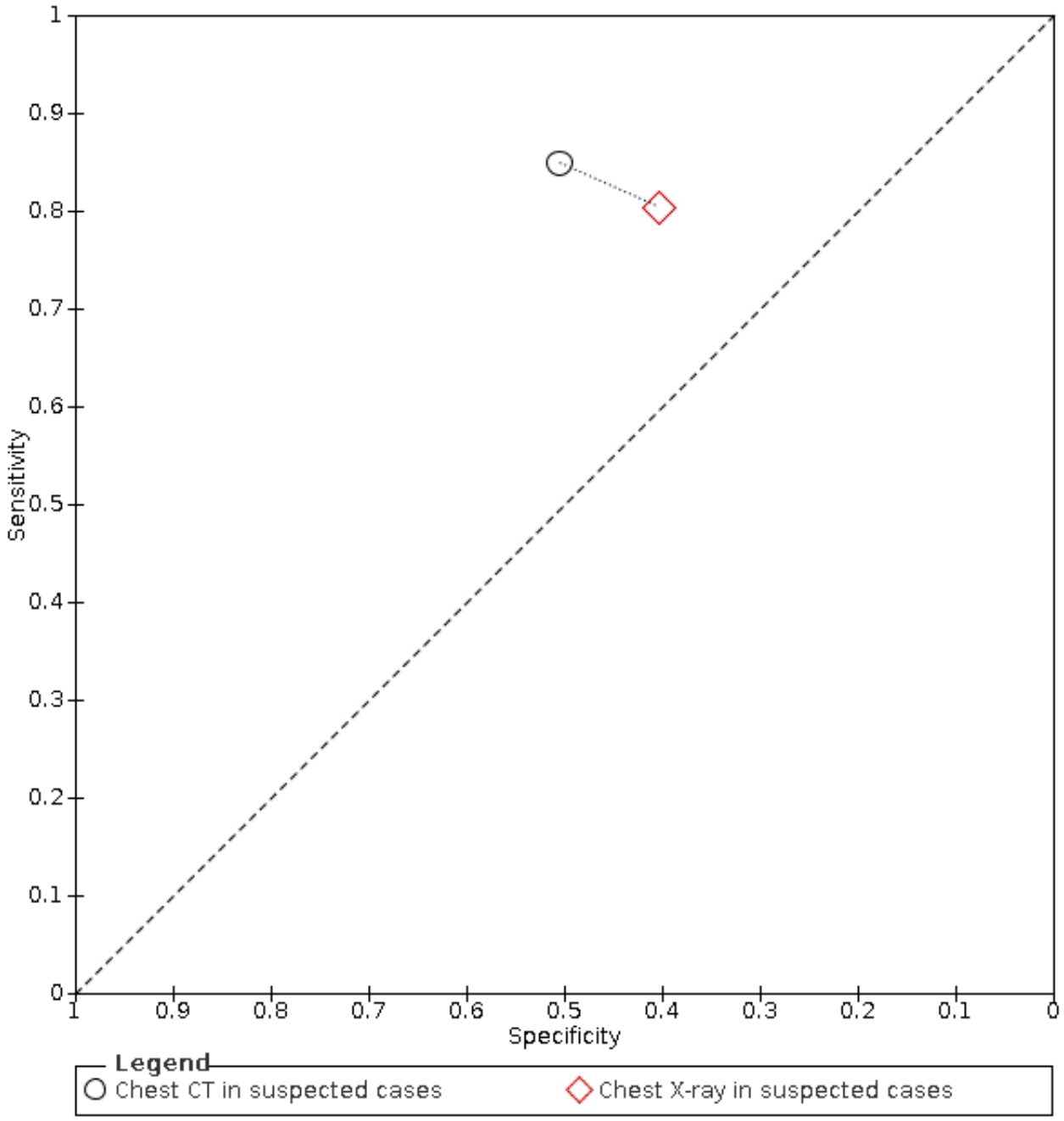


Figure 12. Summary ROC plot of comparative studies: chest CT versus ultrasound of the lungs. The dotted lines connect index tests pairs evaluated in the same study (i.e. a comparative study).

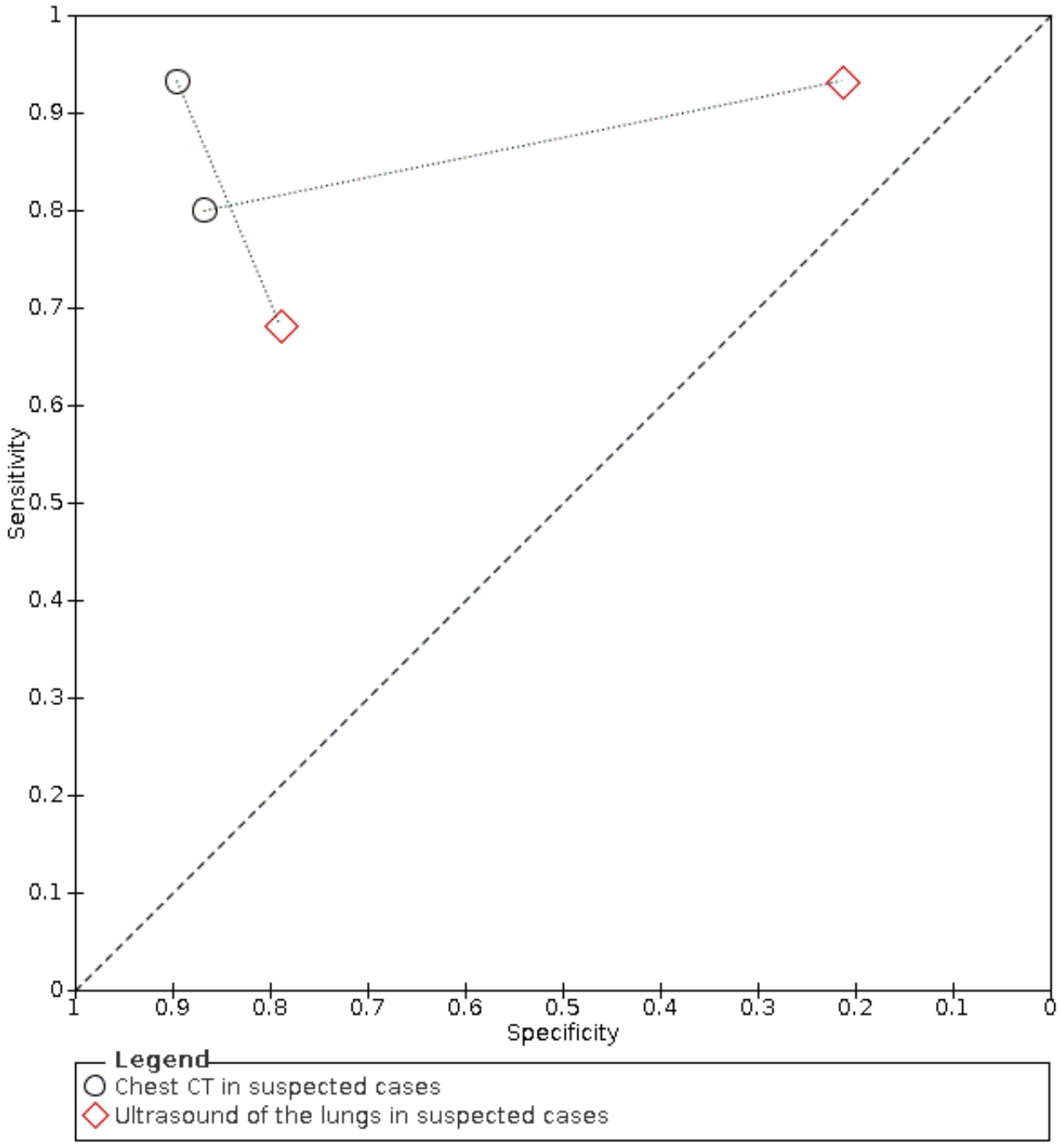
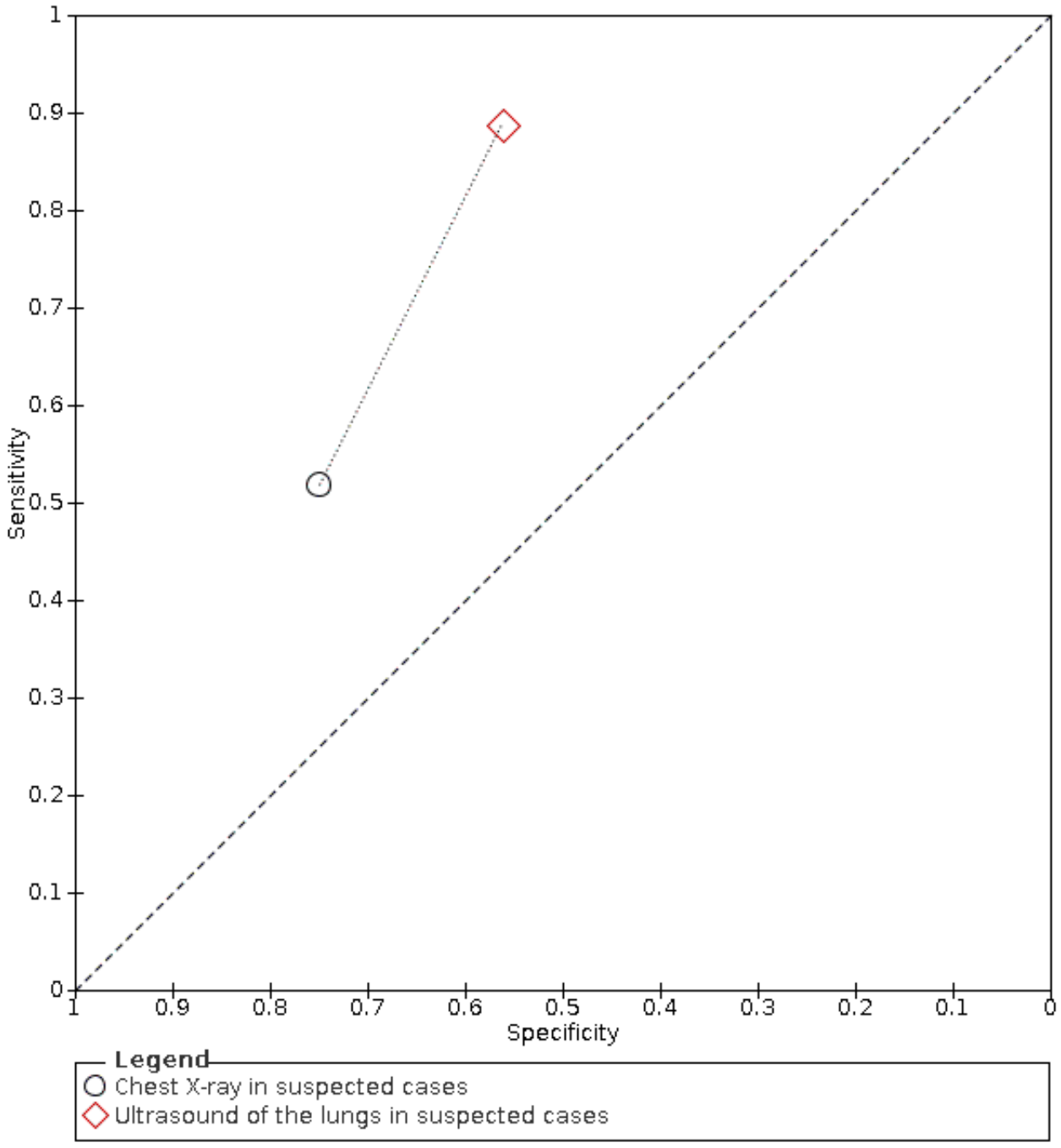


Figure 13. Summary ROC plot of comparative studies: chest X-ray versus ultrasound of the lungs. The dotted line connects index tests pairs evaluated in the same study (i.e. a comparative study).



Indirect comparisons of modalities evaluated across all 51 studies indicated that: chest CT (41 studies) and chest X-ray (9 studies) gave similar sensitivity (P = 0.10) and specificity (P = 0.12) estimates; chest CT and ultrasound (5 studies) gave similar sensitivity estimates (P = 0.77), while chest CT gave higher specificity estimates than ultrasound (P = 0.0052); and chest X-ray and ultrasound gave similar sensitivity (P = 0.43) and specificity (P = 0.13) estimates. These findings are summarised in Table 5.

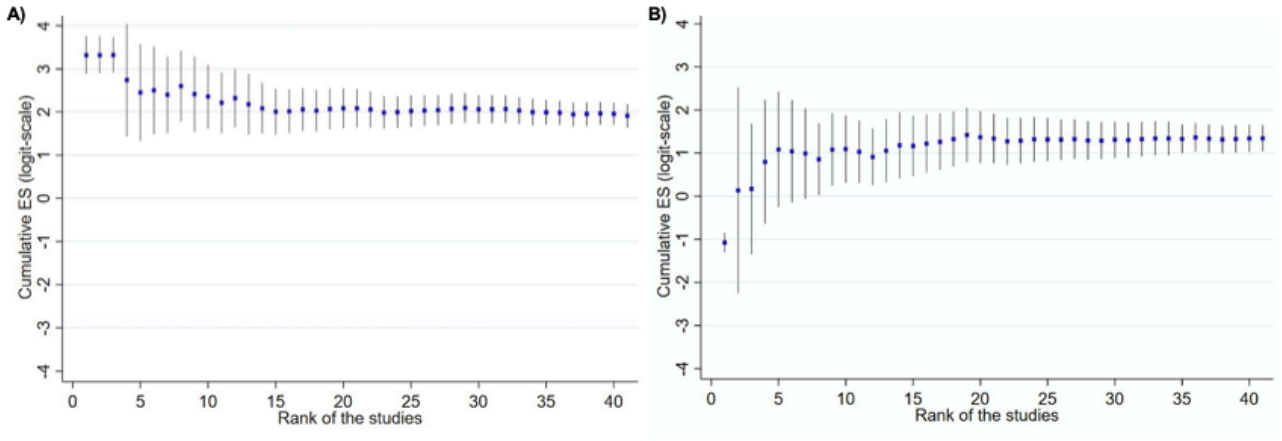
Cumulative meta-analyses and time trends

Cumulative meta-analyses and time trends analyses of the 41 chest CT studies included in this current review indicated a decrease in cumulative estimates of sensitivity and an increase in cumulative estimates of specificity over time; both P values < 0.001.

Figure 14 displays the cumulative meta-analyses of logit-sensitivity and logit-specificity over time, respectively. Based on visual

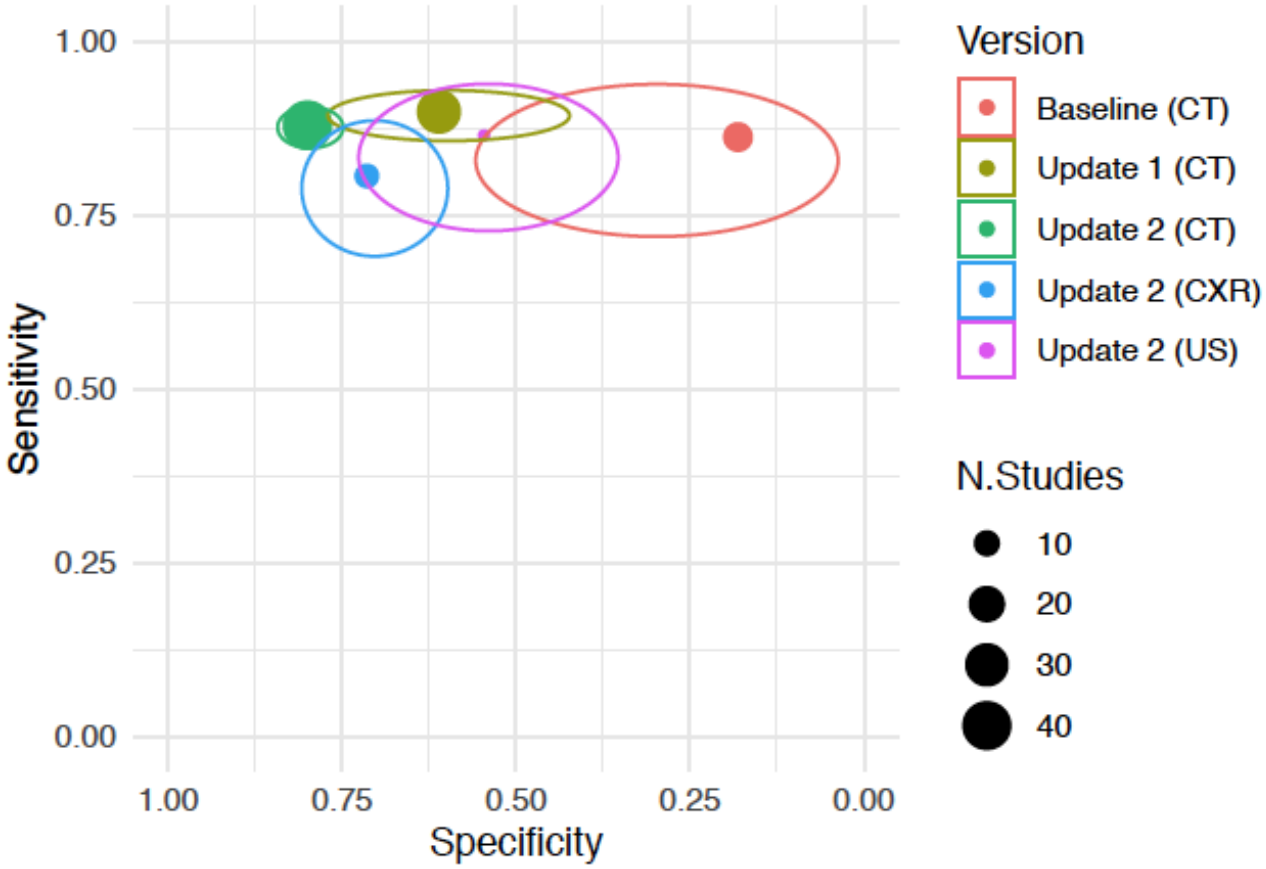
assessment, the meta-analysis estimates for sensitivity and specificity appear to stabilise near their final values after 20 consecutive studies have been included in the meta-analysis.

Figure 14. Cumulative meta-analysis of chest CT A) logit-sensitivity and B) logit-specificity. Includes 41 studies in the current review. Studies are ranked according to date of publication (1 = oldest, 41 = most recent).



Based on the pooled sensitivity and specificity estimates derived across all three versions of this review (i.e. initial review published in September 2020 (Salameh 2020a), first update published in November 2020 (Islam 2020), and this current update), the sensitivity estimates of chest CT appear to be similar across all versions, while the specificity estimates of chest CT appear to increase with each update. Figure 15 displays the pooled sensitivity and specificity estimates with 95% CIs from all versions of this review.

Figure 15. Pooled sensitivity and specificity estimates and 95% confidence intervals across all versions of this review.



DISCUSSION

This is the second update of a Cochrane living systematic review evaluating the diagnostic accuracy of thoracic imaging (CT, chest X-ray and ultrasound) in the evaluation of people suspected to have COVID-19. This version of the review is based on published studies and preprints up to 30 September 2020.

Summary of main results

Chest CT (41 studies, 16,133 participants, 8110 (50%) cases) demonstrated a sensitivity of 87.9% (95% CI 84.6 to 90.6), and a specificity of 80.0% (95% CI 74.9 to 84.3) for the diagnosis of COVID-19 in suspected participants. Compared with the findings of the first update of this 'living' systematic review, in which we determined that chest CT had a sensitivity of 89.9% (95% CI 85.7 to 92.9) and specificity of 61.1% (95% CI 42.3 to 77.1), our current update demonstrates similar sensitivity, but notably higher specificity. Possible explanations for this improved specificity could include the stricter study inclusion criteria applied for this update, particularly the exclusion of studies that report an overview of index test findings in participants with and without the target condition, without explicitly classifying the imaging test as either COVID-19 positive or negative. The improved specificity could also be due to the addition of more studies that used well-developed definitions for index test positivity used by index test readers (e.g. CO-RADS, BSTI COVID-19 Reporting Template, RSNA Expert Consensus on Reporting Chest CT Findings for COVID-19). Furthermore, studies from the early stage of the pandemic were included in our initial review and studies from a later stage were added in this update; thus, the stage of the pandemic during which included studies were conducted have likely influenced these differing specificity estimates through improved knowledge about the pathophysiology and imaging manifestations of COVID-19. This is supported by the results of our cumulative meta-analysis and time trends analysis, which indicate that specificity has increased over time. The influence of these proposed variables on only specificity and not sensitivity remains unexplained and requires exploration in future updates.

Sensitivity analysis for chest CT studies, limiting inclusion to studies published in peer-reviewed journals, gave accuracy estimates similar to those of the overall included studies. Based on this, publication status had a minimal effect on our results. In our previous review, publication status did not appear to contribute to heterogeneity.

There was no statistical evidence of the effect of reference standard conduct on the sensitivity or specificity of chest CT; studies that performed RT-PCR testing at least twice for all initial negative

results and studies that did not perform repeat RT-PCR testing for all initial negative results had similar sensitivities and specificities. These findings align with those of our previous review, in which a sensitivity analysis limiting inclusion to studies that implemented RT-PCR testing at least twice for all initial negative results, gave accuracy estimates similar to those of the overall included studies. However, as the results of this meta-regression in the current review were close to statistical significance, reference standard conduct will continue to be of interest in future updates of this review.

Surprisingly, the definition used for index test positivity in chest CT studies did not appear to affect test accuracy, as studies that used a formal scoring system demonstrated comparable accuracy to those that used radiologists' impressions. A possible explanation is that radiologists using the 'radiologist impression' method may have implicitly used a formal scoring system that they were previously familiar with (e.g. CO-RADS). Thus, there may be minimal differences in the interpretation of chest CT between the formal scoring system and radiologist impression groups.

In chest CT studies that used the CO-RADS scoring system to define index test positivity (9 studies, 2807 participants with 1139 (41%) cases), as expected, when the threshold for index test positivity increased (i.e. from 2 to 5), sensitivity decreased and specificity increased.

Chest X-ray (9 studies, 3694 participants with 2111 (57%) cases) demonstrated a sensitivity of 80.6% (95% CI 69.1 to 88.6), and a specificity of 71.5% (95% CI 59.8 to 80.8) for the diagnosis of COVID-19 in suspected participants. Ultrasound (5 studies, 446 participants with 211 (47%) cases) demonstrated a sensitivity of 86.4% (95% CI 72.7 to 93.9), and a specificity of 54.6% (95% CI 35.3 to 72.6). As meta-analysis was not performed for chest X-ray or ultrasound in our previous review due to a limited number of studies, comparisons between our current and previous findings are not possible. The number of included studies for chest X-ray and ultrasound was not sufficient to conduct meta-regression analyses.

Direct and indirect test comparisons

Based on indirect comparisons of all included studies, chest CT and chest X-ray had similar sensitivities and specificities. Chest CT and ultrasound had similar sensitivities, but chest CT had a higher specificity than ultrasound. Chest X-ray and ultrasound had similar sensitivities and specificities.

The indirect comparisons between chest CT and chest X-ray, and chest CT and ultrasound are supported by the limited evidence provided by studies that performed direct comparisons. The one study that compared chest CT and chest X-ray in the same population showed that sensitivities and specificities for both modalities were similar (Borakati 2020). Both studies that compared chest CT and ultrasound in the same population found that the two modalities had similar sensitivities and chest CT had a higher specificity compared to ultrasound (Fonsi 2020; Narinx 2020). However, the one study that compared chest X-ray and ultrasound in the same population showed that X-ray had a lower sensitivity and a higher specificity compared to ultrasound (Pare 2020), which contradicts the findings of the indirect comparison. These findings require investigation with more direct comparative design studies in future updates.

Cumulative meta-analyses and time trends

With respect to the three versions of this review, the sensitivity estimates of chest CT appear to remain similar across all versions, while the specificity estimates of chest CT appear to increase with each update. With the current number of chest CT studies included in this review, sensitivity and specificity estimates appear to be stable and have narrow confidence intervals. This may suggest that pooled sensitivity and specificity estimates of chest CT will not notably differ in future updates of this review.

Strengths and weaknesses of the review

Our search strategy was broad and allowed for identification of a wide range of articles about COVID-19 diagnosis. The review authors screened records, extracted data, and assessed study methodology independently and in duplicate. Though we are relatively confident in the accuracy and completeness of our findings, please inform us at mmcinnes@toh.ca should errors be found so that we can address them in a future update. Furthermore, compared to our initial review (Salameh 2020a), as well as the first update of our review (Islam 2020), this current update includes a greater number of studies that evaluated accuracy estimates of imaging tests in the diagnosis of suspected COVID-19 participants.

We included studies that involved only symptomatic participants, as well as studies that had a mixed population (i.e. symptomatic and asymptomatic participants). Thus, there may be situations when asymptomatic individuals are suspected of having COVID-19, such as if they have infected contacts or other risk factors for infection. However, not all the studies clearly reported information on participants' symptoms.

We did not identify reference standard conduct or definition for index test positivity as sources of variability for chest CT accuracy. These findings may suggest that the variables we investigated did not significantly contribute to variability; alternatively, there may be confounding variables obscuring our analyses. Due to insufficient granularity of data, we were unable to investigate additional potential sources of variability, particularly participant setting (inpatient versus outpatient) and symptom status (symptomatic versus asymptomatic). Furthermore, we were unable to investigate potential sources of variability of chest X-ray and ultrasound accuracy estimates due to the limited number of studies. We plan to perform these analyses in future updates, when sufficient data become available.

In this update, we addressed our secondary objective of evaluating threshold effects of imaging findings of COVID-19 on accuracy measures, particularly with respect to the CO-RADS scoring system. We could not evaluate threshold effects for studies that used other scoring systems (e.g. BSTI COVID-19 Reporting Template, RSNA Expert Consensus on Reporting Chest CT Findings for COVID-19) due to the limited number of included studies that used other scoring systems.

We explored indirect comparisons of chest CT, chest X-ray and ultrasound of the lungs, and we qualitatively assessed studies that evaluated multiple imaging modalities in the same population (i.e. direct comparisons). However, due to the limited number of studies that evaluated multiple imaging modalities in the same population, we did not formally evaluate direct comparisons of different imaging tests at this stage. We plan to conduct formal

analyses of direct comparisons of imaging tests in future updates, as more studies with comparative designs become available.

We performed the cumulative meta-analyses and time trends analyses of chest CT accuracy estimates using a univariate model, whereas we performed all other meta-analyses in this review using a bivariate model. The univariate model, which analyses sensitivity and specificity separately instead of together, is therefore a limitation of the cumulative meta-analyses and time trends and should be considered when interpreting the results of the analyses. We did not perform cumulative meta-analyses and time trends of chest X-ray and ultrasound accuracy due to the limited number of studies.

We were not able to evaluate accuracy estimates based on specific findings of imaging tests (e.g. ground-glass, consolidation, pleural effusion) or combinations of such findings because of the lack of data granularity reported in included studies; however, we will consider this in future updates of the review.

We were not able to evaluate several planned additional secondary objectives due to insufficient data. Important questions that remain a concern are:

1. the rate of positive thoracic imaging in individuals with initial negative RT-PCR results who have positive RT-PCR results on repeat testing;
2. possible associations between findings on thoracic imaging for patients with COVID-19, the number of days after symptom onset, symptom severity and subsequent hospitalisation; and
3. the screening of asymptomatic individuals.

We hope that in future updates of this review we will be able to evaluate these associations as research on the role of imaging tests in the diagnosis of COVID-19 evolves.

The quality of the primary studies included in this review continues to impact the overall robustness of the review. Several studies failed to describe their participants (e.g. recruitment method), the details of reference standard conduct used for identifying COVID-19 cases, and the definition used for positivity of the imaging tests. Furthermore, of the studies that described reference standard conduct, one study used a composite reference standard including index test findings, which creates the risk of incorporation bias. Future studies need to prioritise scientific rigour and completeness of reporting and we encourage investigators to refer to the STARD 2015 checklist (Bossuyt 2015; Hong 2018).

A limitation of primary studies that may not be captured in our risk of bias evaluation concerns the recruitment of participants. Of the studies that did report recruitment methods, the majority reported including 'consecutive' participants. However, many of these studies did not actually recruit 'consecutive' participants that represent the target population (i.e. individuals suspected of having COVID-19), but instead included all consecutive participants that underwent an imaging test and RT-PCR testing. These studies did not describe whether all suspected patients in the recruitment setting underwent both an imaging test and RT-PCR as a part of standard practice (which would result in a true 'consecutive' recruitment), or whether imaging tests were only performed in patients with specific clinical signs (e.g. severe symptoms). In studies where the latter situation is present, included participants may not represent the target population, and this could create bias.

We recommend that the accuracy estimates reported in this review are interpreted with caution because of the use of RT-PCR as the reference standard. The results of RT-PCR are not always sensitive, and it is possible that chest CT may be more sensitive than the reference standard in some patients. However, our investigations of heterogeneity for chest CT studies did not identify different accuracy estimates between studies that used at least two RT-PCR test results to define disease-negative status versus studies that used only one RT-PCR test result to define disease-negative status. At this stage, despite its limitations, RT-PCR remains the best tool for diagnosing COVID-19. In future updates of this review, we may consider the use of a latent-class bivariate model for meta-analysis, which adjusts for the imperfect accuracy of the reference standard (Butler-Laporte 2021).

Three out of 51 included studies (6%) were only available as preprints at the time of the search. We will update data extracted from these studies in future versions of our review as these studies become published in peer-reviewed journals.

Applicability of findings to the review question

As the studies in our cohort included suspected COVID-19 participants, our findings are applicable to individuals suspected to have COVID-19. Our search did not identify many studies that evaluated the accuracy of chest CT, ultrasound of the lungs, and chest X-ray for the diagnosis of COVID-19 in paediatric populations. Thus, the diagnostic accuracy of these modalities in children is not as well established. In addition, the lack of data available in the included studies pertaining to signs and symptoms of presenting cases, the severity of the symptoms, as well as timing of symptom onset adds complexity to the interpretation of the findings in this review.

AUTHORS' CONCLUSIONS

Implications for practice

The uncertainty resulting from high or unclear risk of bias and the heterogeneity of included studies limit our ability to confidently draw conclusions based on our results. Our findings indicate that chest computed tomography (CT), chest X-ray and ultrasound all give higher proportions of positive results for individuals with COVID-19 as compared to those without. For chest CT, the chances of getting a positive result are 87.9% (95% CI 84.6 to 90.6) in individuals with COVID-19 and 20.0% (95% CI 15.7 to 25.1) in those without. For chest X-ray, the chances of getting a positive result are 80.6% (95% CI 69.1 to 88.6) in individuals with COVID-19 and 28.5% (95% CI 19.2 to 40.2) in those without. For ultrasound of the lungs, the chances of getting a positive result are 86.4% (95% CI 72.7 to 93.9) in individuals with COVID-19 and 45.4% (95% CI 27.4 to 64.7) in those without. Due to the limited availability of data, accuracy estimates of chest X-ray and ultrasound of the lungs for the diagnosis of COVID-19 in suspected participants should be carefully interpreted.

Implications for research

From our current pool of included studies, we can draw limited conclusions regarding the diagnostic performance of thoracic imaging modalities. Additional studies evaluating the accuracy of COVID-19 diagnosis in suspected patients are needed to allow for more reliable findings.

In this update, we were unable to assess several secondary objectives due to the lack of available data required to formally evaluate direct comparisons of different imaging modalities, and the effect of time since onset of symptoms on the diagnostic performance of various index tests. Future studies should ideally pre-define positive imaging findings and include direct comparisons of the various modalities of interest on the same participant population in order to provide robust and reliable data. Furthermore, improved transparency and reporting is necessary for more efficient data extraction in our updated versions of this review. We encourage authors and investigators to refer to the STARD 2015 checklist (Bossuyt 2015; Hong 2018) to ensure that any relevant information is clearly reported in their studies.

We hope that future updates of this review include more informative studies to allow for additional investigations of variability with improved power and further evaluations of secondary objectives.

ACKNOWLEDGEMENTS

Members of the Cochrane COVID-19 Diagnostic Test Accuracy Review Group include the following.

- The project team (Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Leeflang MMG, Spijker R, Hooft L, Van den Bruel A, McInnes MDF, Emperador D, Dittrich S);
- The systematic review teams for each review
  - Molecular, antigen, and antibody tests (Adriano A, Beese S, Dretzke J, Ferrante di Ruffano L, Harris I, Price M, Taylor-Phillips S)
  - Signs and symptoms (Stuyf T, Domen J, Horn S)
  - Routine laboratory markers (Yang B, Langendam M, Ochodo E, Guleid F, Holtman G, Verbakel J, Wang J, Stegeman I)
  - Imaging tests (Islam N, Ebrahimzadeh S, Salameh JP, Kazi S, Fabiano N, Treanor L, Absi M, Hallgrimson Z, Van Der Pol CB, Prager R, Hare SS, Dennie C, Jenniskens K, Korevaar DA, Cohen JF, van de Wijgert J, Damen JAAG, Wang J)
- The wider team of systematic reviewers from University of Birmingham, UK who assisted with title and abstract screening across the entire suite of reviews for the diagnosis of COVID-19 (Agarwal R, Baldwin S, Berhane S, Herd C, Kristunas C, Quinn L, Scholefield B).

We thank Dr Jane Cunningham (World Health Organization) for participation in technical discussions and comments on the manuscript.

The Cochrane Central Editorial Service collaborated with Cochrane Infectious Diseases Group (CIDG) on the management of the editorial process. We thank Helen Wakeford and Anne-Marie Stephani (Managing Editors, Central Editorial Service, Cochrane) and Deirdre Walshe (CIDG) for editorial feedback; Gianni Virgili (Contact Editor), Sophie Beese and Bella Harris (Managing Editors) and Marta Roqué (statistical peer reviewer) from the Cochrane Diagnostic Test Accuracy Reviews Editorial Team; Robin Featherstone (Central Editorial Service, Cochrane) for search peer review; peer reviewers Paul Garner and Robert Walton; consumer referee Shirley Hall; and Denise Mitchell (Central Editorial Service, Cochrane), who copy-edited the review.

The CIDG editorial base is funded by UK aid from the UK government for the benefit of low- and middle-income countries (project number 300342-104). The views expressed do not necessarily reflect the UK government's official policies.

Jonathan Deeks is a UK National Institute for Health Research (NIHR) Senior Investigator Emeritus. Yemisi Takwoingi is supported by a NIHR Postdoctoral Fellowship. Jonathan Deeks, Jacqueline Dinnes, and Yemisi Takwoingi are supported by the NIHR Birmingham Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

REFERENCES

References to studies included in this review

Ai 2020a *{published data only}*

Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology* 2020;**296**(2):E32-E40.

Aslan 2020 *{published data only}***10.5152/dir.2020.20350**

Aslan S, Bekci T, Cakir IM, Ekiz M, Yavuz I, Sahin AM. Diagnostic performance of low-dose chest CT to detect COVID-19: a Turkish population study. *Diagnostic and Interventional Radiology* 2020. [DOI: 10.5152/dir.2020.20350]

Bar 2020 *{published data only}*

Bar S, Lecourtois A, Diouf M, Goldberg E, Bourbon C, Arnaud E, et al. The association of lung ultrasound images with COVID-19 infection in an emergency room cohort. *Anaesthesia* 2020;**75**(12):1620-25.

Barbosa 2020 *{published data only}*

Barbosa PN, Bitencourt AG, de Miranda GD, Almeida MF, Chojniak R. Chest CT accuracy in the diagnosis of SARS-CoV-2 infection: initial experience in a cancer center. *Radiologia Brasileira* 2020;**53**(4):211-5.

Bellini 2020 *{published data only}***10.1007/s00330-020-07273-y**

Bellini D, Panvini N, Rengo M, Vicini S, Lichtner M, Tieghi T, et al. Diagnostic accuracy and interobserver variability of CO-RADS in patients with suspected coronavirus disease-2019: a multireader validation study. *European Radiology* 2020. [DOI: 10.1007/s00330-020-07273-y]

Besutti 2020 *{published data only}*

Besutti G, Giorgi Rossi P, Iotti V, Spaggiari L, Bonacini R, Nitrosi A, et al. Accuracy of CT in a cohort of symptomatic patients with suspected COVID-19 pneumonia during the outbreak peak in Italy. *European Radiology* 2020 Dec;**30**(12):6818-27.

Borakati 2020 *{published data only}*

Borakati A, Perera A, Johnson J, Sood T. Chest X-ray has poor diagnostic accuracy and prognostic significance in COVID-19: a propensity matched database study. *medRxiv [Preprint]* 2020. [DOI: DOI: 10.1101/2020.07.07.20147934]

Cartocci 2020 *{published data only}*

Cartocci G, Colaiacomo MC, Lanciotti S, Andreoli C, De Cicco ML, Brachetti G, et al. Chest CT for early detection and management of coronavirus disease (COVID-19): a report of 314 patients admitted to Emergency Department with suspected pneumonia. *Radiologia Medica* 2020;**125**(10):931-42.

Caruso 2020 *{published data only}*

Caruso D, Zerunian M, Polici M, Pucciarelli F, Polidori T, Rucci C, et al. Chest CT features of COVID-19 in Rome, Italy. *Radiology* 2020;**2962**(2):E79-E85.

Choudhury 2020 *{published data only}*

Choudhury RS, Shahi PK, Sharma S, Dhar R. Utility of chest radiography on admission for initial triaging of COVID-19 in symptomatic patients. *ERJ Open Research* 2020;**6**(3):00357-2020.

Cozzi 2020 *{published data only}*

Cozzi A, Schiaffino S, Arpaia F, Della Pepa G, Tritella S, Bertolotti P, et al. Chest x-ray in the COVID-19 pandemic: Radiologists' real-world reader performance. *European Journal of Radiology* 2020 Nov;**132**:109272.

Debray 2020 *{published data only}*

Debray M-P, Tarabay H, Males L, Mahdjoub E, Pavlovsky T. Observer agreement and clinical significance of chest CT reporting in patients suspected of COVID-19. *medRxiv [Preprint]* 2020. [DOI: doi.org/10.1101/2020.05.07.20094102]

Deng 2020 *{published data only}*

Deng Z, Zhang X, Li Y, Xu H, Gang Y, Wang H, et al. Value of chest CT screening in the early COVID-19 outbreak. *Chinese Journal of Radiology* 2020;**5**(54):430-4.

De Smet 2020 *{published data only}*

Smet KD, Smet DD, Demedts I, Bouckaert B, Ryckaert T, Laridon E, et al. Diagnostic power of chest CT for COVID-19: to screen or not to screen. *medRxiv [Preprint]* 2020. [DOI: doi.org/10.1101/2020.05.18.20097444]

Dini 2020 *{published data only}***10.4081/monaldi.2020.1446**

Dini FL, Bergamini C, Allegrini A, Scopelliti M, Secco G, Miccoli M, et al. Bedside wireless lung ultrasound for the evaluation of COVID-19 lung injury in senior nursing home residents. *Monaldi Archives for Chest Disease* 2020;**90**(3). [DOI: 10.4081/monaldi.2020.1446]

Dofferhoff 2020 *{published data only}*

Dofferhoff AS, Swinkels A, Sprong T, Berk Y, Spanbroek M, Nabuurs-Franssen MH, et al. Diagnostic algorithm for COVID-19 at the ER. *Nederlands Tijdschrift voor Geneeskunde* 2020;**164**:D5042.

Ducray 2020 *{published data only}***10.1007/s00330-020-07154-4**

Ducray V, Vlachomitrou AS, Bouscambert-Duchamp M, Si-Mohamed S, Gouttard S, Mansuy A, et al. Chest CT for rapid triage of patients in multiple emergency departments during COVID-19 epidemic: experience report from a large French university hospital. *European Radiology* 2020;**31**(2):795-803. [DOI: 10.1007/s00330-020-07154-4]

Falaschi 2020 *{published data only}*

Falaschi Z, Danna PS, Arioli R, Pasché A, Zagaria D, Percivale I, et al. Chest CT accuracy in diagnosing COVID-19 during the peak of the Italian epidemic: a retrospective correlation with RT-PCR testing and analysis of discordant cases. *European Journal of Radiology* 2020;**130**:109192.

Fonsi 2020 *{published data only}*

Fonsi GB, Sapienza P, Brachini G, Andreoli C, De Cicco ML, Cirillo B, et al. Is lung ultrasound imaging a worthwhile procedure for severe acute respiratory syndrome coronavirus 2 pneumonia detection? *Journal of Ultrasound in Medicine* 2020. [DOI: 10.1002/jum.15487]

Fujioka 2020 *{published data only}*

Fujioka T, Takahashi M, Mori M, Tsuchiya J, Yamaga E, Horii T, et al. Evaluation of the usefulness of CO-RADS for chest CT in patients suspected of having COVID-19. *Diagnostics* 2020;**10**(9):608.

Gezer 2020 *{published data only}*

Gezer NS, Ergan B, Barış MM, Appak Ö, Sayıner AA, Balcı P, et al. COVID-19 S: a new proposal for diagnosis and structured reporting of COVID-19 on computed tomography imaging. *Diagnostic and Intervional Radiology* 2020;**26**(4):315–22.

Giannitto 2020 *{published data only}*

Giannitto C, Sposta FM, Repici A, Vatteroni G, Casiraghi E, Casari E, et al. Chest CT in patients with a moderate or high pretest probability of COVID-19 and negative swab. *Radiologia Medica* 2020;**125**:1260-70.

Gietema 2020 *{published data only}***10.1371/journal.pone.0235844**

Gietema HA, Zelis N, Nobel JM, Lambriks LJ, Van Alphen LB, Oude Lashof AM, et al. CT in relation to RT-PCR in diagnosing COVID-19 in The Netherlands: a prospective study. *PLoS ONE* 2020;**15**(7):e0235844. [DOI: 10.1371/journal.pone.0235844]

Guillo 2020 *{published data only}*

Guillo E, Bedmar Gomez I, Dangeard S, Bennani S, Saab I, Tordjman M, et al. COVID-19 pneumonia: diagnostic and prognostic role of CT based on a retrospective analysis of 214 consecutive patients from Paris, France. *European Journal of Radiology* 2020;**131**:109209.

He 2020 *{published data only}*

He J-L, Luo L, Luo Z-D, Lyu J-X, Ng M-Y, Shen X-P, et al. Diagnostic performance between CT and initial real-time RT-PCR for clinically suspected 2019 coronavirus disease (COVID-19) patients outside Wuhan, China. *Respiratory Medicine* 2020;**168**:105980.

Hermans 2020 *{published data only}*

Hermans JJ, Groen J, Zwets E, Boxma-De Klerk BM, Van Werkhoven JM, Ong DS, et al. Chest CT for triage during COVID-19 on the emergency department: myth or truth? *Emergency Radiology* 2020;**27**(6):641-51.

Hernigou 2020 *{published data only}*

Hernigou J, Cornil F, Poignard A, El Bouchaibi S, Mani J, Naouri JF, et al. Thoracic computerised tomography scans in one hundred eighteen orthopaedic patients during the COVID-19 pandemic: identification of chest lesions; added values; help in managing patients; burden on the computerised tomography scan department. *International Orthopaedics (SICOT)* 2020;**44**(8):1571–80.

Herpe 2020 *{published data only}*

Herpe G, Lederlin M, Naudin M, Ohana M, Chaumoitre K, Gregory J, et al. Efficacy of chest CT for COVID-19 pneumonia in France. *Radiology* 2020;**298**(2):E81-E87.

Hwang 2020 *{published data only}*

Hwang EJ, Kim H, Yoon SH, Goo JM, Park CM. Implementation of a deep learning-based computer-aided detection system for the interpretation of chest radiographs in patients suspected for COVID-19. *Korean Journal of Radiology* 2020;**21**(10):1150.

Ippolito 2020 *{published data only}*

Ippolito D, Pecorelli A, Maino C, Capodaglio C, Mariani I, Giandola T, et al. Diagnostic impact of bedside chest X-ray features of 2019 novel coronavirus in the routine admission at the emergency department: case series from Lombardy region. *European Journal of Radiology* 2020;**129**:109092.

Korevaar 2020 *{published data only}*

Korevaar DA, Kootte RS, Smits LP, Van den Aardweg JG, Bonta PI, Schinkel J, et al. Added value of chest computed tomography in suspected COVID-19: an analysis of 239 patients. *European Respiratory Journal* 2020;**56**(2):2001377.

Krdzalic 2020 *{published data only}*

Krdzalic J, De Jaegere TM, Kwee RM. Diagnostic performance of chest CT in screening patients with suspected COVID-19 infection in a Western population. *Brtish Jounal of Radiology* 2020;**93**(1113):20200643.

Kuzan 2020 *{published data only}***10.5152/dir.2020.20270**

Kuzan TY, Murzoglu Altıntoprak K, et al. A comparison of clinical, laboratory and chest CT findings of laboratory-confirmed and clinically diagnosed COVID-19 patients at first admission. *Diagnostic and Interventional Radiology [Epub ahead of print]* 2020 Sep 2. [DOI: 10.5152/dir.2020.20270]

Li 2020a *{published data only}*

Li Z, Zeng B, Lei P, Liu J, Fan B, Shen Q, et al. Differentiating pneumonia with and without COVID-19 using chest CT images: from qualitative to quantitative. *Journal of X-ray Science and Technology* 2020;**28**(4):583-9.

Luo 2020a *{published data only}*

Luo L, Luo Z, Jia Y, Zhou C, He J, Lyu J, et al. CT differential diagnosis of COVID-19 and non-COVID-19 in symptomatic suspects: a practical scoring method. *BMC Pulmonary Medicine* 2020;**20**(1):129.

Luo 2020b *{published data only}*

Luo N, Zhang H, Zhou Y, Kong Z, Sun W, Huang N, et al. Utility of chest CT in diagnosis of COVID-19 pneumonia. *Diagnostic and Interventional Radiology* 2020;**26**(5):437–42.

Mei 2020 *{published data only}*

Mei X, Lee H-C, Diao K, Huang M, Lin B, Liu C, et al. Artificial intelligence-enabled rapid diagnosis of COVID-19 patients. medRxiv [Preprint] 2020. [DOI: doi.org/10.1101/2020.04.12.20062661]

Miranda Magalhães Santos 2020 *{published data only}*

Miranda Magalhães Santos JM, Paula Alves Fonseca A, Pinheiro Zarattini Anastacio E, Formagio Minenelli F, Furtado de Albuquerque Cavalcanti C, Borges da Silva Teles G. Initial results of the use of a standardized diagnostic criteria for chest computed tomography findings in coronavirus disease 2019. *Journal of Computer Assisted Tomography* 2020;**44**(5):647-51.

Murphy 2020 *{published data only}*

Murphy K, Smits H, Knoops AJ, Korst MB, Samson T, Scholten ET, et al. COVID-19 on chest radiographs: a multireader evaluation of an artificial intelligence system. *Radiology* 2020;**296**(3):E166-72.

Narinx 2020 *{published data only}*

Narinx N, Smismans A, Symons R, Frans J, Demeyere A, Gillis M. Feasibility of using point-of-care lung ultrasound for early triage of COVID-19 patients in the emergency room. *Emergency Radiology* 2020;**27**(6):663-70.

Pare 2020 *{published data only}***10.5811/westjem.2020.5.47743**

Pare J, Camelo I, Mayo K, Leo M, Dugas J, Nelson K, et al. Point-of-care lung ultrasound is more sensitive than chest radiograph for evaluation of COVID-19. *Western Journal of Emergency Medicine* 2020;**21**(4):771-8. [DOI: 10.5811/westjem.2020.5.47743]

Patel 2020 *{published data only}*

Patel M, Chowdhury J, Zheng M, Abramian O, Verga S, Zhao H, et al. High Resolution CHEST CT(HRCT) evaluation in patients hospitalized with COVID-19 infection. *medRxiv [Preprint]* 2020. [DOI: doi.org/10.1101/2020.05.26.20114082]

Peng 2020 *{published data only}*

Peng D, Zhang J, Xu Y, Liu Z, Wu P. Clinical analysis and early differential diagnosis of suspected pediatric patients with 2019 novel coronavirus infection. *medRxiv [Preprint]* 2020. [DOI: doi.org/10.1101/2020.04.07.20057315]

Prokop 2020 *{published data only}*

Prokop M, Van Everdingen W, Van Rees Vellinga T, Quarles van Ufford H, Stöger L, Beenen L, et al. CO-RADS: a categorical CT assessment scheme for patients suspected of having COVID-19—definition and evaluation. *Radiology* 2020;**296**(2):E97–E104.

Schiaffino 2020 *{published data only}*

Schiaffino S, Tritella S, Cozzi A, Carriero S, Blandi L, Ferraris L, et al. Diagnostic performance of chest X-Ray for COVID-19 pneumonia during the SARS-CoV-2 pandemic in Lombardy, Italy. *Journal of Thoracic Imaging* 2020;**35**(4):W105–6.

Schulze-hagen 2020 *{published data only}*

Schulze-Hagen M, Hübel C, Meier-Schroers M, Yüksel C, Sander A, Sähn M, et al. Low-dose chest CT for the diagnosis of COVID-19—a systematic, prospective comparison with PCR. *Deutsches Aerzteblatt Online* 2020;**117**(22-23):389-95. [DOI: 10.3238/arztebl.2020.0389]

Song 2020a *{published data only}*

Song S, Wu F, Liu Y, Jiang H, Xiong F, Guo X, et al. Correlation between chest CT findings and clinical features of 211 COVID-19 suspected patients in Wuhan, China. *Open Forum Infectious Diseases* 2020;**7**(6):ofaa171.

Steuwe 2020 *{published data only}*

Steuwe A, Rademacher C, Valentin B, Köhler M-H, Appel E, Keitel V, et al. Dose-optimised chest computed tomography for diagnosis of coronavirus disease 2019 (COVID-19) - evaluation of image quality and diagnostic impact. *Journal of Radiological Protection* 2020;**40**(3):877-91.

Stevens 2020 *{published data only}*

Stevens BJ. Reporting radiographers' interpretation and use of the British Society of Thoracic Imaging's coding system when reporting COVID-19 chest X-rays. *Radiography* 2020;**S1078817420300985**.

Wang 2020a *{published data only}*

Wang T, Xiong Z, Zhou H, Luo W, Tang H, Liu J. Design, validation, and clinical practice of standardized imaging diagnostic report for COVID-19. *Zhong Nan da Xue Xue Bao. Yi Xue Ban [Journal of Central South University. Medical Sciences]* 2020;**45**(3):229-35.

Xiong 2020 *{published data only}*

Xiong Z, Fu L, Zhou H, Liu JK, Wang AM, Huang Y, et al. Construction and evaluation of a novel diagnosis pathway for 2019-Corona virus disease. *Zhonghua Yi Xue za Zhi* 2020;**100**(16):1223-9.

References to studies excluded from this review

Ai 2020b *{published data only}*

Ai J, Gong J, Xing L, He R, Tian F, Wang J, et al. Analysis of factors associated early diagnosis in coronavirus disease 2019 (COVID-19). *medRxiv [Preprint]* 2020. [DOI: doi.org/10.1101/2020.04.09.20059352]

Ai 2020c *{published data only}*

Ai J, Chen J, Wang Y, Liu X, Fan W, Qu G, et al. The cross-sectional study of hospitalized coronavirus disease 2019 patients in Xiangyang, Hubei province. *medRxiv [Preprint]* 2020. [DOI: 10.1101/2020.02.19.20025023]

Arentz 2020 *{published data only}*

Arentz M, Yim E, Klaff L, Lokhandwala S, Riedo FX, Chong M, et al. Characteristics and outcomes of 21 critically ill patients with COVID-19 in Washington State. *JAMA* 2020;**323**(16):1612-4.

Bai 2020a *{published data only}*

Bai HX, Hsieh B, Xiong Z, Halsey K, Choi JW, Tran TM, et al. Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT. *Radiology* 2020;**296**(2):E46-E54.

Bai 2020b *{published data only}*

Bai HX, Wang R, Xiong Z, Hsieh B, Chang K, Halsey K, et al. Artificial intelligence augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other origin at chest CT. *Radiology* 2020;**296**(3):E156-65.

Chang 2020 *{published data only}*

Chang D, Lin M, Wei L, Xie L, Zhu G, Dela Cruz CS, et al. Epidemiologic and clinical characteristics of novel coronavirus infections involving 13 patients outside Wuhan, China. *JAMA* 2020;**323**(11):1092-3. [DOI: 10.1001/jama.2020.1623 10.1001/jama.2020.1623.]

Chen 2020a *{published data only}*

Chen S, Wu JJ, Li MZ, Xu DZ, Zhu YZ, Wang HC, et al. Clinical features of 109 cases of novel coronavirus pneumonia. *Chinese Journal of Infectious Diseases* 2020;**38**(12):E015. [DOI: 10.3760 / cma.j.issn.1000-6680.2020.0015]

Chen 2020b *{published data only}*

Chen X, Tang Y, Mo Y, Li S, Lin D, Yang Z, et al. A diagnostic model for coronavirus disease 2019 (COVID-19) based on radiological semantic and clinical features: a multi-center study. *European Radiolology* 2020;**30**(9):4893-902. [DOI: 10.1007/s00330-020-06829-2]

Chen 2020c *{published data only}*

Chen Z, Fan H, Cai J, Li Y, Wu B, Hou Y, et al. High-resolution computed tomography manifestations of COVID-19 infections in patients of different ages. *European Journal of Radiology* 2020;**126**:108972.

Cheng 2020 *{published data only}*

Cheng Z, Lu Y, Cao Q, Qin L, Pan Z, Yan F, et al. Clinical features and chest CT manifestations of coronavirus disease 2019 (COVID-19) in a single-center study in Shanghai, China. *American Journal of Roentgenology* 2020;**215**(1):121-6. [DOI: 10.2214/ajr.20.22959 10.2214/AJR.20.22959.]

Çinkooğlu 2020 *{published data only}*

Çinkooğlu A, Bayraktaroğlu S, Savaş R. Lung changes on chest CT during 2019 novel coronavirus (COVID-19) pneumonia. *European Journal of Breast Health* 2020;**16**(2):89-90. [DOI: 10.5152/ejbh.2020.010420]

Colombi 2020 *{published data only}*

Colombi D, Bodin FC, Petrini M, Maffi G, Morelli N, Milanese G, et al. Well-aerated lung on admitting chest CT to predict adverse outcome in COVID-19 pneumonia. *Radiology* 2020;**296**(2):E86-E96. [DOI: 10.1148/radiol.2020201433]

Dai 2020 *{published data only}*

Dai H, Zhang X, Xia J, Zhang T, Shang Y, Huang R, et al. High-resolution chest CT features and clinical characteristics of patients infected with COVID-19 in Jiangsu, China. *International Journal of Infectious Diseases* 2020;**95**:106-12. [DOI: S1201-9712(20)30218-6]

Ding 2020 *{published data only}*

Ding X, Xu J, Zhou J, Long Q. Chest CT findings of COVID-19 pneumonia by duration of symptoms. *European Journal of Radiology* 2020;**127**:109009. [DOI: 10.1016/j.ejrad.2020.109009]

Dong 2020 *{published data only}*

Dong J, Wu L, Jin Q, Chen J, He J. Chest CT scan of hospitalized patients with COVID-19: a case-control study. *medRxiv [Preprint]* 2020. [DOI: doi.org/10.1101/2020.04.07.20056762]

Guan 2020 *{published data only}*

Guan W, Liu J, Yu C. CT findings of coronavirus disease (COVID-19) severe pneumonia. *American Journal of Roentgenology* 2020;**5**:W85-W6. [DOI: 10.2214/AJR.20.23035]

Hao 2020 *{published data only}*

Hao W, Li M. Clinical diagnostic value of CT imaging in COVID-19 with multiple negative RT-PCR testing. *Travel Medicine and Infectious Disease* 2020;**34**:101627. [DOI: 10.1016/j.tmaid.2020.101627]

Himoto 2020 *{published data only}*

Himoto Y, Sakata A, Kirita M, Hiroi T, Kobayashi K I, Kubo K, et al. Diagnostic performance of chest CT to differentiate COVID-19 pneumonia in non-high-epidemic area in Japan. *Japanese Journal of Radiology* 2020;**38**(5):400-6.

Huang 2020 *{published data only}*

Huang P, Liu T, Huang L, Liu H, Lei M, Xu W, et al. Use of chest CT in combination with negative RT-PCR assay for the 2019 novel coronavirus but high clinical suspicion. *Radiology* 2020;**295**(1):22-3. [DOI: 10.1148/radiol.2020200330 10.1148/radiol.2020200330. Epub 2020 Feb 12.]

Liang 2020 *{published data only}*

Liang Y, Liang J, Zhou Q, Li X, Lin F, Deng Z, et al. Prevalence and clinical features of 2019 novel coronavirus disease (COVID-19) in the Fever Clinic of a teaching hospital in Beijing: a single-center, retrospective study. *medRxiv [Preprint]* 2020. [DOI: doi.org/10.1101/2020.02.25.20027763]

Lu 2020 *{published data only}*

Lu X, Zhang L, Du H, Zhang J, Li Y, Qu J, et al. SARS-CoV-2 infection in children. *New England Journal of Medicine* 2020;**382**(17):1663-5. [DOI: 10.1056/NEJMc2005073]

Mao 2020 *{published data only}*

Mao X, Liu X-P, Xiong M, Yang X, Jin X, Li Z, et al. Development and validation of chest CT-based imaging biomarkers for early stage COVID-19 screening. *medRxiv [Preprint]* 2020. [DOI: doi.org/10.1101/2020.05.15.20103473]

Miao 2020a *{published data only}*

Miao C, Zhuang J, Jin M, Xiong H, Huang P, Zhao Q, et al. A comparative multi-centre study on the clinical and imaging features of confirmed and unconfirmed patients with COVID-19. *medRxiv [Preprint]* 2020. [DOI: doi.org/10.1101/2020.03.22.20040782]

Miao 2020b *{published data only}*

Miao C, Jin M, Miao L, Yang X, Huang P, Xiong H, et al. Early chest computed tomography to diagnose COVID-19 from suspected patients: a multicenter retrospective study. *American Journal of Emergency Medicine* 2020;**S0735-6757**(20):30281-3.

Pakray 2020 *{published data only}*

Pakray A, Walker D, Figacz A, Kilanowski S, Rhodes C, Doshi S, et al. Imaging evaluation of COVID-19 in the emergency department. *Emergency Radiology* 2020;**27**:579-88.

Poggiali 2020 *{published data only}*

Poggiali E, Dacrema A, Bastoni D, Tinelli V, Demichele E, Mateo Ramos P, et al. Can lung US help critical care clinicians in the early diagnosis of novel coronavirus (COVID-19) pneumonia? *Radiology* 2020;**295**(E6):200847. [DOI: 10.1148/radiol.2020200847]

Pu 2020 *{published data only}*

Pu J, Leader J, Bandos A, Shi J, Du P, Yu J, et al. Any unique image biomarkers associated with COVID-19? *European Radiology* 2020;**30**(NA):6221–7.

Siegel 2020 *{published data only}*

Siegel A, Chang PJ, Jarou ZJ, Paushter DM, Harmath CB, Arevalo JB, et al. Lung base findings of coronavirus disease (COVID-19) on abdominal CT in patients with predominant gastrointestinal symptoms. *American Journal of Roentgenology* 2020;**215**(3):607-9. [DOI: 10.2214/AJR.20.23232]

Song 2020b *{published data only}*

Song F, Shi N, Shan F, Zhang Z, Shen J, Lu H, et al. Emerging 2019 novel Coronavirus (2019-nCoV) pneumonia. *Radiology* 2020;**295**(1):210-7. [DOI: 10.1148/radiol.2020200274 10.1148/radiol.2020200274. Epub 2020 Feb 6.]

Tavare 2020 *{published data only}*

Tavare AN, Braddy A, Brill S, Jarvis H, Sivaramakrishnan A, Barnett J, et al. Managing high clinical suspicion COVID-19 inpatients with negative RT-PCR: a pragmatic and limited role for thoracic CT. *Thorax* 2020;**75**(7):537-8. [DOI: thoraxjnl-2020-214916]

Wang 2020b *{published data only}*

Wang Y, Dong C, Hu Y, Li C, Ren Q, Zhang X, et al. Temporal changes of CT findings in 90 patients with COVID-19 pneumonia: a longitudinal study. *Radiology* 2020;**296**(2):E55-E64. [DOI: 10.1148/radiol.2020200843 10.1148/radiol.2020200843.]

Wu 2020a *{published data only}*

Wu J, Feng CL, Xian XY, Qiang J, Zhang J, Mao QX, et al. Novel coronavirus pneumonia (COVID-19) CT distribution and sign features. *Zhonghua Jie He He Hu Xi za Zhi [Chinese Journal of Tuberculosis and Respiratory Diseases]* 2020;**43**(4):321-6. [DOI: 10.3760/cma.j.cn112147-20200217-00106 10.3760/cma.j.cn112147-20200217-00106.]

Wu 2020b *{published data only}*

Wu Q, Xing Y, Shi L, Li W, Gao Y, Pan S, et al. Epidemiological and clinical characteristics of children with coronavirus disease 2019. *medRxiv [Preprint]* 2020. [DOI: 10.1101/2020.03.19.20027078]

Wu 2020c *{published data only}*

Wu X, Sun R, Chen J, Xie Y, Zhang S, Wang X. Radiological findings and clinical characteristics of pregnant women with COVID-19 pneumonia. *International Journal of Gynaecology and Obstetrics* 2020;**150**(1):58-63. [DOI: 10.1002/ijgo.13165]

Wu 2020d *{published data only}*

Wu X, Fu B, Chen L, Feng Y. Serological tests facilitate identification of asymptomatic SARS-CoV-2 infection in Wuhan, China. *Journal of Medical Virology* 2020;**10**:1002/jmv.25904. [DOI: 10.1002/jmv.25904]

Xie 2020 *{published data only}*

Xie S, Lei Z, Chen X, Liu W, Wang X, Dong Y, et al. Chest CT-based differential diagnosis of 28 patients with suspected corona virus disease 2019 (COVID-19). *British Journal of Radiology* 2020;**93**(1112):20200243.

Xu 2020a *{published data only}*

Xu X, Yu C, Qu J, Zhang L, Jiang S, Huang D, et al. Imaging and clinical features of patients with 2019 novel coronavirus SARS-CoV-2. *European Journal of Nuclear Medicine and Molecular Imaging* 2020;**47**(5):1275-80. [DOI: 10.1007/s00259-020-04735-9 10.1007/s00259-020-04735-9.]

Xu 2020b *{published data only}*

Xu X, Yu C, Zhang L, Luo L, Liu J. Imaging features of 2019 novel coronavirus pneumonia. *European Journal of Nuclear Medicine and Molecular Imaging* 2020;**47**(5):1022-23. [DOI: 10.1007/s00259-020-04720-2]

Yang 2020a *{published data only}*

Yang S, Shi Y, Lu H, Xu J, Li F, Qian Z, et al. Clinical and CT features of early-stage patients with COVID-19: a retrospective analysis of imported cases in Shanghai, China. *European Respiratory Journal* 2020;**55**(4):2000407. [DOI: 10.1183/13993003.00407-2020]

Yang 2020b *{published data only}*

Yang W, Cao Q, Qin L, Wang X, Cheng Z, Pan A, et al. Clinical characteristics and imaging manifestations of the 2019 novel coronavirus disease (COVID-19): a multi-center study in Wenzhou city, Zhejiang, China. *Journal of Infection* 2020;**80**(4):388-93.

Yuan 2020 *{published data only}*

Yuan M, Yin W, Tao Z, Tan W, Hu Y. Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China. *PloS One* 2020;**15**(3):e0230548. [DOI: 10.1371/journal.pone.0230548 10.1371/journal.pone.0230548. eCollection 2020.]

Zhifeng 2020 *{published data only}*

Zhifeng J, Feng A, Li T. Consistency analysis of COVID-19 nucleic acid tests and the changes of lung CT. *Journal of Clinical Virology* 2020;**127**:104359.

Additional references

Bossuyt 2015

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. *Radiology* 2015 Oct 28;**277**(3):826–32. [DOI: 10.1148/radiol.2015151516]

BSTI 2020

British Society of Thoracic Imaging. COVID-19 BSTI Reporting Templates and Codes. bsti.org.uk/covid-19-resources/covid-19-bsti-reporting-templates/ 2020 (accessed 4 January 2020).

Butler-Laporte 2021

Butler-Laporte G, Lawandi A, Schiller I, Yao MC, Dendukuri N, McDonald E, et al. Comparison of saliva and nasopharyngeal swab nucleic acid amplification testing for detection of SARS-CoV-2: a systematic review and meta-analysis. *JAMA Internal Medicine* 2021:E1-8. [DOI: 10.1001/jamainternmed.2020.8876]

China National Health Comission 2020

China National Health Comission. Chinese Clinical Guidance for COVID-19 Pneumonia Diagnosis and Treatment (7th Edition). kjfy.meetingchina.org/msite/news/show/cn/3337.html 2020 (accessed 24 January 2020).

Chu 2006

Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology* 2006;**59**(12):1331-2; author reply 1332-3. [DOI: 10.1016/j.jclinepi.2006.06.011]

Cohen 2016

Cohen JF, Korevaar DA, Wang J, Leeflang MM, Bossuyt PM. Meta-epidemiologic study showed frequent time trends in summary estimates from meta-analyses of diagnostic accuracy studies. *Journal of Clinical Epidemiology* 2016;**77**:60–7. [DOI: 10.1016/j.jclinepi.2016.04.013]

Datta 2020

Datta SD, Talwar A, Lee JT. A proposed framework and timeline of the spectrum of disease due to SARS-CoV-2 infection: illness beyond acute infection and public health implications. *JAMA* 2020;**324**(22):2251-2. [DOI: 10.1001/jama.2020.22717]

Deeks 2020

Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Spijker R, Taylor-Phillips S, et al. Antibody tests for identification of current and past infection with SARS-CoV-2. *Cochrane Database of Systematic Reviews* 2020, Issue 6. Art. No: CD013652. [DOI: 10.1002/14651858.CD013652]

Dinnes 2020

Dinnes J, Deeks JJ, Adriano A, Berhane S, Davenport C, Dittrich S, et al. Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database of Systematic Reviews* 2020, Issue 8. Art. No: CD013705. [DOI: 10.1002/14651858.CD013705]

Harbord 2009

Harbord RM, Whiting P. metandi: meta-analysis of diagnostic accuracy using hierarchical logistic regression. *Stata Journal* 2009;**9**(2):211-29. [DOI: doi.org/10.1177/1536867X0900900203]

Hong 2018

Hong PJ, Korevaar DA, McGrath TA, Ziai H, Frank R, Alabousi M, et al. Reporting of imaging diagnostic accuracy studies with focus on MRI subgroup: adherence to STARD 2015. *Journal of Magnetic Resonance Imaging* 2018;**47**(2):523-44. [DOI: 10.1002/jmri.25797]

Irwig 1995

Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *Journal of Clinical Epidemiology* 1995;**48**(1):119-30; discussion 131-2. [DOI: 10.1016/0895-4356(94)00099-c]

Kucirka 2020

Kucirka LM, Lauer SA, Laeyendecker O, Boon D, Lessler J. Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based SARS-CoV-2 tests by time since exposure. *Annals of Internal Medicine* 2020;**173**(4):262-7. [DOI: 10.7326/M20-1495]

Lau 1992

Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *New England Journal of Medicine* 1992;**327**(4):248–54. [DOI: 10.1056/NEJM199207233270406]

Li 2020b

Li Y, Yao L, Li J, Chen L, Song Y, Cai Z, et al. Stability issues of RT-PCR testing of SARS-CoV-2 for hospitalized patients clinically diagnosed with COVID-19. *Journal of Medical Virology* 2020;**92**(7):903-8. [DOI: 10.1002/jmv.25786]

Loeffelholz 2020

Loeffelholz MJ, Tang YW. Laboratory diagnosis of emerging human coronavirus infections - the state of the art. *Emerging Microbes and Infections* 2020;**9**(1):747-56. [DOI: 10.1080/22221751.2020.1745095]

McGrath 2017

McGrath TA, McInnes MD, Langer FW, Hong J, Korevaar DA, Bossuyt PM. Treatment of multiple test readers in diagnostic accuracy systematic reviews-meta-analyses of imaging studies. *European Journal of Radiology* 2017;**93**:59–64. [DOI: doi.org/10.1016/j.ejrad.2017.05.032]

Moher 2009

Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Journal of Clinical Epidemiology* 2009;**62**(10):1006-12. [DOI: 10.1371/journal.pmed.1000097]

Ng 2020

Ng M-Y, Lee EYP, Yang J, Yang F, Li X, Wang H, et al. Imaging Profile of the COVID-19 Infection: Radiologic Findings and Literature Review. *Cardiothoracic Imaging* 2020 Feb 1;**2**(1):e200034. [DOI: 10.1148/ryct.2020200034]

Pedersen 2020

Pedersen, TL. Package 'ggforce'. www.ggforce.data-imaginist.com 2020 (accessed 5 January 2020).

R Core Team 2021 [Computer program]

R Foundation for Statistical Computing R: A language and environment for statistical computing. R Core Team. Vienna, Austria: R Foundation for Statistical Computing, 2020.

Reitsma 2005

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews.

Journal of Clinical Epidemiology 2005;**58**(10):982-90. [DOI: 10.1016/j.jclinepi.2005.02.022]

Rubin 2020

Rubin GD, Ryerson CJ, Haramati LB, Sverzellati N, Kanne JP, Raoof S, et al. The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the Fleischner Society. *Radiology* 2020;**296**(1):172–80. [DOI: 10.1148/radiol.2020201365]

Salameh 2020b

Salameh J-P, Bossuyt PM, McGrath TA, Thombs BD, Hyde CJ, Macaskill P, et al. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. *BMJ* 2020;**370**:m2632. [DOI: 10.1136/bmj.m2632]

Salehi 2020

Salehi S, Abedi A, Balakrishnan S, Gholamrezanezhad A. Coronavirus Disease 2019 (COVID-19): A Systematic Review of Imaging Findings in 919 Patients. *American Journal of Roentgenology* 2020;**215**(1):87-93. [DOI: 10.2214/AJR.20.23034]

Shi 2020

Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *Lancet Infectious Diseases* 2020 Apr;(4):425–34. [DOI: 10.1016/S1473-3099(20)30086-4]

Simpson 2020

Simpson S, Kay FU, Abbara S, Bhalla S, Chung JH, Chung M, et al. Radiological Society of North America expert consensus document on reporting chest CT findings related to COVID-19: endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA. *Radiology: Cardiothoracic Imaging* 2020;**2**(2):e200152. [DOI: 10.1148/ryct.2020200152]

StataCorp 2019 [Computer program]

Stata Statistical Software. StataCorp, Version Release 16. College Station, TX: StataCorp LLC, 2019.

Stegeman 2020

Stegeman I, Ochodo EA, Guleid F, Holtman GA, Yang B, Davenport C, et al. Routine laboratory testing to determine if a patient has COVID-19. *Cochrane Database of Systematic Reviews* 2020, Issue 11. Art. No: CD013787. [DOI: 10.1002/14651858.CD013787]

Struyf 2020

Struyf T, Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Leeflang MM, et al. Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19 disease. *Cochrane Database of Systematic Reviews* 2020, Issue 7. Art. No: CD013665. [DOI: 10.1002/14651858.CD013665]

Thomas 2010 [Computer program]

EPPI-Centre Software EPPI-Reviewer 4.0: software for research synthesis. Thomas J, Brunton J, Graziosi S. London: Social Science Research Unit, Institute of Education, University of London: EPPI-Centre Software, 2010.

WHO 2020

World Health Organization (WHO). WHO COVID-19 Case definition. WHO/2019-nCoV/Surveillance_Case_Definition/2020.1 2020 (accessed 17 October 2020).

Wickham 2016

Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag, 2016.

References to other published versions of this review

Islam 2020

Islam N, Salameh J-P, Leeflang MM, Hooft L, McGrath TA, Pol CB, et al. Thoracic imaging tests for the diagnosis of COVID-19. *Cochrane Database of Systematic Reviews* 2020, Issue 11. Art. No: CD013639. [DOI: 10.1002/14651858.CD013639.pub3]

McInnes 2020

McInnes MD, Leeflang MM, Salameh J-P, McGrath TA, Pol CB, Frank RA, et al. Imaging tests for the diagnosis of COVID-19. *Cochrane Database of Systematic Reviews* 2020, Issue 6. Art. No: CD013639. [DOI: 10.1002/14651858.CD013639]

Salameh 2020a

Salameh J-P, Leeflang MM, Hooft L, Islam N, McGrath TA, Pol CB, et al. Thoracic imaging tests for the diagnosis of COVID-19. *Cochrane Database of Systematic Reviews* 2020, Issue 9. Art. No: CD013639. [DOI: 10.1002/14651858.CD013639.pub2]

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Ai 2020a

Study characteristics	
Patient Sampling	Study design: suspected patients, unclear symptom status
Patient characteristics and setting	Age group: adults only Setting: unclear

Ai 2020a *(Continued)*

Index tests	Index test(s): chest CT Definition for positive diagnosis on CT: unclear Level of training of readers: radiologist Prevalence: 0.6
Target condition and reference standard(s)	Reference standard: RT-PCR, no other details provided
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		

Ai 2020a *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Aslan 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic		
Patient characteristics and setting	Age group: adults only Setting: outpatient		
Index tests	Index test(s): chest CT (non-contrast, low dose) Definition for positive diagnosis on CT: Radiological evidence of COVID-19 pneumonia, including presence of groundglass opacity (GGO), mixed GGO (GGO and consolidation), consolidation, distribution and number of lobes and segment affected by GGO and/or consolidation, etc. Level of training of readers: radiologist Prevalence: 0.8		
Target condition and reference standard(s)	Reference standard: RT-PCR twice, if necessary		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Aslan 2020 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Bar 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, unclear symptom status		
Patient characteristics and setting	Age group: adults only Setting: outpatient		
Index tests	Index test(s): ultrasound of the lungs (POCUS) Defintion for positive diagnosis on ultrasound: unclear Level of training of readers: unclear Prevalence: 0.3		
Target condition and reference standard(s)	Reference standard: RT-PCR twice, if necessary		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	

Bar 2020 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Barbosa 2020

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic
Patient characteristics and setting	Age group: adults only Setting: unclear
Index tests	Index test(s): chest CT, no further details provided Definition for positive diagnosis on CT: RSNA classification Level of training of readers: radiologist Prevalence: 0.3
Target condition and reference standard(s)	Reference standard: RT-PCR, no other details provided
Flow and timing	
Comparative	
Notes	
Methodological quality	

Barbosa 2020 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Bellini 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic		
Patient characteristics and setting	Age group: children and adults Setting: unclear		
Index tests	Index test(s): chest CT (non-contrast) Definition for positive diagnosis on CT: CO-RADS Level of training of readers: radiologist Prevalence: 0.2		
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some; other (clinical signs on follow-up)		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Bellini 2020 *(Continued)*

DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Besutti 2020

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic
Patient characteristics and setting	Age group: adults, perhaps also children Setting: outpatient
Index tests	Index test(s): chest CT (non-contrast) Definition for positive diagnosis on CT: a structured report about the probability of COVID-19 pneumonia Level of training of readers: radiologist Prevalence: 0.9
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some
Flow and timing	
Comparative	
Notes	

Besutti 2020 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		

Besutti 2020 *(Continued)*

Could the patient flow have introduced bias?	Low risk

Borakati 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, symptomatic or asymptomatic		
Patient characteristics and setting	Age group: adults, perhaps also children Setting: outpatient		
Index tests	Index test(s): chest CT (non-contrast, IV contrast); chest x-rays Definition for positive diagnosis (both CT and x-ray): BSTI template Level of training of readers: radiologist Prevalence: 0.6		
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		

Borakati 2020 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Cartocci 2020

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic
Patient characteristics and setting	Age group: adults only Setting: outpatient

Cartocci 2020 *(Continued)*

Index tests	Index test(s): chest CT, no further details provided Definition for positive diagnosis on CT: classification system by Simpson et al.: typical CT pattern, possible CT pattern, inconsistent CT pattern, negative for pneumonia (Simpson 2020) Level of training of readers: radiologist Prevalence: 0.5
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		

Cartocci 2020 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Caruso 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic		
Patient characteristics and setting	Age group: adults only Setting: outpatient		
Index tests	Index test(s): chest CT (non-contrast) Definition for positive diagnosis on CT: pneumonia Level of training of readers: radiologist Prevalence: 0.4		
Target condition and reference standard(s)	Reference standard: RT-PCR twice, if necessary		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		

Caruso 2020 *(Continued)*

Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Choudhury 2020

Study characteristics			

Choudhury 2020 *(Continued)*

Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic
Patient characteristics and setting	Age group: unclear Settinng: inpatient
Index tests	Index test(s): chest X-rays, no further details provided Defintion for positive diagnosis: a previously unvalidated Likert score (scores 1 to 5) based on radiographic features thought to be related to COVID-19, based on format reported by Simspon et al. (Simpson 2020) Level of training of readers: unclear Prevalence: 0.3
Target condition and reference standard(s)	Reference standard: RT-PCR, no other details provided
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
DOMAIN 2: Index Test (Chest X-ray)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Choudhury 2020 *(Continued)*

DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Cozzi 2020

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, symptomatic or asymptomatic
Patient characteristics and setting	Age group: unclear Setting: outpatient
Index tests	Index test(s): chest X-rays Definition for positive diagnosis on X-ray: the presence of interstitial infiltrates with predominantly bilateral and basal distribution Level of training of readers: radiologist Prevalence: 0.8
Target condition and reference standard(s)	Reference standard: RT-PCR, no other details provided; other (follow-up phone call)
Flow and timing	
Comparative	
Notes	

Cozzi 2020 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
DOMAIN 2: Index Test (Chest X-ray)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		

Cozzi 2020 *(Continued)*

Could the patient flow have introduced bias?	Unclear risk

Debray 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, unclear symptom status		
Patient characteristics and setting	Age group: adults only Setting: outpatient		
Index tests	Index test(s): chest CT (non-contrast) Defintion for positive diagnosis on CT: quote: “evocative”: multifocal ground-glass opacities, being nodular or not, or crazy-paving with or without consolidations, with a bilateral, peripheral or mixed distribution and involvement of the posterior zones Level of training of readers: radiologist Prevalence: 0.7		
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Debray 2020 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Deng 2020

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic
Patient characteristics and setting	Age group: children and adults Setting: unclear
Index tests	Index test(s): chest CT (high resolution) Defintion for positive diagnosis on CT: 1. any one of the following: a. single, multiple, or diffuse GGO, with thickened blood vessels and thickened bronchial shadows passing through, with or without localised lobular septal grid thickening b. single or multiple real shadows

Deng 2020 *(Continued)*

	2. re-examination 3-5 days later showed that the original GGO or consolidation range increased, the number increased, or accompanied by pleural effusion on one or both sides Level of training of readers: radiologist Prevalence: 0.7
Target condition and reference standard(s)	Reference standard: RT-PCR once
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		

Deng 2020 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

De Smet 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic		
Patient characteristics and setting	Age group: children and adults Setting: outpatient		
Index tests	Index test(s): chest CT, no further details provided Defintion for positive diagnosis on CT: CO-RADS Level of training of readers: unclear Prevalence: 0.4		
Target condition and reference standard(s)	Reference standard: RT-PCR, no other details provided		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		

De Smet 2020 *(Continued)*

Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Dini 2020

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, symptomatic or asymptomatic
Patient characteristics and setting	Age group: ≥ 70 years of age

Dini 2020 *(Continued)*

	Setting: outpatient
Index tests	Index test(s): ultrasound of lungs (POCUS); no further details provided Definition for positive diagnosis on ultrasound: scoring system: non-coalescent B-lines, coalescent and with hyperechoic non-consolidated state Level of training of readers: unclear Prevalence: 0.6
Target condition and reference standard(s)	Reference standard: RT-PCR, no other details provided
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			

Dini 2020 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Dofferhoff 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, symptomatic or asymptomatic		
Patient characteristics and setting	Age group: adults only Setting: outpatient		
Index tests	Index test(s): chest CT (low dose) Defintion for positive diagnosis on CT: CO-RADS Level of training of readers: unclear Prevalence: 0.5		
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Dofferhoff 2020 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Ducray 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, symptomatic or asymptomatic		
Patient characteristics and setting	Age group: adults only Setting: outpatient		
Index tests	Index test(s): chest CT (IV contrast) Defintion for positive diagnosis on CT: classification system: surely COVID+, possible COVID+, COVID- Level of training of readers: radiologist Prevalence: 0.4		
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			

Ducray 2020 *(Continued)*

DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Falaschi 2020

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic
Patient characteristics and setting	Age group: adults only Setting: outpatient
Index tests	Index test(s): chest CT (non-contrast) Defintion for positive diagnosis on CT: STR/ACR/RSNA Level of training of readers: radiologist Prevalence: 0.6
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some
Flow and timing	
Comparative	
Notes	
Methodological quality	

Falaschi 2020 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Fonsi 2020

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic
Patient characteristics and setting	Age group: adults only Setting: outpatient
Index tests	Index test(s): chest CT (non-contrast); ultrasound of lungs (POCUS) Defintion for positive diagnosis on CT: GGOs; consolidation; a mixed GGO and consolidation pattern; single or multiple solid nodules surrounded by GGOs; a focal or multifocal distribution; GGO and consolidation location; multilobe involvement; a bilateral distribution; interlobular septal thickening; an air bronchogram; the presence of cavitation; bronchial wall thickening; bronchiectasis; mediastinal lymph node enlargement; pleural effusion; and pericardial effusion Definition for positive diagnosis on ultrasound: not reported Level of training of readers: radiologist Prevalence: 0.7
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Fonsi 2020 *(Continued)*

If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Fujioka 2020

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic
Patient characteristics and setting	Age group: adults only

Fujioka 2020 *(Continued)*

	Setting: unclear
Index tests	Index test(s): chest CT, no further details provided Definition for positive diagnosis on CT: CO-RADS Level of training of readers: radiologist Prevalence: 0.5
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		

Fujioka 2020 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Gezer 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic		
Patient characteristics and setting	Age group: adults only Setting: outpatient		
Index tests	Index test(s): chest CT (non-contrast) Defintion for positive diagnosis on CT: unclear Level of training of readers: radiologist Prevalence: 0.4		
Target condition and reference standard(s)	Reference standard: RT-PCR, no other details provided; other (clinical signs and imaging tests)		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		

Gezer 2020 *(Continued)*

Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Giannitto 2020

Study characteristics			

Giannitto 2020 *(Continued)*

Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic
Patient characteristics and setting	Age group: adults only Setting: outpatient
Index tests	Index test(s): chest CT (non-contrast) Definition for positive diagnosis on CT: classification system: suspected COVID-19 pneumonia, non-COVID-19 pneumonia, negative CT Level of training of readers: radiologist Prevalence: 0.3
Target condition and reference standard(s)	Reference standard: RT-PCR twice, if necessary
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			

Giannitto 2020 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Gietema 2020

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic
Patient characteristics and setting	Age group: adults only Setting: outpatient
Index tests	Index test(s): chest CT (non-contrast) Definition for positive diagnosis on CT: standardized imaging reporting system (typical for COVID-19, equivocal, non COVID-19) Level of training of readers: resident Prevalence: 0.4
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

Gietema 2020 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Guillo 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic		
Patient characteristics and setting	Age group: adults only Setting: outpatient		
Index tests	Index test(s): chest CT (IV contrast) Definition for positive diagnosis on CT: a structured report about the probability of COVID-19 pneumonia based on the presence of GGOs with or without crazy-paving pattern, isolated or admixed with perilobular or linear consolidation, their peripheral or central distribution, etc. Level of training of readers: resident Prevalence: 0.6		
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	

Guillo 2020 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

He 2020

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, unclear symptom status
Patient characteristics and setting	Age group: children and adults Setting: unclear
Index tests	Index test(s): chest CT (high-resolution) Defintion for positive diagnosis on CT: GGO with or without consolidation, crazy paving patten, peripheral and diffuse distribution, and bilateral/multilobular involvement Level of training of readers: radiologist Prevalence: 0.4
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some
Flow and timing	

He 2020 *(Continued)*

Comparative
Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		

He 2020 *(Continued)*

Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Hermans 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, symptomatic or asymptomatic		
Patient characteristics and setting	Age group: adults only Setting: outpatient		
Index tests	Index test(s): chest CT, no further details provided Defintion for positive diagnosis on CT: CO-RADS Level of training of readers: radiologist Prevalence: 0.4		
Target condition and reference standard(s)	Reference standard: RT-PCR once		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		

Hermans 2020 *(Continued)*

If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Hernigou 2020

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, symptomatic or asymptomatic
Patient characteristics and setting	Age group: adults only Setting: outpatient
Index tests	Index test(s): chest CT (low dose) Defintion for positive diagnosis on CT: unclear Level of training of readers: radiologist Prevalence: 0.3

Hernigou 2020 *(Continued)*

Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some		
Flow and timing			
Comparative			
Notes			

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern

Hernigou 2020 *(Continued)*

DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Herpe 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic		
Patient characteristics and setting	Age group: children and adults Setting: unclear		
Index tests	Index test(s): chest CT, no further details provided Definition for positive diagnosis on CT: bilateral GGO with peripheral distribution, bilateral crazy paving appearance with intralobular thickening, reverse halo sign, or other signs compatible with organising pneumonia Level of training of readers: radiologist Prevalence: 0.5		
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	

Herpe 2020 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Hwang 2020

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, symptomatic or asymptomatic
Patient characteristics and setting	Age group: adults, perhaps also children Setting: unclear

Hwang 2020 *(Continued)*

Index tests	Index test(s): chest X-rays Definition for positive diagnosis on X-ray: abnormality suggesting pneumonia Level of training of readers: radiologists and resident Prevalence: 0.05
Target condition and reference standard(s)	Reference standard: RT-PCR, no other details provided
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
DOMAIN 2: Index Test (Chest X-ray)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		

Hwang 2020 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Ippolito 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic		
Patient characteristics and setting	Age group: children and adults Setting: outpatient		
Index tests	Index test(s): chest X-rays Defintion for positive diagnosis on X-ray: reticulations, alveolar opacities or both Level of training of readers: radiologist Prevalence: 0.4		
Target condition and reference standard(s)	Reference standard: RT-PCR, no other details provided		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		

Ippolito 2020 *(Continued)*

Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
DOMAIN 2: Index Test (Chest X-ray)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Korevaar 2020

Study characteristics			

Korevaar 2020 *(Continued)*

Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic
Patient characteristics and setting	Age group: adults only Setting: outpatient
Index tests	Index test(s): chest CT (low dose) Defintion for positive diagnosis on CT: CO-RADS Level of training of readers: radiologist Prevalence: 0.5
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			

Korevaar 2020 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Krdzalic 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic		
Patient characteristics and setting	Age group: adults only Setting: unclear		
Index tests	Index test(s): chest CT Defintion for positive diagnosis on CT: CO-RADS Level of training of readers: radiologist Prevalence: 0.5		
Target condition and reference standard(s)	Reference standard: RT-PCR twice, if necessary		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Krdzalic 2020 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Kuzan 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic		
Patient characteristics and setting	Age group: adults only Setting: outpatient		
Index tests	Index test(s): chest CT (non-contrast) Defintion for positive diagnosis on CT: BSTI version 2 Level of training of readers: radiologist Prevalence: 0.6		
Target condition and reference standard(s)	Reference standard: RT-PCR twice, if necessary		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			

Kuzan 2020 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Li 2020a

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic
Patient characteristics and setting	Age group: adults, perhaps also children Setting: unclear
Index tests	Index test(s): chest CT Defintion for positive diagnosis on CT: specific scoring criteria based on literature findings Level of training of readers: radiologist Prevalence: 0.5
Target condition and reference standard(s)	Reference standard: RT-PCR, no other details provided
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

Li 2020a *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Luo 2020a

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic		
Patient characteristics and setting	Age group: children and adults Setting: outpatient		
Index tests	Index test(s): chest CT, no further details provided Defintion for positive diagnosis on CT: scoring system was developed (with scores from −4 to +7) Level of training of readers: radiologist Prevalence: 0.4		
Target condition and reference standard(s)	Reference standard: RT-PCR twice, if necessary		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			

Luo 2020a *(Continued)*

DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Luo 2020b

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, unclear symptom status
Patient characteristics and setting	Age group: adults only Setting: outpatient
Index tests	Index test(s): chest CT, no further details provided Defintion for positive diagnosis on CT: unclear Level of training of readers: radiologist Prevalence: 0.6
Target condition and reference standard(s)	Reference standard: RT-PCR, no other details provided
Flow and timing	
Comparative	
Notes	
Methodological quality	

Luo 2020b *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Mei 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, symptomatic or asymptomatic		
Patient characteristics and setting	Age group: children and adults Setting: unclear		
Index tests	Index test(s): chest CT, no further details provided Defintion for positive diagnosis on CT: unclear Level of training of readers: radiologist Prevalence: 0.5		
Target condition and reference standard(s)	Reference standard: RT-PCR twice, if necessary		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Mei 2020 *(Continued)*

DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Miranda Magalhães Santos 2020

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic
Patient characteristics and setting	Age group: children and adults Setting: outpatient
Index tests	Index test(s): chest CT, no further details provided Defintion for positive diagnosis on CT: RSNA classification Level of training of readers: radiologist Prevalence: 0.5
Target condition and reference standard(s)	Reference standard: RT-PCR, no other details provided
Flow and timing	
Comparative	
Notes	

Miranda Magalhães Santos 2020 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		

Miranda Magalhães Santos 2020 *(Continued)*

Could the patient flow have introduced bias?	Low risk

Murphy 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic		
Patient characteristics and setting	Age group: children and adults Setting: outpatient		
Index tests	Index test(s): chest X -rays Defintion for positive diagnosis on X-ray: Classification system: normal, no finding (category 0); abnormal but no lung opacity consistent with pneumonia (category 1); lung opacity consistent with pneumonia (unlikely COVID-19) (category 2); lung opacity consistent with pneumonia (consistent with COVID-19) (category 3). Sensitivities matched to AI reading. Level of training of readers: radiologist Prevalence: 0.5		
Target condition and reference standard(s)	Reference standard: RT-PCR, no other details provided		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
DOMAIN 2: Index Test (Chest X-ray)			

Murphy 2020 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Unclear		
Could the patient flow have introduced bias?		Unclear risk	

Narinx 2020

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic
Patient characteristics and setting	Age group: adults, perhaps also children Setting: outpatient
Index tests	Index test(s): chest CT (low dose); ultrasound of lungs (POCUS) Defintion for positive diagnosis on CT: scored as suggestive for or inconsistent with COVID-19 infection based on the presence of clinical manifestations as presented by Ng et al. (Ng 2020) and Shi et al. (Shi 2020)

Narinx 2020 *(Continued)*

	Defintion for positive diagnosis on ultrasound: positive if one or more BLUE points showed a positive B-line parameter Level of training of readers: radiologist Prevalence: 0.2
Target condition and reference standard(s)	Reference standard: RT-PCR, no other details provided
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	

Narinx 2020 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Pare 2020

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic
Patient characteristics and setting	Age group: adults, perhaps also children Setting: outpatient
Index tests	Index test(s): chest X-rays; ultrasound of lungs (POCUS) Defintion for positive diagnosis on X-ray: if the report included infection in the differential, as defined by words such as opacity, consolidation, or airspace disease; negative if no abnormality was noted, an abnormality was noted but attributed to a non-infectious etiology, or was inconclusive for infectious process Definition for positive diagnosis on ultrasound: positive if any B-lines were detected. Level of training of readers: unclear Prevalence: 0.6
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some
Flow and timing	

Pare 2020 *(Continued)*

Comparative
Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
DOMAIN 2: Index Test (Chest X-ray)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Ultrasound of the lungs)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		

Pare 2020 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Patel 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic		
Patient characteristics and setting	Age group: children and adults Setting: outpatient		
Index tests	Index test(s): chest CT (high resolution) Defintion for positive diagnosis on CT: scoring system: consistent with multifocal pneumonia (category 1); indeterminate for multifocal pneumonia (category 2); not consistent with multifocal pneumonia (category 3) Level of training of readers: radiologist Prevalence: 0.5		
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		

Patel 2020 *(Continued)*

Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Peng 2020

Study characteristics			

Peng 2020 *(Continued)*

Patient Sampling	Study design: patients with suspected COVID-19, symptomatic or asymptomatic
Patient characteristics and setting	Age group: children only Setting: unclear
Index tests	Index test(s): chest CT Definition for positive diagnosis on CT: GGO, consolidations with surrounding halo sign, nodules, residual fibre strips, lymphadenopathy Level of training of readers: radiologist Prevalence: 0.5
Target condition and reference standard(s)	Reference standard: RT-PCR, no other details provided; other (positive contacts)
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			

Peng 2020 *(Continued)*

DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Prokop 2020

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic
Patient characteristics and setting	Age group: children and adults Setting: outpatient
Index tests	Index test(s): chest CT, no further details provided Defintion for positive diagnosis on CT: CO-RADS Level of training of readers: radiologist Prevalence: 0.5
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some
Flow and timing	
Comparative	
Notes	
Methodological quality	

Prokop 2020 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Schiaffino 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, unclear symptom status		
Patient characteristics and setting	Age group: children and adults Setting: outpatient		
Index tests	Index test(s): chest X-rays Defintion for positive diagnosis on X-ray: unclear Level of training of readers: radiologist Prevalence: 0.8		
Target condition and reference standard(s)	Reference standard: RT-PCR twice, if necessary		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
DOMAIN 2: Index Test (Chest X-ray)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	

Schiaffino 2020 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Schulze-hagen 2020

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic
Patient characteristics and setting	Age group: adults only Setting: unclear
Index tests	Index test(s): chest CT (low dose) Defintion for positive diagnosis on CT: CO-RADS Level of training of readers: radiologist Prevalence: 0.4
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some
Flow and timing	
Comparative	
Notes	

Schulze-hagen 2020 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		

Schulze-hagen 2020 *(Continued)*

Could the patient flow have introduced bias?	Low risk

Song 2020a

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic		
Patient characteristics and setting	Age group: adults only Setting: unclear		
Index tests	Index test(s): chest CT, no further details provided Defintion for positive diagnosis on CT: diagnosis of viral pneumonia according to: multiple bilateral, ill-defined GGOs or mixed consolidation with diffuse peripheral distribution or bilateral pulmonary consolidation Prevalence: 0.5		
Target condition and reference standard(s)	Reference standard: RT-PCR twice, if necessary		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Song 2020a *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Steuwe 2020

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic
Patient characteristics and setting	Age group: adults only Setting: unclear
Index tests	Index test(s): chest CT (low dose) Defintion for positive diagnosis on CT: unclear; based on typical COVID-19 findings reported by Salehi et al. (Salehi 2020). Level of training of readers: unclear Prevalence: 0.2
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some

Steuwe 2020 *(Continued)*

Flow and timing
Comparative
Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Steuwe 2020 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Stevens 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, all symptomatic		
Patient characteristics and setting	Age group: adults only Setting: outpatient		
Index tests	Index test(s): chest X-rays Defintion for positive diagnosis on X-ray: BSTI template Level of training of readers: radiologist Prevalence: 0.8		
Target condition and reference standard(s)	Reference standard: RT-PCR once; twice in some		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			

Stevens 2020 *(Continued)*

DOMAIN 2: Index Test (Chest X-ray)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Wang 2020a

Study characteristics	
Patient Sampling	Study design: patients with suspected COVID-19, symptomatic or asymptomatic
Patient characteristics and setting	Age group: children and adults Setting: unclear
Index tests	Index test(s): chest CT (no further details provided) Defintion for positive diagnosis on CT: standardised imaging reporting system: : infectious disease, viral pneumonia is highly likely (class 1), infectious lesions, viral pneumonia (class 2), infectious

Wang 2020a *(Continued)*

	lesions, pathogens to be investigated (class 3), infectious lesions (class 4) Level of training of readers: unclear Prevalence: 0.2
Target condition and reference standard(s)	RT-PCR twice, if necessary
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		

Wang 2020a *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Xiong 2020

Study characteristics			
Patient Sampling	Study design: patients with suspected COVID-19, unclear symptom status		
Patient characteristics and setting	Age group: children and adults Setting: inpatient		
Index tests	Index test(s): chest CT, no further details provided Definition for positive diagnosis on CT: subpleural GGO without pleural effusion, bronchial changes or lymphadenopathy Level of training of readers: radiologist Prevalence: 0.4		
Target condition and reference standard(s)	Reference standard: RT-PCR, no other details provided		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		

Xiong 2020 *(Continued)*

Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (Chest CT)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (Chest X-ray)			
DOMAIN 2: Index Test (Ultrasound of the lungs)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Unclear		
Could the patient flow have introduced bias?		High risk	

ACR: American College of Radiology; **AI:** artificial intelligence; **BSTI:** British Society of Thoracic Imaging; **CO-RADS:** COVID-19 Reporting and Data System; **CT:** computed tomography; **GGO:** ground-glass opacity; **IV:** intravenous; **POCUS:** point-of-care ultrasound; **RSNA:** Radiological Society of North America; **RT-PCR:** reverse transcriptase polymerase chain reaction; **STR:** Society of Thoracic Radiology

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Ai 2020b	Ineligible study design
Ai 2020c	Ineligible setting
Arentz 2020	Ineligible patient population
Bai 2020a	Ineligible study design
Bai 2020b	Ineligible study design
Chang 2020	< 10 participants
Chen 2020a	Ineligible outcomes
Chen 2020b	Ineligible outcomes
Chen 2020c	Ineligible patient population
Cheng 2020	Ineligible outcomes
Çinkooğlu 2020	Ineligible study design
Colombi 2020	Ineligible outcomes
Dai 2020	Ineligible outcomes
Ding 2020	Ineligible outcomes
Dong 2020	Ineligible study design
Guan 2020	< 10 participants
Hao 2020	< 10 participants
Himoto 2020	Ineligible study design
Huang 2020	< 10 participants
Liang 2020	Ineligible study design
Lu 2020	Ineligible patient population
Mao 2020	Ineligible study design
Miao 2020a	Ineligible study design
Miao 2020b	Ineligible study design
Pakray 2020	Ineligible study design
Poggiali 2020	Ineligible outcomes
Pu 2020	Ineligible study design
Siegel 2020	Ineligible study design

Study	Reason for exclusion
Song 2020b	Ineligible outcomes
Tavare 2020	Ineligible study design
Wang 2020b	Ineligible patient population
Wu 2020a	Ineligible setting
Wu 2020b	Ineligible setting
Wu 2020c	Ineligible patient population
Wu 2020d	Ineligible patient population
Xie 2020	Ineligible study design
Xu 2020a	Ineligible outcomes
Xu 2020b	< 10 participants
Yang 2020a	Ineligible setting
Yang 2020b	Ineligible study design
Yuan 2020	Ineligible target condition
Zhifeng 2020	Ineligible study design

DATA

Presented below are all the data for all of the tests entered into the review.

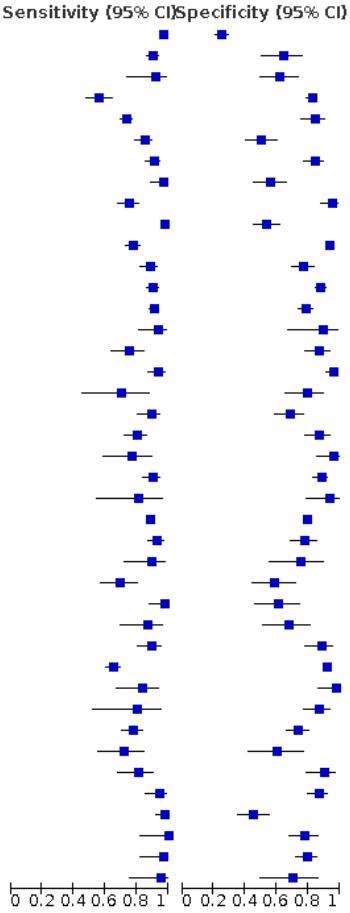
Table Tests. Data tables by test

Test	No. of studies	No. of participants
1 Chest CT in suspected cases	41	16133
2 Chest X-ray in suspected cases	9	3694
3 Ultrasound of the lungs in suspected cases	5	446
4 CT CO-RADS 2	7	2560
5 CT CO-RADS 3	9	2807
6 CT CO-RADS 4	7	2560
7 CT CO-RADS 5	7	2560

Test 1. Chest CT in suspected cases

Chest CT in suspected cases

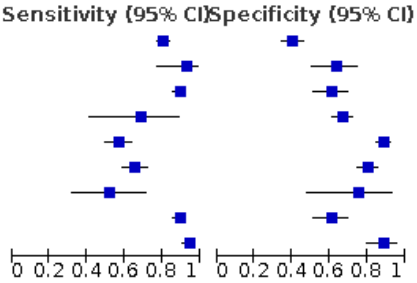
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ai 2020a	580	308	21	105	0.97 [0.95, 0.98]	0.25 [0.21, 0.30]
Aslan 2020	226	20	24	36	0.90 [0.86, 0.94]	0.64 [0.50, 0.77]
Barbosa 2020	23	25	2	41	0.92 [0.74, 0.99]	0.62 [0.49, 0.74]
Bellini 2020	80	76	62	354	0.56 [0.48, 0.65]	0.82 [0.78, 0.86]
Besutti 2020	438	16	158	84	0.73 [0.70, 0.77]	0.84 [0.75, 0.91]
Borakati 2020	162	55	29	56	0.85 [0.79, 0.90]	0.50 [0.41, 0.60]
Cartocci 2020	147	24	15	128	0.91 [0.85, 0.95]	0.84 [0.77, 0.90]
Caruso 2020	60	42	2	54	0.97 [0.89, 1.00]	0.56 [0.46, 0.66]
Debray 2020	120	4	40	77	0.75 [0.68, 0.81]	0.95 [0.88, 0.99]
Deng 2020	423	71	10	83	0.98 [0.96, 0.99]	0.54 [0.46, 0.62]
De Smet 2020	279	33	79	468	0.78 [0.73, 0.82]	0.93 [0.91, 0.95]
Dofferhoff 2020	136	36	18	122	0.88 [0.82, 0.93]	0.77 [0.70, 0.84]
Ducray 2020	259	49	28	358	0.90 [0.86, 0.93]	0.88 [0.84, 0.91]
Falaschi 2020	419	66	43	245	0.91 [0.88, 0.93]	0.79 [0.74, 0.83]
Fonsi 2020	41	2	3	17	0.93 [0.81, 0.99]	0.89 [0.67, 0.99]
Fujioka 2020	57	10	19	68	0.75 [0.64, 0.84]	0.87 [0.78, 0.94]
Gezer 2020	92	5	6	119	0.94 [0.87, 0.98]	0.96 [0.91, 0.99]
Giannitto 2020	14	10	6	38	0.70 [0.46, 0.88]	0.79 [0.65, 0.90]
Gietema 2020	74	35	9	75	0.89 [0.80, 0.95]	0.68 [0.59, 0.77]
Guillo 2020	103	11	26	74	0.80 [0.72, 0.86]	0.87 [0.78, 0.93]
He 2020	26	2	8	46	0.76 [0.59, 0.89]	0.96 [0.86, 0.99]
Hermans 2020	120	22	13	164	0.90 [0.84, 0.95]	0.88 [0.83, 0.92]
Hernigou 2020	13	2	3	29	0.81 [0.54, 0.96]	0.94 [0.79, 0.99]
Herpe 2020	1999	525	250	2050	0.89 [0.88, 0.90]	0.80 [0.78, 0.81]
Korevaar 2020	117	25	9	88	0.93 [0.87, 0.97]	0.78 [0.69, 0.85]
Krdzalic 2020	25	7	3	21	0.89 [0.72, 0.98]	0.75 [0.55, 0.89]
Kuzan 2020	48	21	21	30	0.70 [0.57, 0.80]	0.59 [0.44, 0.72]
Li 2020a	42	19	1	30	0.98 [0.88, 1.00]	0.61 [0.46, 0.75]
Luo 2020a	26	14	4	29	0.87 [0.69, 0.96]	0.67 [0.51, 0.81]
Luo 2020b	70	7	8	55	0.90 [0.81, 0.95]	0.89 [0.78, 0.95]
Mei 2020	274	39	145	447	0.65 [0.61, 0.70]	0.92 [0.89, 0.94]
Miranda Magalhães Santos 2020	30	1	6	38	0.83 [0.67, 0.94]	0.97 [0.87, 1.00]
Narinx 2020	12	10	3	65	0.80 [0.52, 0.96]	0.87 [0.77, 0.93]
Patel 2020	125	41	36	115	0.78 [0.70, 0.84]	0.74 [0.66, 0.80]
Peng 2020	28	13	11	20	0.72 [0.55, 0.85]	0.61 [0.42, 0.77]
Prokop 2020	43	5	10	47	0.81 [0.68, 0.91]	0.90 [0.79, 0.97]
Schulze-hagen 2020	65	16	4	106	0.94 [0.86, 0.98]	0.87 [0.80, 0.92]
Song 2020a	108	55	3	45	0.97 [0.92, 0.99]	0.45 [0.35, 0.55]
Steuwe 2020	19	19	0	67	1.00 [0.82, 1.00]	0.78 [0.68, 0.86]
Wang 2020a	28	33	1	128	0.97 [0.82, 1.00]	0.80 [0.72, 0.85]
Xiong 2020	19	8	1	19	0.95 [0.75, 1.00]	0.70 [0.50, 0.86]



Test 2. Chest X-ray in suspected cases

Chest X-ray in suspected cases

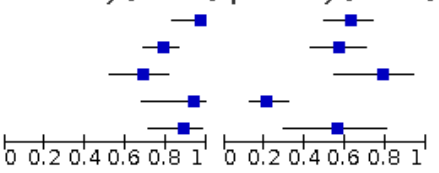
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Borakati 2020	441	186	107	126	0.80 [0.77, 0.84]	0.40 [0.35, 0.46]
Choudhury 2020	27	25	2	43	0.93 [0.77, 0.99]	0.63 [0.51, 0.75]
Cozzi 2020	363	50	45	77	0.89 [0.86, 0.92]	0.61 [0.52, 0.69]
Hwang 2020	11	105	5	211	0.69 [0.41, 0.89]	0.67 [0.61, 0.72]
Ippolito 2020	116	35	88	279	0.57 [0.50, 0.64]	0.89 [0.85, 0.92]
Murphy 2020	146	45	77	186	0.65 [0.59, 0.72]	0.81 [0.75, 0.85]
Pare 2020	14	4	13	12	0.52 [0.32, 0.71]	0.75 [0.48, 0.93]
Schiaffino 2020	363	50	45	77	0.89 [0.86, 0.92]	0.61 [0.52, 0.69]
Stevens 2020	234	8	14	64	0.94 [0.91, 0.97]	0.89 [0.79, 0.95]



Test 3. Ultrasound of the lungs in suspected cases

Ultrasound of the lungs in suspected cases

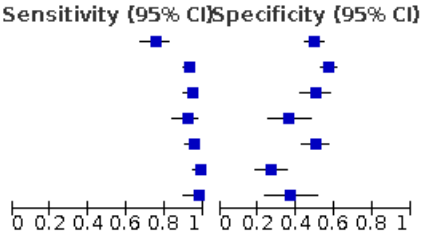
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bar 2020	30	26	1	43	0.97 [0.83, 1.00]	0.62 [0.50, 0.74]
Dini 2020	74	24	20	32	0.79 [0.69, 0.86]	0.57 [0.43, 0.70]
Fonsi 2020	30	4	14	15	0.68 [0.52, 0.81]	0.79 [0.54, 0.94]
Narinx 2020	14	59	1	16	0.93 [0.68, 1.00]	0.21 [0.13, 0.32]
Pare 2020	24	7	3	9	0.89 [0.71, 0.98]	0.56 [0.30, 0.80]



Test 4. CT CO-RADS 2

CT CO-RADS 2

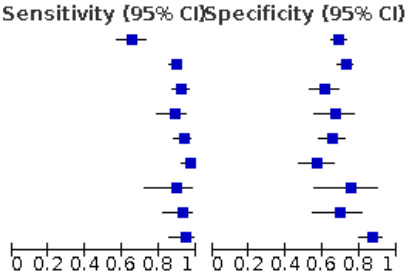
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bellini 2020	107	216	35	214	0.75 [0.67, 0.82]	0.50 [0.45, 0.55]
De Smet 2020	331	215	27	286	0.92 [0.89, 0.95]	0.57 [0.53, 0.61]
Dofferhoff 2020	145	79	9	79	0.94 [0.89, 0.97]	0.50 [0.42, 0.58]
Fujioka 2020	70	50	6	28	0.92 [0.84, 0.97]	0.36 [0.25, 0.48]
Hermans 2020	127	93	6	93	0.95 [0.90, 0.98]	0.50 [0.43, 0.57]
Korevaar 2020	124	83	2	30	0.98 [0.94, 1.00]	0.27 [0.19, 0.36]
Prokop 2020	52	33	1	19	0.98 [0.90, 1.00]	0.37 [0.24, 0.51]



Test 5. CT CO-RADS 3

CT CO-RADS 3

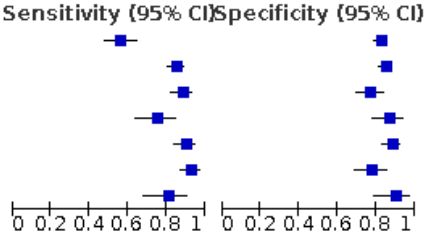
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bellini 2020	93	134	49	296	0.65 [0.57, 0.73]	0.69 [0.64, 0.73]
De Smet 2020	319	138	39	363	0.89 [0.85, 0.92]	0.72 [0.68, 0.76]
Dofferhoff 2020	142	61	12	97	0.92 [0.87, 0.96]	0.61 [0.53, 0.69]
Fujioka 2020	67	26	9	52	0.88 [0.79, 0.94]	0.67 [0.55, 0.77]
Hermans 2020	124	64	9	122	0.93 [0.88, 0.97]	0.66 [0.58, 0.72]
Korevaar 2020	122	49	4	64	0.97 [0.92, 0.99]	0.57 [0.47, 0.66]
Krdzalic 2020	25	7	3	21	0.89 [0.72, 0.98]	0.75 [0.55, 0.89]
Prokop 2020	49	16	4	36	0.92 [0.82, 0.98]	0.69 [0.55, 0.81]
Schulze-hagen 2020	65	16	4	106	0.94 [0.86, 0.98]	0.87 [0.80, 0.92]



Test 6. CT CO-RADS 4

CT CO-RADS 4

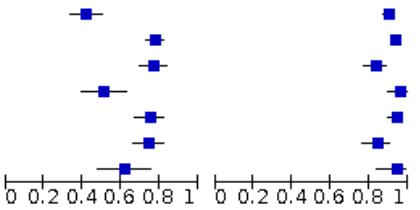
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bellini 2020	80	76	62	354	0.56 [0.48, 0.65]	0.82 [0.78, 0.86]
De Smet 2020	304	76	54	425	0.85 [0.81, 0.88]	0.85 [0.81, 0.88]
Dofferhoff 2020	136	36	18	122	0.88 [0.82, 0.93]	0.77 [0.70, 0.84]
Fujioka 2020	57	10	19	68	0.75 [0.64, 0.84]	0.87 [0.78, 0.94]
Hermans 2020	120	22	13	164	0.90 [0.84, 0.95]	0.88 [0.83, 0.92]
Korevaar 2020	117	25	9	88	0.93 [0.87, 0.97]	0.78 [0.69, 0.85]
Prokop 2020	43	5	10	47	0.81 [0.68, 0.91]	0.90 [0.79, 0.97]



Test 7. CT CO-RADS 5

CT CO-RADS 5

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bellini 2020	59	41	83	389	0.42 [0.33, 0.50]	0.90 [0.87, 0.93]
De Smet 2020	279	33	79	468	0.78 [0.73, 0.82]	0.93 [0.91, 0.95]
Dofferhoff 2020	119	26	35	132	0.77 [0.70, 0.84]	0.84 [0.77, 0.89]
Fujioka 2020	39	3	37	75	0.51 [0.40, 0.63]	0.96 [0.89, 0.99]
Hermans 2020	100	11	33	175	0.75 [0.67, 0.82]	0.94 [0.90, 0.97]
Korevaar 2020	94	18	32	95	0.75 [0.66, 0.82]	0.84 [0.76, 0.90]
Prokop 2020	33	3	20	49	0.62 [0.48, 0.75]	0.94 [0.84, 0.99]



ADDITIONAL TABLES

Table 1. Summary of included studies

Study ID	Country of corresponding author	Study design (symptom status)	Age group	Setting	Index test(s)	Definition for index test positivity	Training level of readers	Reference standard	Proportion of initial negative results with repeat RT-PCR	Prevalence
Ai 2020a	China	People with suspected COVID-19 (unclear)	Adults only	Unclear	Chest CT	Unclear	Radiologist	RT-PCR, no other details provided	Unclear	0.6
Aslan 2020	Turkey	People with suspected COVID-19 (all symptomatic)	Adults only	Outpatient	Chest CT (low dose)	Radiological evidence of COVID-19 pneumonia, including presence of GGO, mixed GGO (GGO and consolidation), consolidation, distribution and number of lobes and segment affected by GGO and/or consolidation, etc	Radiologist	RT-PCR twice, if necessary	1	0.8
Bar 2020	France	People with suspected COVID-19 (unclear)	Adults only	Outpatient	Ultrasound of the lungs (POCUS)	Unclear	Unclear	RT-PCR twice, if necessary	1	0.3
Barbosa 2020	Brazil	People with suspected COVID-19 (all symptomatic)	Adults only	Unclear	Chest CT	RSNA classification	Radiologist	RT-PCR, no other details provided	Unclear	0.3
Bellini 2020	Italy	People with suspected COVID-19 (all symptomatic)	Children and adults	Unclear	Chest CT (non-contrast)	CO-RADS	Radiologist	RT-PCR once; twice in some; other (clinical signs on follow-up)	Unclear	0.2
Besutti 2020	Italy	People with suspected	Adults, perhaps	Outpatient	Chest CT (non-contrast)	Structured report about the probability of COVID-19 pneumonia: high-	Radiologist	RT-PCR once; twice in some	0.26	0.9

Table 1. Summary of included studies *(Continued)*

		COVID-19 (all symptomatic)	also children			ly suggestive, suggestive, non-suggestive				
Borakati 2020	UK	People with suspected COVID-19 (symptomatic or asymptomatic)	Adults, perhaps also children	Outpatient	Chest CT (IV contrast); chest radiographs / Chest X-rays	BSTI template	Radiologist	RT-PCR once; twice in some	Unclear	0.6
Cartocci 2020	Italy	People with suspected COVID-19 (all symptomatic)	Adults only	Outpatient	Chest CT	Classification system by Simpson 2020: typical CT pattern, possible CT pattern, inconsistent CT pattern, negative for pneumonia	Radiologist	RT-PCR once; twice in some	Unclear	0.5
Caruso 2020	Italy	People with suspected COVID-19 (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast)	Pneumonia	Radiologist	RT-PCR twice, if necessary	1	0.4
Choudhury 2020	India	People with suspected COVID-19 (all symptomatic)	Unclear	Inpatient	Chest X-rays	A previously unvalidated Likert score (scores 1-5) based on radiographic features thought to be related to COVID-19, based on format reported by Simpson 2020	Unclear	RT-PCR, no other details provided	Unclear	0.3
Cozzi 2020	Italy	People with suspected COVID-19 (symptomatic or asymptomatic)	Unclear	Outpatient	Chest X-rays	The presence of interstitial infiltrates with predominantly bilateral and basal distribution	Radiologist	RT-PCR, no other details provided; other (follow-up phone call)	Unclear	0.8
Debray 2020	France	People with suspected COVID-19 (unclear)	Adults only	Outpatient	Chest CT (non-contrast)	"Evocative": multifocal GGOs, being nodular or not, or crazy-paving with or without consolidations, with a bilateral, peripheral or mixed distribution and	Radiologist	RT-PCR once; twice in some	0.24	0.7

Table 1. Summary of included studies *(Continued)*

						involvement of the posterior zones				
Deng 2020	China	People with suspected COVID-19 (all symptomatic)	Children and adults	Unclear	Chest CT (high resolution)	Any one of the following: • single, multiple, or diffuse GGO, with thickened blood vessels and thickened bronchial shadows passing through, with or without localised lobular septal grid thickening; • single or multiple real shadows Re-examination 3-5 days later showed that the original GGO or consolidation range increased, the number increased, or accompanied by pleural effusion on 1 or both sides	Radiologist	RT-PCR once	0	0.7
De Smet 2020	Belgium	People with suspected COVID-19 (all symptomatic)	Children and adults	Outpatient	Chest CT	CO-RADS classification; threshold not pre-specified	Unclear	RT-PCR, no other details provided	Unclear	0.4
Dini 2020	Italy	People with suspected COVID-19 (symptomatic or asymptomatic)	≥ 70 years of age	Outpatient (LTC)	Ultrasound of lungs (POCUS)	Classification system: non-coalescent B-lines in > 3 zones (score 1), coalescent B-lines in > 3 zones (score 2), and with hyperechoic non-consolidated state (score 3)	Unclear	RT-PCR, no other details provided	Unclear	0.6
Dofferhoff 2020	The Netherlands	People with suspected COVID-19 (symptomatic or asymptomatic)	Adults only	Outpatient	Chest CT (low dose)	CO-RADS classification; threshold not pre-specified	Unclear	RT-PCR once; twice in some	Unclear	0.5

Table 1. Summary of included studies *(Continued)*

Ducray 2020	France	People with suspected COVID-19 (symptomatic or asymptomatic)	Adults only	Outpatient	Chest CT (IV contrast)	Classification system: surely COVID+, possible COVID+, COVID-	Radiologist	RT-PCR once; twice in some	0.048	0.4
Falaschi 2020	Italy	People with suspected COVID-19 (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast)	STR/ACR/RSNA	Radiologist	RT-PCR once; twice in some	Unclear	0.6
Fonsi 2020	Italy	People with suspected COVID-19 (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast); ultrasound of lungs (POCUS)	Chest CT: GGOs; consolidation; a mixed GGO and consolidation pattern; single or multiple solid nodules surrounded by GGOs; a focal or multifocal distribution; GGO and consolidation location; multilobe involvement; a bilateral distribution; interlobular septal thickening; an air bronchogram; the presence of cavitation; bronchial wall thickening; bronchiectasis; mediastinal lymph node enlargement; pleural effusion; and pericardial effusion. Ultrasound: not reported	Radiologist	RT-PCR once; twice in some	Unclear	0.7
Fujioka 2020	Japan	People with suspected COVID-19 (all symptomatic)	Adults only	Unclear	Chest CT	CO-RADS	Radiologist	RT-PCR once; twice in some	Unclear	0.5
Gezer 2020	Turkey	People with suspected COVID-19 (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast)	Unclear	Radiologist	RT-PCR, no other details provided; other (clinical signs and imaging tests)	Unclear	0.4

Table 1. Summary of included studies *(Continued)*

Giannitto 2020	Italy	People with suspected COVID-19 (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast)	Classification system: suspected COVID-19 pneumonia, non-COVID-19 pneumonia, negative CT	Radiologist	RT-PCR twice, if necessary	1	0.3
Gietema 2020	The Netherlands	People with suspected COVID-19 (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast)	Standardised imaging reporting system (typical for COVID-19, equivocal, non COVID-19)	Resident	RT-PCR once; twice in some	Unclear	0.4
Guillo 2020	France	People with suspected COVID-19 (all symptomatic)	Adults only	Outpatient	Chest CT (IV contrast)	A structured report about the probability of COVID-19 pneumonia based on the presence of GGOs with or without crazy-paving pattern, isolated or admixed with perilobular or linear consolidation, their peripheral or central distribution, etc	Resident	RT-PCR once; twice in some	0.27	0.6
He 2020	China	People with suspected COVID-19 (unclear)	Children and adults	Unclear	Chest CT (high resolution)	Ground-glass opacity with or without consolidation, crazy paving patten, peripheral and diffuse distribution, and bilateral/multilobular involvement	Radiologist	RT-PCR once; twice in some	Unclear	0.4
Hermans 2020	The Netherlands	People with suspected COVID-19 (symptomatic or asymptomatic)	Adults only	Outpatient	Chest CT	CO-RADS	Radiologist	RT-PCR once	0	0.4
Hernigou 2020	Belgium	People with suspected COVID-19 (symptomatic or asymptomatic)	Adults only	Outpatient	Chest CT (low dose)	Unclear	Radiologist	RT-PCR once; twice in some	Unclear	0.3

Table 1. Summary of included studies *(Continued)*

Herpe 2020	France	People with suspected COVID-19 (all symptomatic)	Children and adults	Unclear	Chest CT	Structured report based on bilateral GGOs with peripheral distribution, bilateral crazy paving appearance with intralobular thickening, reverse halo sign, or other signs compatible with organizing pneumonia	Radiologist	RT-PCR once; twice in some	0.04	0.5
Hwang 2020	Korea	People with suspected COVID-19 (symptomatic or asymptomatic)	Adults, perhaps also children	Unclear	Chest X-rays	Any abnormality suggesting pneumonia	Radiologist and Resident	RT-PCR, no other details provided	Unclear	0.05
Ippolito 2020	Italy	People with suspected COVID-19 (all symptomatic)	Children and adults	Outpatient	Chest X-rays	Reticulations, alveolar opacities or both	Radiologist	RT-PCR, no other details provided	Unclear	0.4
Korevaar 2020	The Netherlands	People with suspected COVID-19 (all symptomatic)	Adults only	Outpatient	Chest CT (low dose)	CO-RADS	Radiologist	RT-PCR once; twice in some	0.61	0.5
Krdzalic 2020	The Netherlands	People with suspected COVID-19 (all symptomatic)	Adults only	Unclear	Chest CT	CO-RADS	Radiologist	RT-PCR twice, if necessary	1	0.5
Kuzan 2020	Turkey	People with suspected COVID-19 (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast)	BSTI template (version 2)	Radiologist	RT-PCR twice, if necessary	1	0.6
Li 2020a	China	People with suspected COVID-19 (all symptomatic)	Adults, perhaps also children	Unclear	Chest CT	Specific scoring criteria based on literature findings	Radiologist	RT-PCR, no other details provided	Unclear	0.5

Table 1. Summary of included studies *(Continued)*

Luo 2020a	China	People with suspected COVID-19 (all symptomatic)	Children and adults	Unclear	Chest CT	Scoring system was developed (with scores from −4 to +7)	Radiologist	RT-PCR twice, if necessary	1	0.4
Luo 2020b	China	People with suspected COVID-19 (unclear)	Adults only	Outpatient	Chest CT	Unclear	Radiologist	RT-PCR, no other details provided	Unclear	0.6
Mei 2020	USA	People with suspected COVID-19 (symptomatic or asymptomatic)	Children and adults	Unclear	Chest CT	Unclear	Radiologist	RT-PCR twice, if necessary	1	0.5
Miranda Magalhães Santos 2020	Brazil	People with suspected COVID-19 (all symptomatic)	Children and adults	Outpatient	Chest CT	RSNA classification	Radiologist	RT-PCR, no other details provided	Unclear	0.5
Murphy 2020	The Netherlands	People with suspected COVID-19 (all symptomatic)	Children and adults	Outpatient	Chest X-rays	Classification system: normal, no finding (category 0); abnormal but no lung opacity consistent with pneumonia (category 1); lung opacity consistent with pneumonia (unlikely COVID-19) (category 2); lung opacity consistent with pneumonia (consistent with COVID-19) (category 3). Sensitivities matched to AI reading.	Radiologist	RT-PCR, no other details provided	Unclear	0.5
Narinx 2020	Belgium	People with suspected COVID-19 (all symptomatic)	Adults, perhaps also children	Outpatient	Chest CT (low dose); ultrasound of lungs (POCUS)	Chest CT: scored as suggestive for or inconsistent with COVID-19 infection based on Ng et al. (Ng 2020) and Shi et al. (Shi 2020). Ultrasound: positive if ≥ 1 BLUE points	Radiologist	RT-PCR, no other details provided	Unclear	0.2

Table 1. Summary of included studies *(Continued)*

						showed a positive B-line parameter				
Pare 2020	USA	People with suspected COVID-19 (all symptomatic)	Adults, perhaps also children	Outpatient	Chest X-rays; ultrasound of lungs (POCUS)	Chest X-ray: positive if the report included infection in the differential, as defined by words such as opacity, consolidation, or airspace disease; negative if no abnormality was noted, an abnormality was noted but attributed to a non-infectious aetiology, or was inconclusive for infectious process. Ultrasound: positive if any B-lines were detected.	Unclear	RT-PCR once; twice in some	0.25	0.6
Patel 2020	USA	People with suspected COVID-19 (all symptomatic)	Children and adults	Outpatient	Chest CT (high resolution)	Scoring system: consistent with multifocal pneumonia (category 1); indeterminate for multifocal pneumonia (category 2); not consistent with multifocal pneumonia (category 3)	Radiologist	RT-PCR once; twice in some	0.2	0.5
Peng 2020	China	People with suspected COVID-19 (symptomatic or asymptomatic)	Children only	Unclear	Chest CT	GGO, consolidations with surrounding halo sign, nodules, residual fibre strips, lymphadenopathy	Radiologist	RT-PCR, no other details provided; other (positive contacts)	Unclear	0.5
Prokop 2020	The Netherlands	People with suspected COVID-19 (all symptomatic)	Children and adults	Outpatient	Chest CT	CO-RADS classification; threshold not pre-specified	Radiologist	RT-PCR once; twice in some	Unclear	0.5
Schiaffino 2020	Italy	People with suspected COVID-19 (unclear)	Children and adults	Outpatient	Chest X-rays	Unclear	Radiologist	RT-PCR twice, if necessary	1	0.8

Table 1. Summary of included studies *(Continued)*

Schulze-hagen 2020	Germany	People with suspected COVID-19 (all symptomatic)	Adults only	Unclear	Chest CT (low dose)	CO-RADS	Radiologist	RT-PCR once; twice in some	Unclear	0.4
Song 2020a	China	People with suspected COVID-19 (all symptomatic)	Adults only	Unclear	Chest CT	Viral pneumonia according to: multiple bilateral, ill-defined GGOs or mixed consolidation with diffuse peripheral distribution or bilateral pulmonary consolidation	Radiologist	RT-PCR twice, if necessary	1	0.5
Steuwe 2020	Germany	People with suspected COVID-19 (all symptomatic)	Adults only	Unclear	Chest CT (low dose)	Unclear; based on typical COVID-19 findings reported by Salehi et al. (Salehi 2020).	Unclear	RT-PCR once; twice in some	Unclear	0.2
Stevens 2020	UK	People with suspected COVID-19 (all symptomatic)	Adults only	Outpatient	Chest X-rays	BSTI template	Radiographer and Radiologist	RT-PCR once; twice in some	Unclear	0.8
Wang 2020a	China	People with suspected COVID-19 (symptomatic or asymptomatic)	Children and adults	Unclear	Chest CT	Standardised imaging reporting system: infectious disease, viral pneumonia is highly likely (class 1), infectious lesions, viral pneumonia (class 2), infectious lesions, pathogens to be investigated (class 3), infectious lesions (class 4)	Unclear	RT-PCR twice, if necessary	1	0.2
Xiong 2020	China	People with suspected COVID-19 (unclear)	Children and adults	Inpatient	Chest CT	Subpleural GGO without pleural effusion, bronchial changes or lymphadenopathy	Radiologist	RT-PCR, no other details provided	Unclear	0.4

AI: artificial intelligence; **BSTI:** British Society of Thoracic Imaging; **CO-RADS:** COVID-19 Reporting and Data System; **CT:** computed tomography; **GGO:** ground-glass opacity; **IV:** intravenous; **LTC:** long-term care; **POCUS:** point-of-care ultrasound; **RSNA:** Radiological Society of North America; **RT-PCR:** reverse transcription polymerase chain reaction

Table 2. Sensitivity analyses for chest CT of suspected cases

Analysis	Studies (n)	Number of participants (cases)	Sensitivity (95% CI)	Specificity (95% CI)
Published in peer-reviewed journals[a]	38	15,442 (7719)	88.5% (85.2 to 91.2)	81.2% (76.0 to 85.3)

CI: confidence interval;**CT:** computed tomography

[a]The publication status of studies has been updated as of 1 November 2020.

Table 3. Meta-regression analyses for chest CT of suspected cases

Test, analysis group	Studies (n)	Number of participants (cases)	Sensitivity (95% CI)	Specificity (95% CI)
Reference standard conduct				
RT-PCR testing at least twice for all initial negative results	9	2087 (1018)	88.1% (78.5 to 93.8)	71.3% (59.9 to 80.6)
RT-PCR testing not repeated for all initial negative results	22	11,344 (5779)	86.7% (82.3 to 90.1)	82.6% (77.8 to 86.6)
P value			0.74	0.05
Definition for index test positivity				
Radiologist impression	13	7000 (3565)	90.3% (84.5 to 94.1)	77.2% (67.0 to 84.9)
Formal scoring system	23	6805 (3333)	85.9% (81.2 to 89.2)	80.0% (75.0 to 84.2)
P value			0.15	0.58

CI: confidence interval;**CT:** computed tomography; **RT-PCR:** reverse transcription polymerase chain reaction

Table 4. Analyses of 'threshold' effects for chest CT studies of suspected cases that used the COVID-19 Reporting and Data System (CO-RADS)

CO-RADS threshold	Studies (n)	Number of participants (cases)	Sensitivity (95% CI)	Specificity (95% CI)
5	7	2560 (1042)	67.0% (56.4 to 76.2)	91.3% (87.6 to 94.0)
4	7	2560 (1042)	83.5% (74.4 to 89.7)	83.6% (80.5 to 86.4)
3	9	2807 (1139)	90.7% (85.2 to 94.3)	69.4% (63.3 to 74.9)
2	7	2560 (1042)	94.3% (88.6 to 97.2)	44.1% (36.5 to 52.0)
1[a]	-	-	-	-

CI: confidence interval;**CT:** computed tomography

[a]We did not perform meta-analysis for a CO-RADS threshold of 1 since at this threshold all sensitivity values are equal to 1, and all specificity values are equal to 0.

Table 5. Indirect comparisons of sensitivity and specificity of chest CT, chest X-ray and ultrasound

Indirect comparisons of sensitivity and specificity of chest CT, chest X-ray and ultrasound					
			P value (comparison of sensitivities) P value (comparisons of specificities)		
			Chest CT	**Chest X-ray**	**Ultrasound of the lungs**
		Studies (participants)	41 (16,133)	9 (3694)	5 (446)
	Studies (participants)	**Sensitivity (95% CI)** **Specificity (95% CI)**	87.9% (84.6 to 90.6) 80.0% (74.9 to 84.3)	80.6% (69.1 to 88.6) 71.5% (59.8 to 80.8)	86.4% (72.7 to 93.9) 54.6% (35.3 to 72.6)
Chest CT	41 (16,133)	87.9% (84.6 to 90.6) 80.0% (74.9 to 84.3)	-	-	-
Chest CT/X-ray	9 (3694)	80.6% (69.1 to 88.6) 71.5% (59.8 to 80.8)	P = 0.10 P = 0.12	-	-
Ultrasound of the lungs	5 (446)	86.4% (72.7 to 93.9) 54.6% (35.3 to 72.6)	P = 0.77 P = 0.0052	P = 0.43 P = 0.13	-
CI: confidence interval;**CT:** computed tomography					

APPENDICES

Appendix 1. Glossary

Terminology/acronyms

- **COVID-19:** coronavirus disease 2019, the clinical manifestations/symptoms caused by infection with SARS-CoV-2, name given to the disease associated with the virus SARS-CoV-2
- **COVID-19 pneumonia:** COVID-19 that presents as infection-inflammation of the lungs
- **Index test:** the test that is being assessed (the index test will often be a new test)
- **False negative:** the test does not detect a condition in someone when it is present
- **False positive:** the test detects a condition in someone when it is not present
- **Negative predictive value:** the probability that someone who has tested negative for the target condition with the index test will really not have it (a true negative)
- **Positive predictive value:** the probability that someone who has tested positive for the target condition with the index test will actually have it (a true positive)
- **Reference standard:** the most reliable method for determining if the target condition is present or absent, used to verify index test results. This could be a combination of tests.
- **RT-PCR:** reverse transcription polymerase chain reaction (RT-PCR) is a laboratory technique that combines reverse transcription of RNA into DNA and amplification of specific DNA targets using polymerase chain reaction. In this context it is used to detect the presence of SARS-CoV-2 RNA

- **SARS-CoV-2:** severe acute respiratory syndrome coronavirus 2, the name given to the 2019 novel coronavirus
- **SARS-CoV-2 infection:** people infected with severe acute respiratory syndrome coronavirus 2, but who may or may not have any clinical manifestations of infection
- **Secondary care:** medical care that is provided by a specialist or facility upon referral by a primary care physician and that requires more specialised knowledge, skill, or equipment than the primary care physician can provide
- **Sensitivity:** the proportion of people with the target condition (with disease) that are correctly identified by the index test
- **Specificity:** the proportion of people without the target condition (without disease) that are correctly identified by the index test
- **Tertiary care:** specialised care, usually for inpatients and on referral from a primary or secondary health professional, in a facility that has personnel and facilities for advanced medical investigation and treatment
- **Target condition:** the disease or condition of interest
- **True negative:** a correct diagnosis of a condition being absent
- **True positive:** a correct diagnosis of a condition being present

Appendix 2. QUADAS-2

QUADAS-2	
Index test(s):	Imaging studies of the chest (computed tomography (CT), chest X-ray and ultrasound) for diagnosis of COVID-19
Participants (setting, intended use of index test, presentation, prior testing):	People with suspected COVID-19 All settings, in particular secondary care, emergency care and intensive care units (ICUs) In people presenting with suspected COVID-19; suspicion may be based on prior testing, such as general lab testing Signs and symptoms often used for triage or referral
Reference standard and target condition:	A positive diagnosis for COVID-19 by the following. 1. A positive reverse transcriptase polymerase chain reaction (RT-PCR) test for SARS-CoV-2 infection, from any manufacturer in any country, from any source, including nasopharyngeal swabs or aspirates, oropharyngeal swabs, bronchoalveolar lavage fluid (BALF), sputum, saliva, serum, urine, rectal or faecal samples 2. Positive on WHO criteria for COVID-19, which includes some testing RT-PCR-negative 3. Positive on China CDC criteria for COVID-19, which includes some testing RT-PCR-negative 4. Positive serology in addition to consistent symptomatology 5. Positive on study-specific list of criteria for COVID-19, which includes some testing RT-PCR-negative 6. Other criteria (symptoms, imaging findings, other tests) A negative diagnosis for COVID-19 by the following. 1. People with suspected COVID-19 with negative RT-PCR test results, whether tested once or more than once 2. Current healthy or with another disease (no RT-PCR test) This list is not exhaustive, as we anticipate that studies will use a variety of reference standards and we plan to include all of them, at least for the earlier versions of the review. Although RT-PCR is considered the best available test, it is suspected of missing a substantial proportion of cases, and thus may not be the ideal reference standard if used as a standalone test (Li 2020b; Loeffelholz 2020). Therefore, we are likely to use alternative reference standards, such as a combination of RT-PCR, and symptoms or imaging findings, or both. We will judge how likely each reference standard definition is to correctly classify individuals in the assessment of methodological quality. All reference standards are likely to be imperfect in some way; details of reference standard evaluation are provided in the 'Risk of bias' tool below. We will

(Continued)

	use a consensus process to agree the classification of the reference standard as to what we regard as good, moderate and poor. 'Good' reference standards need to have very little change of misclassification, 'moderate', a small but acceptable risk, 'poor', a larger and probably unacceptable risk.
Participant selection	
Was a consecutive or random sample of patients enrolled?	YES: if a study explicitly states that all participants within a certain time frame were included; that this was done consecutively; or that a random selection was done. NO: if it is clear that a different selection procedure was employed; e.g. selection based on clinician's preference, or based on institutions (i.e. 'convenience' series) UNCLEAR: if the selection procedure is not clear or not reported at all.
Was a case-control design avoided?	YES: if a study explicitly states that all participants came from the same group of (suspected) patients. NO: if it is clear that a different selection procedure was employed for the participants depending on their COVID-19 status (e.g. proven infected patients in one group and proven non-infected patients in the other group). UNCLEAR: if the selection procedure is not clear or not reported at all.
Did the study avoid inappropriate in- or exclusions?	This needs to be addressed on a case-to-case basis. YES: If all eligible patients were ore or less equally suspected of having COVID-19 and were included and if the numbers in the flow chart show not too many excluded participants (a maximum of 20% of eligible patients excluded without reasons). NO: If over 20% of eligible patients were excluded without providing a reason; if only proven patients were included, or only proven non-patients were included; if in a retrospective study participants without index test or reference standard result were excluded; if exclusion was based on severity assessment post-factum or comorbidities (cardiovascular disease, diabetes, immunosuppression). If the study oversampled patients with particular characteristics likely to affect estimates of accuracy. UNCLEAR: if the exclusion criteria are not reported.
Could the selection of patients have introduced bias?	HIGH: if one or more signalling questions were answered with NO, as any deviation from the selection process may lead to bias. LOW: if all signalling questions were answered with YES. UNCLEAR: all other instances
Is there concern that the included patients do not match the review question?	This needs to be addressed on a case-to-case basis, based on the objective the included study answers to. HIGH: if accuracy was assessed in a case-control design, or the study was able to only estimate sensitivity or specificity. LOW: any situation where imaging is generally available. UNCLEAR: if a description about the participants is lacking.
Index tests	
Were the index test results interpreted without knowledge of the results of the reference standard?	YES: if blinding was explicitly stated or index test was recorded before the results from the reference standard were available NO: if it was explicitly stated that the index test results were interpreted with knowledge of the results of the reference standard

(Continued)

	UNCLEAR: if blinding was unclearly reported.
If a threshold was used, was it prespecified?	YES: for any of these index tests it is highly unlikely that any numerical threshold is used. Still we expect studies to report their criteria for test-positivity (e.g. the constellation of imaging findings used). If these criteria are reported in the methods section, we will score 'YES' for this question. NO: if the optimal criterion for test-positivity was based on the reported data (for example, different scores on a quantitative scoring system) we will score 'NO'. UNCLEAR: if the criteria for test positivity were not or unclearly reported.
Could the conduct or interpretation of the index test have introduced bias?	HIGH: if one or more signalling questions were answered with NO. LOW: if all signalling questions were answered with YES. UNCLEAR: all other instances Note: For studies that use formal scoring systems with clearly defined thresholds, even if the signalling question about using a 'prespecified threshold' is 'unclear' or 'no', this domain should not be considered as having a 'unclear' or 'high' risk of bias based on the aforementioned question.
Is there concern that the index test, its conduct, or interpretation differ from the review question?	There is not a huge amount of variability from a technical perspective. Therefore, this question will probably be answered 'LOW' in all cases except when assessments are made using personnel not available in practice, or personnel not trained for the job, or using modalities that are uncommon in practice. We will consult expert clinicians on a case-to-case basis to judge this question.
Reference standard	
Is the reference standard likely to correctly classify the target condition?	YES: for COVID-19: RT-PCR, done by trained personnel, and repeated after a first negative RT-PCR, following guidelines for confirmed cases and done with an assay targeting minimum 2 targets in the genes N, E, S or RdRP (one target even acceptable in zone with known transmission). To clarify, a low risk of bias reference standard for true negative would require 2 (or more) negative RT-PCR results. NO: any other test UNCLEAR: if no reference standard was reported, or if it was just reported that RT-PCR was done.
Were the reference standard results interpreted without knowledge of the results of the index test?	YES: if it was explicitly stated that the reference standard results were interpreted without knowledge of the results of the index test, or if the result of the index test was obtained after the reference standard. NO: if it was explicitly stated that the reference standard results were interpreted with knowledge of the results of the index test or if the index test was used to make the final diagnosis (incorporation bias). UNCLEAR: if blinding was unclearly reported.
Could the conduct or interpretation of the reference standard have introduced bias?	HIGH: if one or more signalling questions were answered with NO. LOW: if all signalling questions were answered with YES. UNCLEAR: all other instances Note: For studies that use RT-PCR testing as the reference standard, even if this signalling question about 'blinding' is 'unclear' or 'no', this domain should not be considered as having a 'unclear' or 'high' risk of bias based on the aforementioned question.
Is there concern that the target condition as defined by the reference standard does	HIGH: there is a high concern regarding applicability of the reference standard if the reference standard actually measures a different target condition than the one we are interested in for the review. For example, if the diagnosis is only based on clinical picture, without excluding other possi-

(Continued)

not match the review question?	ble causes of this clinical picture (e.g. other respiratory pathogens), then there is considerable concern that the reference standard is actually measuring something else than COVID-19. In addition, a positive RT-PCR only measures SARS-CoV-2 infection and not COVID-19 and therefore the reference standard for COVID-19 is a combination of positive RT PCR and symptoms and/or imaging findings. LOW: if above situations not present UNCLEAR: if intention for testing is not reported in the study
Flow and timing	
Was there an appropriate interval between index test(s) and reference standard?	YES: as the situation of a patient, including clinical presentation and disease progress, evolves rapidly and new/ongoing exposure can result in case status change. On the other hand, negative PCR results need to be repeated for several days. Therefore, an appropriate time interval will be within 7 days. NO: if there is more than 7 days between the index test and the reference standard or if patients are otherwise reported to be assessed with the index versus reference standard test at moments of different severity. UNCLEAR: if the time interval is not reported
Did all participants receive a reference standard?	YES: if all patients received a reference standard (clearly no partial verification) NO: if only (part of) the index test positives or index test negatives received the complete reference standard UNCLEAR: if it is not reported.
Did all participants receive the same reference standard?	YES: if all patients received the same reference standard (clearly no differential verification). Verification of negative PCR result with a second PCR measurement is considered to be one reference standard. NO: if (part of) the index test positives or index test negatives received a different reference standard UNCLEAR: If it is not reported.
Were all participants included in the analysis?	YES: if all included participants were included in the analyses as well NO: if after the inclusion/exclusion process, participants were removed from the analyses for different reasons: no reference standard done, no index test done, intermediate results of both index test or reference standard, indeterminate results of both index test or reference standard, samples unusable. UNCLEAR: If this is not clear from the reported numbers.
Could the patient flow have introduced bias?	HIGH: if one or more signalling questions were answered with NO, or if one question answered with NO was judged to have little impact on the methodological quality of the study (this should be justified in the scoring). LOW: if all signalling questions were answered with YES. UNCLEAR: all other instances

CT: computed tomography; **CXR:** chest X-ray; **ICU:** intensive care unit; **RT-PCR:** reverse transcriptase polymerase chain reaction; **SARS-CoV-2:** severe acute respiratory syndrome coronavirus 2; **US:** ultrasound

Appendix 3. Search classification model

A more efficient approach was required to keep up with the rapidly increasing volume of COVID-19 literature. A classification model for COVID-19 diagnostic studies was built with the model building function within Eppi Reviewer, which uses the standard SGCClassifier in Scikit-learn on word trigrams (Thomas 2010). As outputs, new documents receive a percentage (from the predict_proba function) where scores close to 100 indicate a high probability of belonging to the class 'relevant document' and scores close to 0 indicate a low probability of belonging to the class 'relevant document'. We used three iterations of manual screening (title and abstract screening, followed by full-text review) to build and test classifiers. The final included studies were used as relevant documents, while the remainder of the COVID-19 studies were used as irrelevant documents. The classifier was trained on the first round of selected articles, and tested and retrained on the second round of selected articles. Testing on the second round of selected articles revealed poor positive predictive value but 100% sensitivity at a cut-off of 10. The poor positive predictive value is mainly due to the broad scope of our topic (all diagnostic studies in COVID-19), poor reporting in abstracts, and a small set of included documents. The model was retrained using the articles selected of the second and third rounds of screening, which added a considerable number of additional documents. This led to a large increase in positive predictive value, at the cost of a lower sensitivity, which led us to reduce the cut-off to 5. The largest proportion of documents had a score between 0-5. This set did not contain any of the relevant documents. This version of the classifier with a cut-off 5 was used in subsequent rounds and accounted for approximately 80% of the screening burden.

Appendix 4. Search strategies

1. Living search from the University of Bern

27 April 2020

From 27 April 2020, we retrieved the curated bioRxiv/medRxiv dataset link

26 March 2020 to 27 April 2020

MEDLINE: (\"Wuhan coronavirus\" [Supplementary Concept] OR \"COVID-19\" OR \"2019 ncov\"[tiab] OR ((\"novel coronavirus\"[tiab] OR \"new coronavirus\"[tiab]) AND (wuhan[tiab] OR 2019[tiab])) OR 2019-nCoV[All Fields] OR (wuhan[tiab] AND coronavirus[tiab])))))

Embase: (nCoV or 2019-nCoV or ((new or novel or wuhan) adj3 coronavirus) or covid19 or covid-19 or SARS-CoV-2).mp

bioRxiv/medRxiv: ncov or corona or wuhan or COVID or SARS-CoV-2

With the kind support of the Public Health & Primary Care Library PHC, and following guidance of the Medical Library Association

01 January 2020 to 27 April 2020

MEDLINE: ("Wuhan coronavirus" [Supplementary Concept] OR "COVID-19" OR "2019 ncov"[tiab] OR (("novel coronavirus"[tiab] OR "new coronavirus"[tiab]) AND (wuhan[tiab] OR 2019[tiab])) OR 2019-nCoV[All Fields] OR (wuhan[tiab] AND coronavirus[tiab])))))

Embase: ncov OR (wuhan AND corona) OR COVID

bioRxiv/medRxiv: ncov or corona or wuhan or COVID

2. Cochrane COVID-19 Study Register searches

Source	Strategy
ClinicalTrials.gov	COVID-19 OR 2019-nCoV OR SARS-CoV-2 OR 2019 novel coronavirus OR severe acute respiratory syndrome coronavirus 2 OR Wuhan coronavirus OR coronavirus
WHO International Clinical Trials Registry Platform	We screen the entire COVID-19.csv file available from www.who.int/emergencies/diseases/novel-coronavirus-2019
PubMed	(2019 nCoV[tiab] OR 2019nCoV[tiab] OR corona virus[tiab] OR corona viruses[tiab] OR coronavirus[tiab] OR coronaviruses[tiab] OR COVID[tiab] OR COVID19[tiab] OR nCov 2019[tiab] OR SARS-CoV2[tiab] OR SARS CoV-2[tiab] OR SARSCoV2[tiab] OR SARSCoV-2[tiab] OR "Coronavirus"[Mesh:NoExp] OR "COVID-19"[nm] OR "COVID-19 drug treatment"[nm] OR "COVID-19 diagnostic testing"[nm] OR "COVID-19 serotherapy"[nm] OR "COVID-19 vaccine"[nm] OR "LAMP assay"[nm] OR "severe acute respiratory syndrome coronavirus 2"[nm] OR "spike protein, SARS-CoV-2"[nm]) NOT ("animals"[mh] NOT "humans"[mh]) NOT (editorial[pt] OR newspaper article[pt])

3. CDC Library, COVID-19 Research Articles Downloadable Database

Embase records from the Stephen B. Thacker CDC Library, Covid-19 Research articles Downloadable database.

Records were obtained by the CDC Library by searching Embase through Ovid using the following search strategy.

Source	Strategy
Embase	(coronavir* OR corona virus* OR betacoronavir* OR covid19 OR covid 19 OR nCoV OR novel CoV OR CoV 2 OR CoV2 OR sarscov2 OR 2019nCoV OR wuhan virus*).mp. OR ((wuhan OR hubei OR huanan) AND (severe acute respiratory OR pneumonia*) AND outbreak*).mp. OR Coronavirus infection/ OR coronavirinae/ OR exp betacoronavirus/ Limits: 2020- OR (novel coronavir* OR novel corona virus* OR covid19 OR covid 19 OR nCoV OR novel CoV OR CoV 2 OR CoV2 OR sarscov2 OR 2019nCoV OR wuhan virus*).mp. OR ((wuhan OR hubei OR huanan) AND (severe acute respiratory OR pneumonia*) AND outbreak*).mp. OR ((wuhan OR hubei OR huanan) AND (coronavir* OR betacoronavir*)).mp. Limits: 2019-

WHAT'S NEW

Date	Event	Description
10 March 2021	New citation required and conclusions have changed	The results for chest X-ray and ultrasound have changed.
9 February 2021	New search has been performed	This is a 'living' systematic review'; searches are run and screened every few months. The last search date was 30 September 2020. Results of all new studies identified have been incorporated. The conclusions of this Cochrane Review are therefore considered up to date.

HISTORY

Protocol first published: Issue 6, 2020
Review first published: Issue 9, 2020

Date	Event	Description
23 October 2020	New search has been performed	This is a 'living' systematic review'; searches are run and screened monthly. The last search date was 22 June 2020. Results of all new studies identified have been incorporated. The conclusions of this Cochrane Review are therefore considered up to date.
23 October 2020	New citation required and conclusions have changed	The results for chest computed tomography (CT) have changed.

CONTRIBUTIONS OF AUTHORS

All authors reviewed, edited, contributed to, and approved this review.

The search was performed by RS, MMGL and LH.

DECLARATIONS OF INTEREST

Nayaar Islam has no known conflicts of interest.

Sanam Ebrahimzadeh has no known conflicts of interest.

Jean-Paul Salameh has no known conflicts of interest.

Sakib Kazi has no known conflicts of interest.

Nicholas Fabiano has no known conflicts of interest.

Lee Treanor has no known conflicts of interest.

Marissa Absi has no known conflicts of interest.

Zachary Hallgrimson has no known conflicts of interest.

Mariska MG Leeflang has no known conflicts of interest.

Lotty Hooft has no known conflicts of interest.

Christian B van der Pol has no known conflicts of interest.

Ross Prager has no known conflicts of interest.

Samanjit Singh Hare has no known conflicts of interest.

Carole Dennie has no known conflicts of interest.

René Spijker: the Dutch Cochrane Centre (DCC) has received grants for performing commissioned systematic reviews. In no situation did the commissioner have any influence on the results of the work.

Jonathan J Deeks has no known conflicts of interest.

Jacqueline Dinnes has no known conflicts of interest.

Kevin Jenniskens has no known conflicts of interest.

Daniel Korevaar has no known conflicts of interest.

Jérémie F Cohen has no known conflicts of interest.

Ann Van den Bruel has no known conflicts of interest.

Yemisi Takwoingi has no known conflicts of interest.

Janneke van de Wijgert has no known conflicts of interest.

Johanna AAG Damen has no known conflicts of interest.

Junfeng Wang received a consultancy fee from Biomind, an Artificial Intelligence (AI) company providing machine intelligence solutions in medical imaging. The consultancy service was about design of clinical studies, not related to this review. The company had no influence on the results of the work.

Matthew McInnes has no known conflicts of interest.

SOURCES OF SUPPORT

Internal sources

- Liverpool School of Tropical Medicine, UK

External sources

- Foreign, Commonwealth and Development Office (FCDO), UK

  Project number: 300342-104
- National Institute for Health Research (NIHR), UK
- Government of Ontario Ministry of Health COVID-19 Rapid Response Research Grant program, Canada

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

Inclusion criteria

The exclusion of case-control studies, as well as studies that report an overview of index test findings in participants with and without the target condition, without explicitly classifying the imaging test as either COVID-19 positive or negative, are modifications from the study protocol and the previous versions of this review. These changes were made prior to initiating the update with approval by the Cochrane COVID-19 Diagnostic Test Accuracy Group, as well as all of the review authors.

Risk of bias assessment

The criteria for the index test and reference standard domains of the QUADAS-2 tool were modified for this update (Appendix 2). For studies that used formal scoring systems with clearly defined thresholds, even if the signalling question about using a 'prespecified threshold' was 'unclear' or 'no', the index test domain was not considered to have a 'unclear' or 'high' risk of bias based on the 'prespecified threshold' signalling question. For studies that used RT-PCR testing as the reference standard, even if this signalling question about 'blinding' was 'unclear' or 'no', the reference standard domain was not considered to have a 'unclear' or 'high' risk of bias based on the 'blinding' signalling question. These changes were approved by the Cochrane COVID-19 Diagnostic Test Accuracy Group, as well as all of the review authors.

Secondary objectives

We did not address several planned secondary objectives due to insufficient available data (McInnes 2020). These objectives include: evaluating the rate of positive imaging in patients with initial RT-PCR-negative results who have a positive result on a follow-up RT-PCR test; determining if there is an association between number of days after symptom onset, symptom severity and the findings on thoracic imaging for patients with COVID-19; and determining the rate of alternative diagnoses identified by thoracic imaging.

Sensitivity analyses

We had planned to undertake additional sensitivity analyses to determine whether low risk of bias for all QUADAS-2 domains had an effect on findings. However, since the majority of included studies had an overall high or unclear risk of bias due to study design and only two studies had an overall low risk of bias, it was not possible to undertake these analyses.

Investigations of heterogeneity

Our protocol included additional sources of heterogeneity to be evaluated, such as disease prevalence, participant symptoms (severity), timing of symptom onset, participant co-morbidities and other potential candidate variables. Due to the lack of available data, we did not investigate these covariates.

Limitations of previous review and changes in this update

The previous version of this review included studies of cross-sectional or case-control designs that either:

1. reported specific criteria for index test positivity (i.e. used a scoring system, such as CO-RADS);
2. did not report specific criteria, but had the index test reader(s) explicitly classify the imaging test result as either COVID-19 positive or negative; or
3. reported an overview of index test findings, without having the index test reader(s) explicitly classify index tests as either COVID-19 positive or negative.

The inclusion of case-control studies may have been a source of bias as the disease prevalence in the sample of these types of studies do not represent the prevalence in the target population. The inclusion of studies that only reported an overview of index test findings (i.e. studies not intended to be 'diagnostic test accuracy studies') was a possible source of bias identified by sensitivity analysis in our previous review and may have limited our ability to evaluate the sensitivity and specificity of chest CT, chest X-ray and ultrasound. In this update, we excluded studies with case-control designs, and studies that only reported an overview of index test findings without having the index test reader(s) explicitly classify index tests as either COVID-19 positive or negative. The body of evidence has grown to the point that sufficient studies that meet these preferred criteria are now available.

Investigations of variability were limited in the previous review due to limited available data. The assessment of secondary objectives such as the association between number of days after symptom onset, symptom severity and the findings on thoracic imaging for patients with COVID-19 was also not possible. In this update, we evaluated the impact of reference standard conduct (RT-PCR, performed at least twice

in all initial negative results versus RT-PCR, not performed at least twice in all initial negative results) and definition used for index test positivity (formal scoring system versus radiologist impression), but we were unable to conduct further investigations of variability due to limited available data. We also formally evaluated the impact of threshold effects on accuracy estimates in this update, particularly for studies that used the CO-RADS scoring system. We were unable to evaluate threshold effects in other types of formal scoring systems due to the limited number of included studies that used other systems.

Of the studies included in the previous review, several failed to clearly report key information about their study design, as well as their methods for recruiting participants and delivering the reference standard. Therefore, data derived from these studies may have a high risk of bias and this quality of reporting and weaknesses in the primary studies reflected the overall degree of robustness of our study. In this update, several included studies also failed to report key information and had a high or unclear risk of bias with respect to participant selection, index test, reference standard, and participant flow.

The interpretation of the accuracy estimates in the previous review involved several uncertainties. While RT-PCR is considered the best available test, the results of the RT-PCR are not always sensitive; sensitivity depends on the timing of specimen collection, with high sensitivity around the onset of symptoms and during the symptomatic period but lower sensitivity before and after that window (Kucirka 2020), and collection of an appropriate specimen for testing can also be challenging. RT-PCR alone may not be the ideal reference standard (Li 2020b; Loeffelholz 2020), and it is possible that chest CT may be more sensitive than the reference standard in some patients, as some patients identified as having a false-positive diagnosis on CT may have been missed by the RT-PCR test. In this update, similar uncertainties with respect to the use of RT-PCR as the reference standard exist. However, our meta-regression analyses for studies that performed RT-PCR testing at least twice for all participants with initial negative results (i.e. studies that addressed, to some extent, the low sensitivity of RT-PCR testing by conducting at least two RT-PCR tests to define disease-negative status) compared with studies that did not perform repeat RT-PCR testing for all participants with initial negative results, did not identify significantly different accuracy estimates between the groups. The quality of reporting and the design of the included studies also affected the generalisability and ability to assess the validity of our findings.

About a quarter of the studies (9/34; 26%) included in the previous review were only available as preprints at the time of the search and had not yet been through the peer-review process; of the four preprint studies that were included in the previous review and also included in this update, two have since been published (publication statuses are updated as of 1 November 2020). Compared to the previous review, this update includes a notably smaller proportion of preprint studies (3/51; 6%). We will update data extracted from these studies and include them in future versions of our review as these studies become published in peer-reviewed journals.

INDEX TERMS

Medical Subject Headings (MeSH)

Bias; Confidence Intervals; COVID-19 [*diagnostic imaging]; COVID-19 Nucleic Acid Testing [standards]; Lung [diagnostic imaging]; *Radiography, Thoracic [standards] [statistics & numerical data]; Reference Standards; Sensitivity and Specificity; *Tomography, X-Ray Computed [standards] [statistics & numerical data]; *Ultrasonography [standards] [statistics & numerical data]

MeSH check words

Adolescent; Adult; Aged; Child; Humans; Middle Aged; Young Adult