Cochrane 'living' systematic review on diagnostic accuracy of imaging for COVID-19

Nayaar Islam

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for the
Degree of Master of Science in Epidemiology

School of Epidemiology and Public Health
Faculty of Medicine
University of Ottawa

Preface

All components of this thesis were conducted primarily by Nayaar Islam, under the supervision of Dr. Matthew McInnes at the University of Ottawa. The University of Ottawa Research Ethics Board (REB) policies indicate that ethics review is not required for the use of previously collected, publicly available, anonymously collected data; thus, REB approval was not required for any of the projects comprising this thesis.

Chapter 2 is a shortened version of a paper that has been published in the Cochrane Database of Systematic Reviews [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. 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] and represents the third version of an ongoing 'living' systematic review. Nayaar Islam was the lead investigator responsible for updating methodologies based on the prior versions of the 'living' systematic review, conducting data collection and statistical analyses, and preparing the manuscript. Matthew McInnes was the supervising author responsible for developing the initial concept and providing guidance throughout the project and preparation of the manuscript. The remaining authors provided technical and clinical expertise, supported the data collection process, and edited the manuscript.

Chapter 3 is a shortened version of a paper that has been published in the Cochrane Database of Systematic Reviews [citation: Islam N*, Ebrahimzadeh S*, Dawit H, Salameh J-P, Kazi S, Fabiano N, Treanor L, Absi M, Ahmad F, Rooprai P, Al Khalil A, Harper K, Kamra N, 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, Wang J, Pena E, Sabongui S, McInnes MDF. Thoracic imaging tests for the diagnosis of COVID-19. Cochrane Database of Systematic Reviews 2022, Issue 5. Art. No.: CD013639. DOI: 10.1002/14651858.CD013639.pub5] and represents the fourth version of an ongoing 'living' systematic review. Nayaar Islam was the co-lead investigator alongside Sanam Ebrahimzadeh and was responsible for updating methodologies based on the prior versions of the 'living' systematic review, conducting data collection and statistical analyses, and preparing the manuscript. Matthew McInnes was the supervising author responsible for developing the initial concept and providing guidance throughout the project and preparation of the manuscript. The remaining authors provided technical and clinical expertise, supported the data collection process, and edited the manuscript.

Chapter 4 is an original work by Islam N, Alghita MK, Ebrahimzadeh S, Dawit H, Prager R, Alvarez GG, Cohen JF, Korevaar DA, Deeks JJ, Verbakel JY, Damen JAAG, Ochodo EA, Nair AV, Dinnes J, Dennie C, Van den Bruel A, van de Wijgert J, Sikora L, Spijker R, Hare SS, and McInnes MDF. This work is a protocol currently accepted for publication (with minor revisions) in the Cochrane Database of Systematic Reviews. Nayaar Islam was the lead investigator responsible for designing methodologies and preparing the manuscript. Matthew McInnes was the supervising author responsible for concept formation, providing guidance throughout the

project and preparation of the manuscript. The remaining authors provided technical and clinical expertise, supported the data collection process, and edited the manuscript.

Acknowledgements

This thesis would not have been possible without the support of my supervisor, colleagues, collaborators, advisors, friends, and family. I am grateful to each one of them for their contributions towards the success of this work.

I am deeply grateful to Dr. Matthew McInnes, who has given me invaluable guidance and mentorship, and has shown me constant support and sincerity throughout this journey. I feel lucky to have had the opportunity to work with Dr. McInnes.

I am thankful to all my colleagues and collaborators whose expertise and contributions were essential to the completion of this work. Thank you to Jean-Paul Salameh for his mentorship and advice throughout the pursuit of my degree. Thank you to the members of the Cochrane COVID-19 Diagnostic Test Accuracy Group for their assistance through this unique process of conducting a 'living' systematic review series in real-time during the COVID-19 pandemic. Thank you to the members of my thesis advisory committee, Dr. Dean Fergusson, and Dr. Gonzalo Alvarez, and my thesis examiners, Dr. David Moher, and Dr. Melissa Brouwers.

Finally, I would like to thank my mom, my dad, my sister, and my husband, who have always been there to cheer me on.

The amazing people I have met, the relationships I have built, and the skills I have developed, have made this a truly memorable experience.

As a graduate student, I have received generous financial support from several sources, including the following scholarships and grants:

- Frederick Banting and Charles Best Canada Graduate Scholarship - Master's (Canadian Institutes of Health Research)
- Mitacs' Research Training Award
- University of Ottawa Excellence Scholarship
- University of Ottawa Graduate Scholarship
- Government of Ontario Ministry of Health COVID-19 Rapid Response Research Grant
- University of Ottawa Faculty of Medicine COVID-19 Pandemic Response Funding Program

Abstract

Background: The coronavirus disease 2019 (COVID-19) presents diagnostic evaluation and patient management challenges, including uncertainty regarding the role of imaging tests. This series of reviews from the suite of Cochrane 'living' systematic reviews aims to evaluate the accuracy of chest imaging (computed tomography (CT), X-ray and ultrasound) for diagnosis and management of individuals with suspected COVID-19.

Methods: The Bern COVID-19 Living Database, Cochrane COVID-19 Register, and CDC Library were searched through 30 September 2020 (for review version 3) and 17 February 2021 (for review version 4). Diagnostic accuracy studies involving participants with suspected COVID-19 were included. Screening, data extraction, and risk of bias assessments (using the QUADAS-2 tool) were completed independently, in duplicate. Pooled accuracy estimates and 95% confidence intervals (CIs) were determined using a bivariate random effects model.

Results: In the third version of the review, chest CT (41 studies, 16133 participants, 8110 (50%) cases) had a pooled sensitivity of 87.9% (95%CI 84.6-90.6) and specificity of 80.0% (74.9-84.3). Chest X-ray (9 studies, 3694 participants, 2111 (57%) cases) had a pooled sensitivity of 80.6% (69.1-88.6) and specificity of 71.5% (59.8-80.8). Ultrasound of the lungs (5 studies, 446 participants, 211 (47%) cases) had a pooled sensitivity of 86.4% (72.7-93.9) and specificity of 54.6% (35.3-72.6). Indirect comparisons showed that chest CT gave higher specificity (P=0.0052) and similar sensitivity (P=0.77) compared to ultrasound. There were no differences (P≥0.05) in accuracy between CT and X-ray, or X-ray and ultrasound. In the fourth version of the review, chest CT (69 studies, 28285 participants, 14342 (51%) cases) had a pooled sensitivity

of 86.9% (83.6-89.6) and specificity of 78.3% (73.7-82.3). Chest X-ray (17 studies, 8530 participants, 5303 (62%) cases) had a pooled of sensitivity=73.1% (64.1-80.5) and specificity of 73.3% (61.9-82.2). Ultrasound of the lungs (15 studies, 2410 participants, 1158 (48%) cases) had a pooled sensitivity of 88.9% (84.9-92.0) and specificity of 72.2% (58.8-82.5). Indirect comparisons showed that chest CT and ultrasound had similar sensitivities (P=0.42), and each gave higher sensitivities than X-ray (P=0.0003 and P=0.001, respectively). All modalities had similar specificities (P≥0.05).

Conclusion: The most recent evidence indicates that both chest CT and ultrasound of the lungs are sensitive and moderately specific for diagnosing individuals with suspected COVID-19, while chest X-ray is moderately sensitive and moderately specific. Chest CT and ultrasound may be useful for ruling out COVID-19, but not for distinguishing COVID-19 from other illnesses. Research assessing the prognostic value of imaging for predicting morbidity and mortality in individuals with COVID-19 is underway and will also be published in the suite of Cochrane 'living' systematic reviews.

Table of Contents

Chapter 1. General introduction

On 31 December 2019, Chinese public health authorities were informed of several cases of severe pneumonia of unknown etiology in Wuhan City, Hubei Province of China; a novel coronavirus was identified as the cause of this outbreak at the beginning of January 2020 WHO 2021). On 30 January 2020, the World Health Organization (WHO) declared the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and resulting coronavirus disease 2019 (COVID-19) a global public health emergency, and then on 11 March 2020, a worldwide pandemic (WHO 2021).

SARS-CoV-2 infection and the COVID-19 pandemic have presented diagnostic and prognostic evaluation challenges. The 'gold' standard for diagnosing COVID-19 is laboratory confirmation of COVID-19 infection, such as a positive reverse transcriptase polymerase chain reaction (RT-PCR) result (WHO 2020). Other tests, including imaging tests, may also be used to assist with diagnosis and patient management. The value of imaging tests in the diagnostic and prognostic pathways for COVID-19 remains undefined.

Decisions about management pathways for patients with COVID-19 vary according to health services and settings, available resources, outbreaks in different settings, and vaccine coverage in the population. These pathways will change over time, as accurate tests, effective treatments, and vaccines are identified. The decision points between these pathways vary, but all require knowledge on the accuracy of diagnostic and prognostic information to inform medical decisions. Therefore, it is essential to understand the accuracy of tests to develop effective diagnostic and management pathways for different contexts. This supports: 1) strategies to identify those who are infected; 2) strategies to limit the spread of the disease; and 3) management of patients either through isolation precautions, contact tracing, quarantine, hospital

admission, admission to the intensive care unit, admission to a specialized facility, or initiation of specific therapies.

Research on the role of imaging in COVID-19 patients has evolved rapidly throughout the COVID-19 pandemic, and this 'living' systematic review has been updated periodically to reflect changes in the evidence base over time and address additional research objectives as they become relevant. Estimates of accuracy from this review series will help inform diagnostic, screening, isolation, and patient-management decisions. An explanation of terminology and acronyms is included in Appendix 1.1.

1.1 Objectives

This thesis aims to assess the accuracy of thoracic imaging (chest computed tomography (CT), chest X-ray and ultrasound) for the diagnosis and management of patients with COVID-19.

The primary objective across all versions of the 'living' systematic review (Chapter 2 and 3) was to evaluate the diagnostic accuracy of thoracic imaging in people with suspected COVID-19. Additional primary objectives were: to assess the rate of positive imaging in individuals with initial RT-PCR negative results and positive RT-PCR results on follow-up (Chapter 3); to evaluate the accuracy of thoracic imaging for screening asymptomatic individuals (Chapter 3); and to develop a protocol to evaluate the accuracy of thoracic imaging for predicting morbidity and mortality (Chapter 4). The secondary objectives were to investigate sources of heterogeneity and evaluate threshold effects of index test positivity on accuracy estimates (Chapter 2 and 3).

1.2 Outline

Chapter 2. Cochrane 'living' systematic review on diagnostic accuracy of imaging for COVID-19: evidence until September 2020

This section represents the third version of the 'living' systematic review and focuses on the evaluation of the diagnostic accuracy of chest CT, chest X-ray and ultrasound in people with suspected COVID-19 based on compiled evidence until 30 September 2020. Sources of heterogeneity, 'threshold' effects of index test positivity on accuracy, and time trends of accuracy for chest CT are also assessed.

Chapter 3. Cochrane 'living' systematic review on diagnostic accuracy of imaging for COVID-19: evidence until February 2021

This section represents the fourth and most recent version of the 'living' systematic review and focuses on the evaluation of the diagnostic accuracy of chest CT, chest X-ray and ultrasound in people with suspected COVID-19 based on compiled evidence until 17 February 2021. Sources of heterogeneity, and 'threshold' effects of index test positivity on accuracy are also assessed. Additionally, this section explores the updated evidence base and evaluates the rate of positive imaging in individuals with initial RT-PCR negative results and positive RT-PCR results on follow-up, and the accuracy of imaging for screening asymptomatic individuals.

Chapter 4. Prognostic accuracy of imaging findings for predicting morbidity and mortality in patients with COVID-19: systematic review protocol

This section outlines a protocol for a systematic review that will aim to evaluate the prognostic accuracy of imaging findings (obtained with chest CT, chest X-ray and ultrasound)

for predicting short-term morbidity (including complications, escalation of care, and hospital admission length) and mortality in patients with COVID-19. This future research will identify the potential value of imaging in patient management.

Chapter 5. Conclusion

This section summarizes the previous chapters in this work and provides a reflection on the unique processes that were employed to conduct a 'living' systematic review series in real-time during the COVID-19 pandemic.

References

WHO 2020
World Health Organization (WHO). WHO COVID-19 Case definition. www.who.int/publications-detail-redirect/WHO-2019-nCoV-Surveillance_Case_Definition-2022.1 (accessed 12 March 2022).

WHO 2021
World Health Organization (WHO). Timeline: WHO's COVID-19 response. www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline (accessed 10 March 2022).

Chapter 2. Cochrane 'living' systematic review on diagnostic accuracy of imaging for COVID-19: evidence until September 2020

List of Authors

1. Nayaar Islam, BHSc. School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa. Department of Radiology, University of Ottawa. Clinical Epidemiology Program, Ottawa Hospital Research Institute.
2. Sanam Ebrahimzadeh, PhD. Department of Radiology, University of Ottawa. Clinical Epidemiology Program, Ottawa Hospital Research Institute.
3. Jean-Paul Salameh, MSc. Department of Radiology, University of Ottawa.
4. Sakib Kazi, BSc. Department of Radiology, University of Ottawa.
5. Nicholas Fabiano, BSc. Department of Radiology, University of Ottawa.
6. Lee Treanor, BSc. Department of Radiology, University of Ottawa.
7. Marissa Absi, BSc. Department of Radiology, University of Ottawa.
8. Zachary Hallgrimson, BSc. Department of Radiology, University of Ottawa.
9. Mariska MG Leeflang, PhD. Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam University Medical Centers, University of Amsterdam.
10. Lotty Hooft, PhD. Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University.
11. Christian B van der Pol, MD. Department of Radiology, McMaster University.
12. Ross Prager, MD. Department of Medicine, University of Ottawa.
13. Samanjit S Hare, MD. Department of Radiology, Royal Free London NHS Trust.
14. Carole Dennie, MD. Department of Radiology, University of Ottawa. Department of Medical Imaging, The Ottawa Hospital.
15. René Spijker, PhD. Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam University Medical Centers, University of Amsterdam.

Medical Library, Amsterdam UMC, University of Amsterdam, Amsterdam Public Health.

16. Jonathan J Deeks, PhD. Test Evaluation Research Group, Institute of Applied Health Research, University of Birmingham. NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham.
17. Jacqueline Dinnes, PhD. Test Evaluation Research Group, Institute of Applied Health Research, University of Birmingham. NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham.
18. Kevin Jenniskens, PhD. Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam University Medical Centers, University of Amsterdam.
19. Daniël A Korevaar, MD, PhD. Department of Respiratory Medicine, Amsterdam UMC, University of Amsterdam.
20. Jérémie F Cohen, MD, PhD. Obstetrical, Perinatal and Pediatric Epidemiology Research Team (EPOPé), Centre of Research in Epidemiology and StatisticS (CRESS), Inserm UMR1153, Université de Paris.
21. Ann Van den Bruel, MD, PhD. Academic of Primary Care, KU Leuven.
22. Yemisi Takwoingi, PhD. Test Evaluation Research Group, Institute of Applied Health Research, University of Birmingham. NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham.
23. Janneke van de Wijgert, MD, MPH, PhD. Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam University Medical Centers, University of Amsterdam. Institute of Infection, Veterinary, and Ecological Sciences, University of Liverpool

24. Johanna AAG Damen, PhD. Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam University Medical Centers, University of Amsterdam
25. Junfeng Wang, PhD. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht.
26. Matthew DF McInnes, MD, PhD. School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa. Department of Radiology, University of Ottawa. Clinical Epidemiology Program, Ottawa Hospital Research Institute.

Preface

Objective

To assess the diagnostic accuracy of thoracic imaging (computed tomography (CT), chest X-ray and ultrasound) in the evaluation of people with suspected COVID-19, based on compiled evidence available until 30 September 2020.

Funding

1. Government of Ontario Ministry of Health COVID-19 Rapid Response Research Grant.
2. The Cochrane Infectious Diseases Group (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).

Role of funders

None of the listed funding bodies had any role in designing or conducting the study; collecting, managing, analyzing, or interpretating the data; or preparing, reviewing, or approving the manuscript.

Appendices

Five appendices (2.1 to 2.5) are included at the end of this document.

Ethics approval

This study did not require ethics approval.

Contributions of co-authors

- Concept and design: Islam, Ebrahimzadeh, Salameh, McInnes.
- Search strategy: Spijker.
- Acquisition of data: Islam, Ebrahimzadeh, Salameh, Kazi, Fabiano, Treanor, Absi, Hallgrimson, van der Pol, Prager, Jenniskens, Korevaar, Cohen, Damen, Wang, McInnes.
- Analysis and interpretation of data: Islam, Takwoingi, Cohen.
- Drafting of the manuscript: Islam, Ebrahimzadeh, McInnes.
- Critical revision of the manuscript for important intellectual content: Leeflang, Hooft, van der Pol, Prager, Hare, Dennie, Deeks, Dinnes, Korevaar, Cohen, Van den Bruel, van de Wijgert.
- Administrative, technical, or material support: McInnes.

Citation details

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. 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 previous version of this review showed thoracic (chest) imaging to be sensitive and moderately specific in the diagnosis of coronavirus disease 2019 (COVID-19), based on evidence until 22 June 2020. This updated version includes new relevant studies, and excludes studies with case-control designs, and those not intended to be diagnostic test accuracy studies.

Objectives: To evaluate the diagnostic accuracy of thoracic imaging (chest computed tomography (CT), X-ray and ultrasound) in people with suspected COVID-19. The secondary objectives were to assess sources of heterogeneity and evaluate threshold effects of index test positivity on accuracy.

Methods: The Bern COVID-19 Living Database, Cochrane COVID-19 Register, CDC Library, and repositories of COVID-19 publications were searched through to 30 September 2020. Language restrictions were not applied. 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 were included. The review authors independently and in duplicate screened articles, extracted data and assessed risk of bias and applicability concerns using QUADAS-2. A bivariate meta-analysis model was used where appropriate to determine pooled estimates of sensitivity and specificity and 95% confidence intervals (CIs).

Results: Fifty-one studies were included with 19775 participants with suspected COVID-19, of whom 10155 (51%) had a final diagnosis of COVID-19. Forty-seven studies evaluated one

imaging modality each, and four studies evaluated two each. All studies used RT-PCR as the reference standard for diagnosing COVID-19. Risk of bias was high or unclear in thirty-two (63%) studies for participant selection, 40 (78%) studies for reference standard, 30 (59%) studies for index test, and 24 (47%) studies for participant flow. Chest CT (41 studies, 16133 participants, 8110 (50%) cases) had a pooled sensitivity of 87.9% (95%CI 84.6-90.6) and specificity of 80.0% (74.9-84.3). There was no statistical evidence indicating that reference standard conduct and definition for index test positivity were sources of heterogeneity for CT studies. Chest X-ray (9 studies, 3694 participants, 2111 (57%) cases) had a pooled sensitivity of 80.6% (69.1-88.6) and specificity of 71.5% (59.8-80.8). Ultrasound of the lungs (5 studies, 446 participants, 211 (47%) cases) had a pooled sensitivity of 86.4% (72.7-93.9) and specificity of 54.6% (35.3-72.6). Based on an indirect comparison using all included studies, chest CT had a higher specificity (P=0.0052) and similar sensitivity (P=0.77) compared to ultrasound. There were no differences (P≥0.05) in specificity or sensitivity identified between chest CT and chest X-ray, or chest X-ray and ultrasound.

Conclusion: 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. The uncertainty resulting from risk of bias and heterogeneity of included studies limit the confidence of these results.

2.1 Introduction

The value of imaging tests in the diagnostic pathway for COVID-19 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).

This is the third version of a review from the suite of Cochrane 'living' systematic reviews that summarizes up-to-date evidence on the accuracy of different imaging tests (chest computed tomography (CT), chest X-ray and ultrasound) and diagnostic features in participants regardless of their symptoms. Details on the target condition (i.e. COVID-19), index tests (i.e. chest CT, chest X-ray and ultrasound of the lungs), and clinical pathway are included in Appendix 2.1.

Summary of previous versions of the review

The first version of the review included studies until 5 May 2020 and found that studies with only confirmed cases of COVID-19 reported high pooled sensitivities for chest CT and X-ray: 93.1% (95% confidence interval (CI) 90.2-95.0) and 82.1% (95%CI 62.5-92.7), respectively (Salameh 2020a). Thirteen studies that assessed chest CT in participants with suspected COVID-19 demonstrated a sensitivity of 86.2% (95%CI 71.9-93.8) but a low specificity of 18.1% (95%CI 3.71-55.8). 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. Data on the accuracy chest X-ray and ultrasound of the lungs in participants with suspected COVID-19 were not available at the time of this review version.

The second version the review included studies published until 22 June 2020 and focused on people suspected of having COVID-19, while excluding 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 85.7-92.9) and a pooled specificity of 61.1% (95%CI 42.3-77.1). 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. Pooled accuracy estimates for chest X-ray and ultrasound of the lungs in participants with suspected COVID-19 were not evaluated due to limited available data. The value of formal scoring systems for the evaluation of index tests, and 'threshold' effects of index test positivity were explored; however, formal analyses were not performed due limited available data.

Compared to the second review version, this third version 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. More studies were included in this version and both chest X-ray and ultrasound of the lungs were evaluated in addition to chest CT. Furthermore, this version formally assesses the value of formal scoring systems, 'threshold' effects, time trends of accuracy estimates of chest CT, and 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 first 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 second version of this review, the value of formal scoring systems was explored but could not be formally analyzed due to limited data. The current and future versions of this review will assess the value of formal scoring systems and 'threshold' effects (Irwig 1995) of index test positivity on accuracy estimates of imaging tests.

2.2 Objectives

The primary objective is to evaluate the diagnostic accuracy of thoracic imaging (chest CT, chest X-ray and ultrasound) in people with suspected COVID-19.

The secondary objectives are to evaluate: 1) sources of heterogeneity for accuracy estimates; 2) 'threshold' effects of index test positivity on accuracy; and 3) time trends of accuracy for chest CT.

2.3 Methods

The protocol for this review has been published in the Cochrane Database (McInnes 2020). Protocol deviations are outlined in Appendix 2.2.

Criteria for considering studies for this review

Types of studies

Studies of all designs, except for case-control studies, that included participants suspected of having the target condition and produced estimates of test accuracy or provided 2x2 data (true positive (TP), true negative (TN), false positive (FP), false negative (FN)) were included. All settings, languages, and all variations of a test were accepted. If data were not available, study authors were contacted 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

Studies with individuals suspected of having COVID-19 as outlined in the 'Target condition' section in Appendix 2.1; studies could have '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 tests' section in Appendix 2.1. 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.

Only index tests interpreted by humans were included. Studies involving interpretation by an algorithm (machine learning/artificial intelligence (AI)) were included only if they also provided data pertaining to diagnostic accuracy of human interpretation.

Definitions of index test positivity

Inclusion was limited to 'diagnostic test accuracy studies' in which the index test readers either (1) used a radiological scoring system (e.g. CO-RADS), or (2) explicitly classified patients as being positive or negative for COVID-19 based on imaging. 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 considered the combined effect of specific findings (a 'gestalt' approach), and some have used formal scoring systems, such as CO-RADS (5 categories; Prokop 2020). As such, inclusion was not limited to predefined definition or threshold for positivity. Instead, the definition for positivity used in each study was extracted.

Target conditions

The target condition was COVID-19 (Appendix 2.1). However, all studies reporting data on COVID-19 or COVID-19 pneumonia that provided data relevant to the review objectives were included.

Reference standards

A positive diagnosis for COVID-19 had to be confirmed by one or more of the following.

- 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

- positive on WHO criteria for COVID-19
- positive on China CDC criteria for COVID-19
- positive serology for SARS-CoV-2 antibodies in addition to consistent symptomatology
- positive on study-specific list of criteria for COVID-19 (e.g., symptoms, imaging findings, other tests, infected contacts)

A negative diagnosis for COVID-19 had to be confirmed by one or more of the following.

- 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)

Methodological quality was assessed based on how likely the reference standard definition used in each study correctly classified 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 2.3.

Search methods for identification of studies

Electronic searches through 30 September 2020 were conducted using three main resources:

1. Living search from the University of Bern

2. Cochrane COVID-19 Study Register searches
3. The Stephen B. Thacker CDC Library, COVID-19 Research Articles Downloadable Database

The following repositories of COVID-19 publications were checked against these search results.

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, company and product websites for studies about test accuracy were searched.
4. Companies were contacted to ask for further information about studies.
5. Research groups completing test evaluations (e.g. UK Public Health England-funded studies, Foundation for Innovative New Diagnostics (FIND) studies) were also contacted.

The search strategies are outlined in Appendix 2.4 and were devised with the help of an experienced Cochrane Information Specialist with diagnostic test accuracy 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; the searches used no terms that specifically focused on an index test, diagnostic accuracy, or study methodology.

A search classification model using artificial intelligence text analysis was implemented from 25 May 2020 and onwards, due to the increased volume of published and preprint articles (Appendix 2.4).

Selection of studies

The review authors screened studies independently, in duplicate. A third, experienced review author resolved disagreements about initial title and abstract screening. Disagreements about eligibility assessments during full-text screening were resolved 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, 2x2 contingency tables (numbers of TP, TN, FP, FN) were extracted. If a study reported accuracy data for more than one index test reader, the average of the data from all readers were taken 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, only the 2x2 contingency table corresponding to the radiologist accuracy data was extracted. If a study used multiple reference standards, but 2x2 contingency tables that included only RT-PCR as the reference standard could be computed, the latter were extracted and analyzed. If a study reported accuracy data for multiple thresholds of index test positivity (e.g. studies that used the CO-RADS scoring system), 2x2 contingency tables for all available thresholds were extracted.

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

In addition, the following items were extracted.

- 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
- Study design
- Imaging timing relative to disease course
- CT, chest X-ray and ultrasound findings
- Criteria for 'positive' diagnosis of COVID-19 on imaging
- Index test technical parameters
- 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)
- Details regarding interpretation of the index test (level of training, number of readers, the inter-observer variability)
- The number of TP, TN, FP, FN, or summary statistics from which they can be computed
- 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 tool (Appendix 2.2) which includes four domains: participant

selection, index test(s), reference standard(s), flow and timing. Three review authors resolved any disagreements through discussion.

Statistical analysis and data synthesis

Estimates of sensitivity and specificity were presented using paired forest plots and pooled estimates summarized in tables. Data were analyzed on a participant level, not a lesion or lung-segment level, since this is what determines care.

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

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

Investigations of heterogeneity

Heterogeneity was assessed by visual inspection of paired forest plots and summary receiver operating characteristics (SROC) plots. Meta-regression was performed using meqrlogit in STATA (StataCorp 2019) 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. The variable of interest was added as a covariate to a bivariate model;

using the model parameters, postestimation command was used to compute absolute differences in pooled sensitivity and specificity and 95% CIs were obtained using the delta method. P values were obtained using the Wald test.

For chest CT studies, the impact of reference standard conduct (RT-PCR performed at least twice in all participants with initial negative results versus RT-PCR not done twice), and definition for index test positivity (radiologist impression versus formal scoring system) on accuracy estimates were evaluated using meta-regression.

Threshold effects

Meta-analyses were performed 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). ggplot2 and ggforce in R were used 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

Comparisons of test accuracy between imaging modalities were undertaken, regardless of whether or not studies compared imaging modalities head-to-head in the same study population (i.e. indirect comparison). This was performed 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. P values were obtained using the Wald test. In future updates, as more data become available, test comparisons that are restricted to only comparative studies (i.e. direct comparisons) will also be performed.

Cumulative meta-analyses and time trends

For chest CT studies, univariate, cumulative, random-effects meta-analyses of sensitivity and specificity with logit-transforms were performed using STATA (StataCorp 2019). Included studies were incorporated 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). Time trends were then assessed by fitting a weighted linear regression model in which the summary estimate up to and including a given 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. Plots displaying changes in cumulative logit sensitivity and cumulative logit specificity over time were generated. These cumulative meta-analyses were restricted to the chest CT studies included in the current review.

Additionally, 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) were generated using ggplot2 and ggforce in R (Wickham 2016; Pedersen 2020; R Core Team 2021).

Assessment of reporting bias

Tests for publication bias and formal assessments of reporting bias were not performed (Salameh 2020b; Moher 2009).

2.4 Results

Results of the search

The PRISMA flow diagram is shown in Figure 2.1 (Salameh 2020b; Moher 2009). The search identified 4734 results, 773 unique references (published or preprint studies) were screened for inclusion, and 358 records were selected for full-text assessment. A total of 51 studies were included in this review, of which four have been included since the first review version (Salameh 2020a), and 12 have been included since the second review version (Islam 2020).

Description of included studies

Fifty-one 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) were included, 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 (4%), both radiologists and residents in one (2%), and radiographers in one (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%), high-resolution chest CT was used

in three (7%), low-dose CT with or without contrast was used in seven (17%) and CT with IV contrast was used in three (7%).

Manuscripts of three (6%) of the studies were published as preprints at the time of the search. The publication status of all four of the preprint studies included in the previous version of the review were updated 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 was re-extracted for analysis; and one study remained as a preprint without any updated versions.

Characteristics of the included studies are summarized in Table 2.1.

Participant characteristics

Twenty-six studies included only adult participants (aged 16 years and over), 21 included both children and adults (although in most cases, only a minority of included patients were children), one included only children, one included participants aged 70 years and older, and the remaining two 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%) involved symptomatic and asymptomatic participants, and seven (14%) 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 using only RT-PCR as the reference standard and four 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 tested some participants at least twice, if necessary, 11 tested all

participants at least twice, if necessary, and 20 did not report on the frequency of testing per participant.

Two studies included inpatients, 32 included outpatients, while the remaining 17 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 in nine studies, and ultrasound of the lungs in five studies.

Methodological quality of included studies

Figure 2.2 provides a summary of the overall methodological quality assessment using the QUADAS-2 tool for all 51 included studies. Risk of bias because of participant selection was high in 10 (20%) studies due to inappropriate exclusions (n=10). Risk of bias because of concerns regarding application of chest CT (41 studies) was high in five (12%) studies that 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 because of concerns regarding application of: chest X-ray (9 studies) was unclear in six (67%) studies; and ultrasound (5 studies) was unclear in four (80%) studies. Risk of bias based on concerns about the reference standard was high in 20 (39%)

studies that 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%) studies that 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). Figure 2.A1 in Appendix 2.5 displays a study-level quality assessment.

Findings

Pooled estimates

The sensitivity of CT in 41 studies (8110 (50%) cases in 16133 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 84.6-90.6) and the pooled specificity was 80.0% (74.9-84.3). The scatter of the study points in ROC space on the SROC plot (Figure 2.3) shows substantial variability in sensitivity and specificity. The forest plot for chest CT is presented in Figure 2.A2 in Appendix 2.5.

The sensitivity of chest X-ray in nine studies (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 69.1-88.6) and the pooled specificity was 71.5% (95%CI 59.8-80.8). The scatter of the study points in ROC space on the SROC plot (Figure 2.4) shows substantial variability in sensitivity and specificity for chest X-ray.

The sensitivity of ultrasound of the lungs in five studies (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 72.7-93.9) and the pooled specificity was 54.6% (95%CI 35.3-72.6). The scatter of the study points in ROC space on the SROC plot

(Figure 2.5) shows substantial variability in sensitivity and specificity for ultrasound of the lungs. The forest plots for chest X-ray and ultrasound of the lungs are presented in Figure 2.A3 in Appendix 2.5.

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. After excluding the three studies published as preprints, 38 studies published in peer-reviewed journals (7719 (50%) cases in 15442 participants) had a pooled sensitivity of 88.5% (95%CI 85.2-91.2) and a pooled specificity of 81.2% (95%CI 76.0-85.3). The publication status of studies has been updated as of 1 November 2020.

Investigations of heterogeneity

The results of the investigations of heterogeneity are outlined in Table 2.2. Reference standard conduct (RT-PCR testing at least twice versus not twice for all participants with initial negative results) did not have an effect ($P \geq 0.05$) on accuracy estimates for chest CT studies. 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.

Definition for index test positivity (defined based on radiologist's impressions versus using a formal scoring system), did not impact ($P \geq 0.05$) accuracy estimates for chest CT studies. For studies that used a formal scoring system, 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)) was used in the analysis.

Threshold effects

Nine studies that evaluated CT used the CO-RADS scoring system (1139 (41%) cases amongst 2807 participants) to define index test positivity. For seven studies, 2x2 data at all five CO-RADS thresholds were obtained; two studies only reported 2x2 data at a CO-RADS threshold of 3, and the authors could not provide additional data.

As the CO-RADS threshold increased from 2 to 5, pooled sensitivity decreased, and pooled specificity increased; Table 2.3 summarizes these results; Figure 2.A4 in Appendix 2.5 displays these results graphically. The forest plots of chest CT studies that used CO-RADS are presented by threshold in Figure 2.A5 in Appendix 2.5.

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 2.6. 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. Due to the limited number of studies, formal analyses could not be performed.

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). Chest X-ray and ultrasound gave similar sensitivity (P=0.43) and specificity (P=0.13) estimates. These findings are summarized in Table 2.4.

Cumulative meta-analyses and time trends

Cumulative meta-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 2.7 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 2.8 displays the pooled sensitivity and specificity estimates with 95% CIs from all versions of this review (i.e. initial review published in September 2020 (Salameh 2020a), second version 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.

2.5 Discussion

This is the third version of a Cochrane 'living' systematic review evaluating the diagnostic accuracy of thoracic imaging in people with suspected COVID-19.

Summary of main results

In the current version of the review, chest CT (41 studies) gave 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 diagnosing COVID-19 in suspected participants. Compared with the findings of the previous version of the review (Islam 2020), the current version demonstrated similar sensitivity, but notably higher specificity for chest CT. Possible explanations for this improved specificity could include the stricter study inclusion criteria applied for this update, particularly the exclusion of studies that reported 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 had index test readers use well-developed definitions for index test positivity (e.g. CO-RADS). Furthermore, studies from the early stage of the pandemic were included in the 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 the 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.

Based on the sensitivity analysis performed with chest CT studies, publication status had a minimal effect on accuracy estimates. In the previous review, publication status did not appear to contribute to heterogeneity.

There was no statistical evidence indicating an effect of reference standard conduct on chest CT accuracy; 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 the previous review version, 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 analysis in the current review was close to the P=0.05 cut-off for statistical significance, reference standard conduct will continue to be of interest in future updates of the review.

Surprisingly, the definition used for index test positivity did not appear to affect chest CT accuracy, as studies that used a formal scoring system demonstrated comparable sensitivity and specificity 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, when the threshold for index test positivity increased (i.e. from 2 to 5), sensitivity decreased and specificity increased, as expected.

Chest X-ray (9 studies) 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), and ultrasound (5 studies) 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) for diagnosing COVID-19 in suspected participants. As meta-analyses were not performed for chest X-ray or ultrasound in the previous review version due to a limited number of studies, comparisons between 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 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 versions of this review.

Cumulative meta-analyses and time trends

Across all three versions of this review, the sensitivity estimates of chest CT appear to be similar, 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

The search strategy was broad and allowed for identification of a wide range of articles about COVID-19 diagnosis. The eligibility criteria was inclusive, and although screening and exclusion of articles that had been retracted or published in predatory journals were not conducted in this review, risk of bias assessment was performed for each included study to judge the quality of the study design and methods. The review authors screened records, extracted data, and assessed study methodology independently and in duplicate. Furthermore, compared to the first (Salameh 2020a) and second (Islam 2020) versions of the review, this current update includes a greater number of studies that evaluated accuracy of imaging tests in the diagnosis of suspected COVID-19 participants.

Studies with only symptomatic participants, and studies with mixed populations were included. 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 included studies clearly reported information on participants' symptoms.

Reference standard conduct and definition for index test positivity were not identified as sources of variability for chest CT accuracy. These findings may suggest that these factors did not significantly contribute to variability; alternatively, there may be confounding factors obscuring the analyses. Due to insufficient granularity of data, it was not possible to investigate additional potential sources of variability, particularly participant setting (inpatient versus outpatient) and symptom status (symptomatic versus asymptomatic). Furthermore, it was not possible to investigate potential sources of variability for chest X-ray and ultrasound accuracy estimates due to the limited number of studies. These analyses will be performed in future versions of the review when sufficient data become available.

This was the first review version that formally addressed the secondary objective of evaluating threshold effects of index test positivity on accuracy, particularly with respect to the CO-RADS scoring system. Threshold effects for studies that used other scoring systems could not be conducted due to the limited number of included studies.

Indirect comparisons of chest CT, chest X-ray and ultrasound of the lungs were formally assessed. Due to the limited number of these studies that evaluated multiple imaging modalities in the same population, direct comparisons were qualitatively assessed, but could not be formally analyzed. Formal evaluation of direct comparisons of imaging tests are planned for future versions of the review, as more studies with comparative designs become available.

Cumulative meta-analyses were performed for chest CT accuracy estimates using a univariate model, whereas all other meta-analyses in this review were conducted using a bivariate model. The univariate model, which analyzes sensitivity and specificity separately

instead of together, is therefore a limitation of the cumulative meta-analyses and should be considered when interpreting the results of the analyses. The cumulative meta-analyses were not applicable for chest X-ray and ultrasound accuracy, due to the limited number of studies.

It was not possible 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, this will be considered this in future versions of the review.

It was not possible to evaluate several planned additional secondary objectives (McInnes 2020) due to insufficient data. Important questions that remain a concern are:

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

Future versions of this review will explore these objectives 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 impacts 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; investigators are encouraged to refer to the STARD 2015 checklist (Bossuyt 2015; Hong 2018).

A limitation of primary studies that may not be captured in the risk of bias evaluation concerns the recruitment of participants. Of the studies that reported recruitment methods, the majority reported including 'consecutive' participants. However, 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.

In all included studies, RT-PCR was used alone or combination with other criteria to compose 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. As such, the accuracy estimates reported in this review should be interpreted with caution. However, the investigations of heterogeneity for chest CT studies in this review did not identify differing accuracy estimates between studies that used at least two RT-PCR test results to define disease status versus studies that did not have repeat RT-PCR testing. At this stage, despite its limitations, RT-PCR remains the best tool for diagnosing COVID-19.

Three out of 51 included studies (6%) were only available as preprints at the time of the search. In future versions of this review, data extracted from these studies will be updated as they become published in peer-reviewed journals.

Applicability of findings to the review question

The review findings are applicable to individuals suspected to have COVID-19. The search did not identify many studies in pediatric populations and most participants evaluated across all included studies were adults. The results of this review are applicable to adults, as the diagnostic accuracy of these modalities in children is not as well-established. It should be noted that the results apply mostly to imaging interpreted by radiologists. Considering country of study origin as a proxy for the race of participants, the results of this review are likely most generalizable to individuals who are Caucasian. In addition, the lack of data available in the included studies pertaining to signs and symptoms of presenting cases, severity of symptoms, and timing of symptom onset adds complexity to the interpretation of the findings in this review.

Conclusion

Based on evidence until 30 September 2020, chest CT was found to be sensitive and moderately specific for the diagnosis of COVID-19; Chest X-ray was moderately sensitive and moderately specific; and ultrasound was sensitive but not specific. 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. The uncertainty resulting from risk of bias and heterogeneity of included studies limit the confidence in these results.

Figures

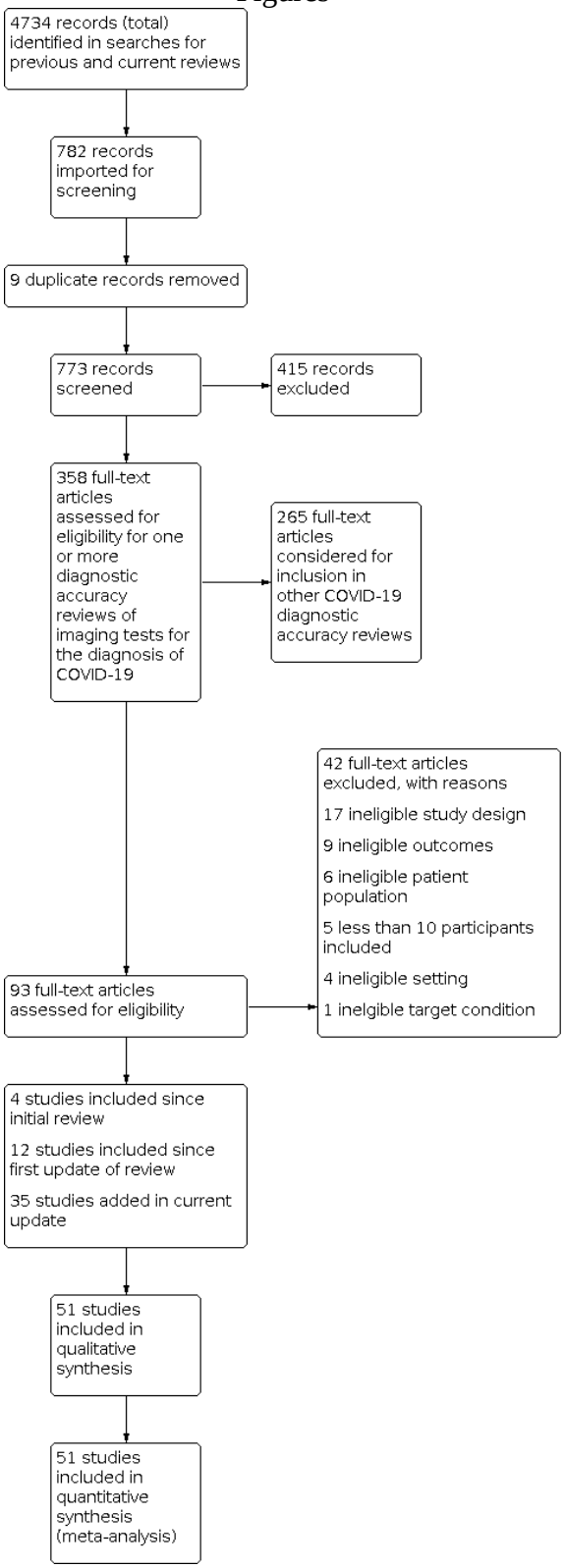


Figure 2.1. Study flow diagram.

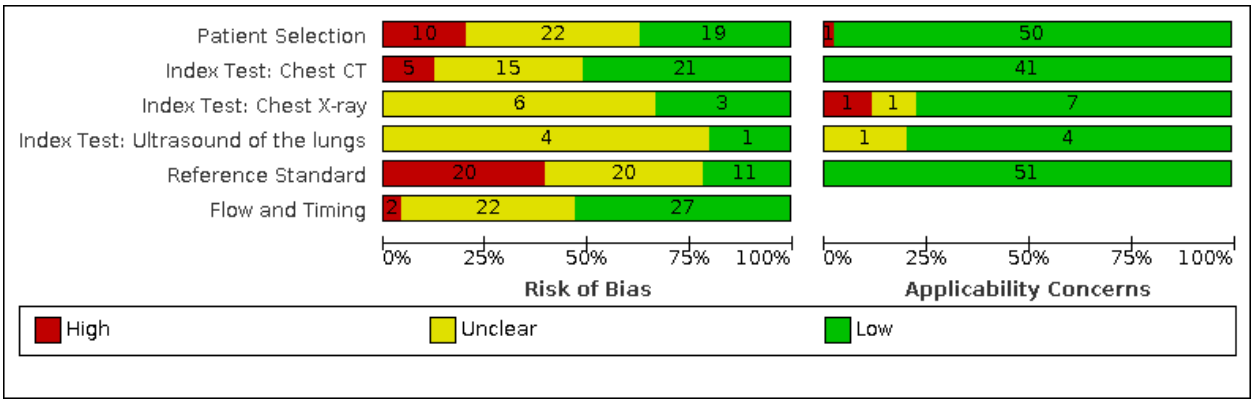


Figure 2.2. Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies.

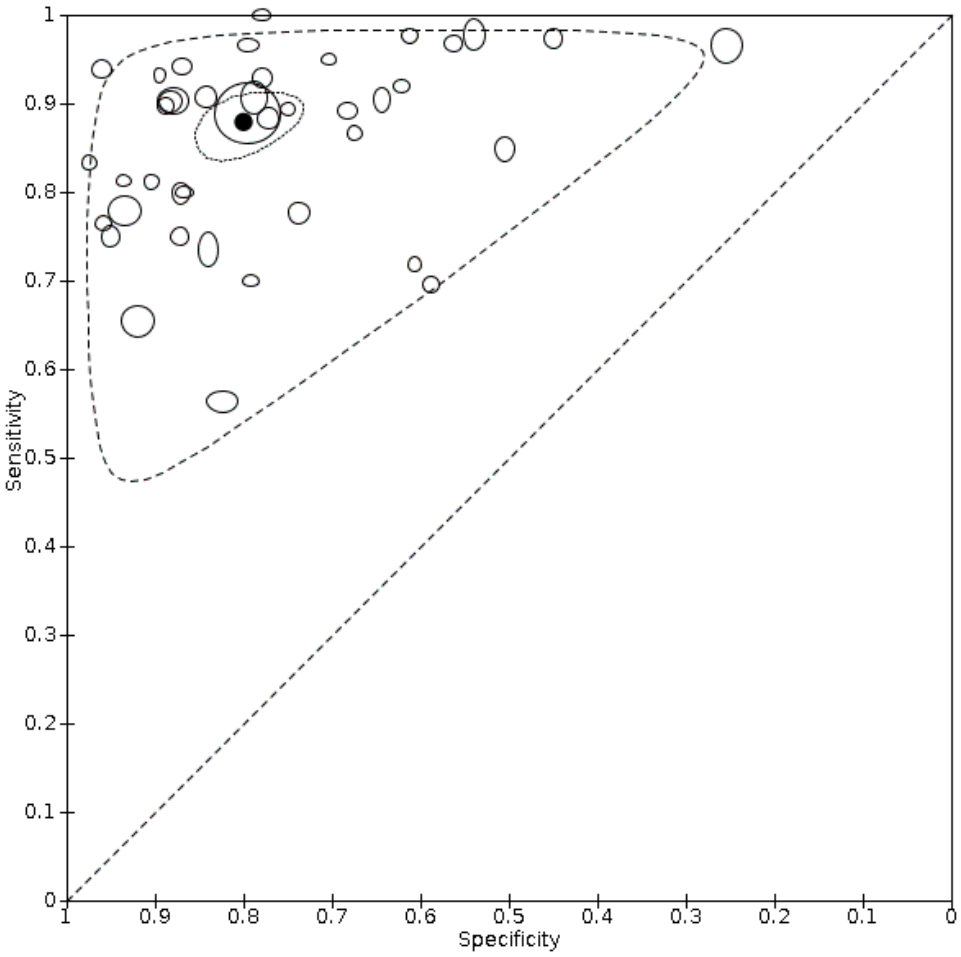


Figure 2.3. Summary ROC plot of chest CT in suspected cases. The summary point is indicated by the solid black circle, individual studies are indicated by outlined circles (scale=study sample size). The dotted border and the dashed border represent 95% confidence regions and 95% prediction regions, respectively.

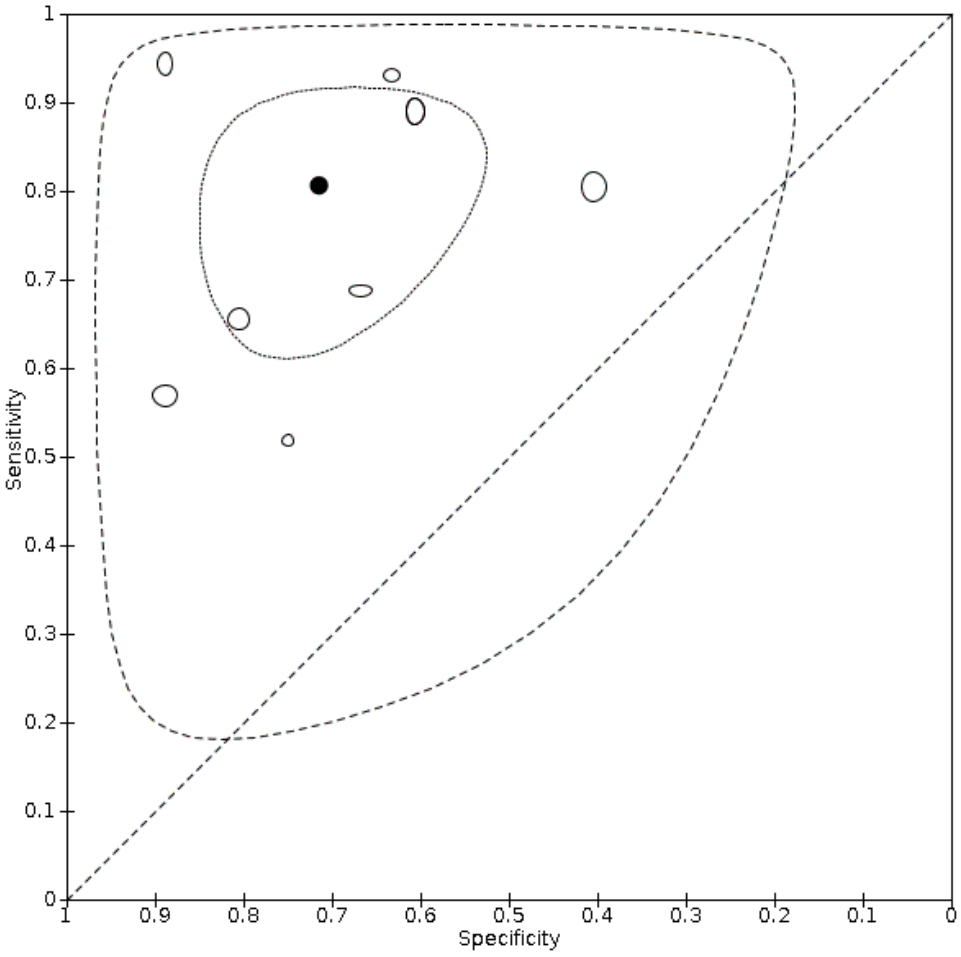


Figure 2.4. Summary ROC plot of chest X-ray in suspected cases. The summary point is indicated by the solid black circle, individual studies are indicated by outlined circles (scale=study sample size). The dotted border and the dashed border represent 95% confidence regions and 95% prediction regions, respectively.

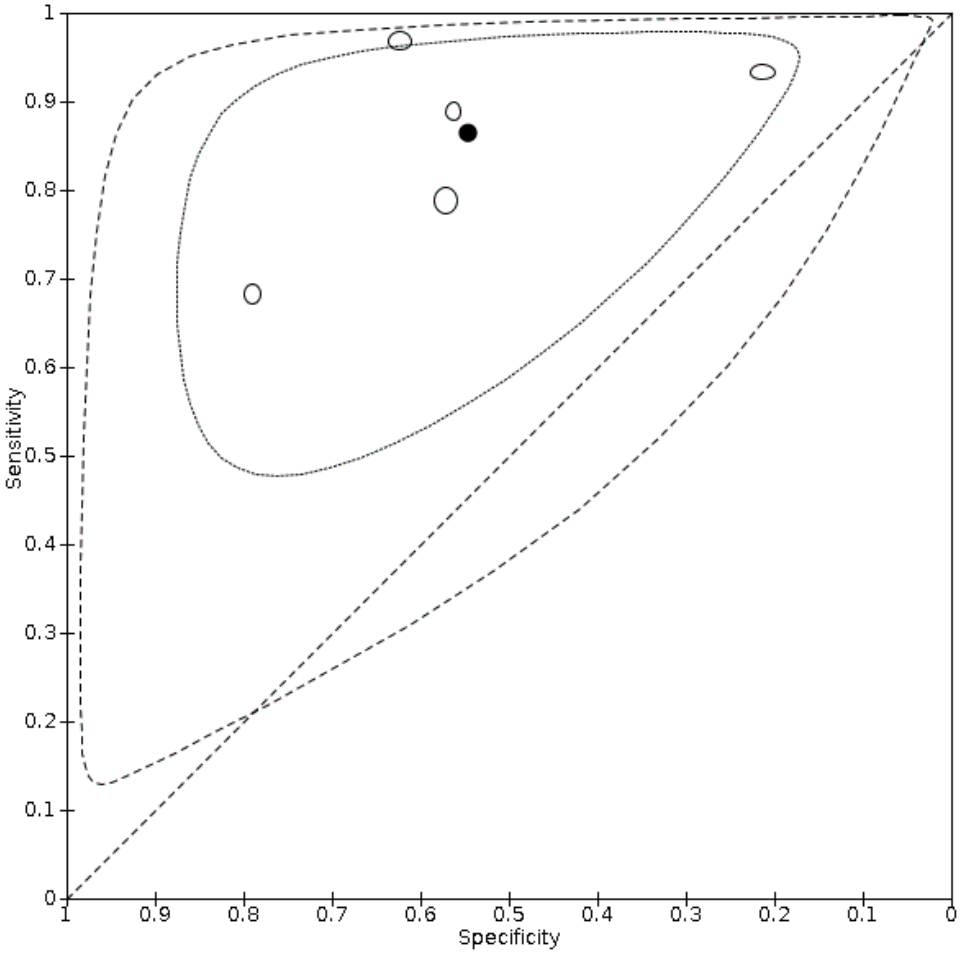


Figure 2.5. Summary ROC plot of ultrasound of the lungs in suspected cases. The summary point is indicated by the solid black circle, individual studies are indicated by outlined circles (scale=study sample size). The dotted border and the dashed border represent 95% confidence regions and 95% prediction regions, respectively.

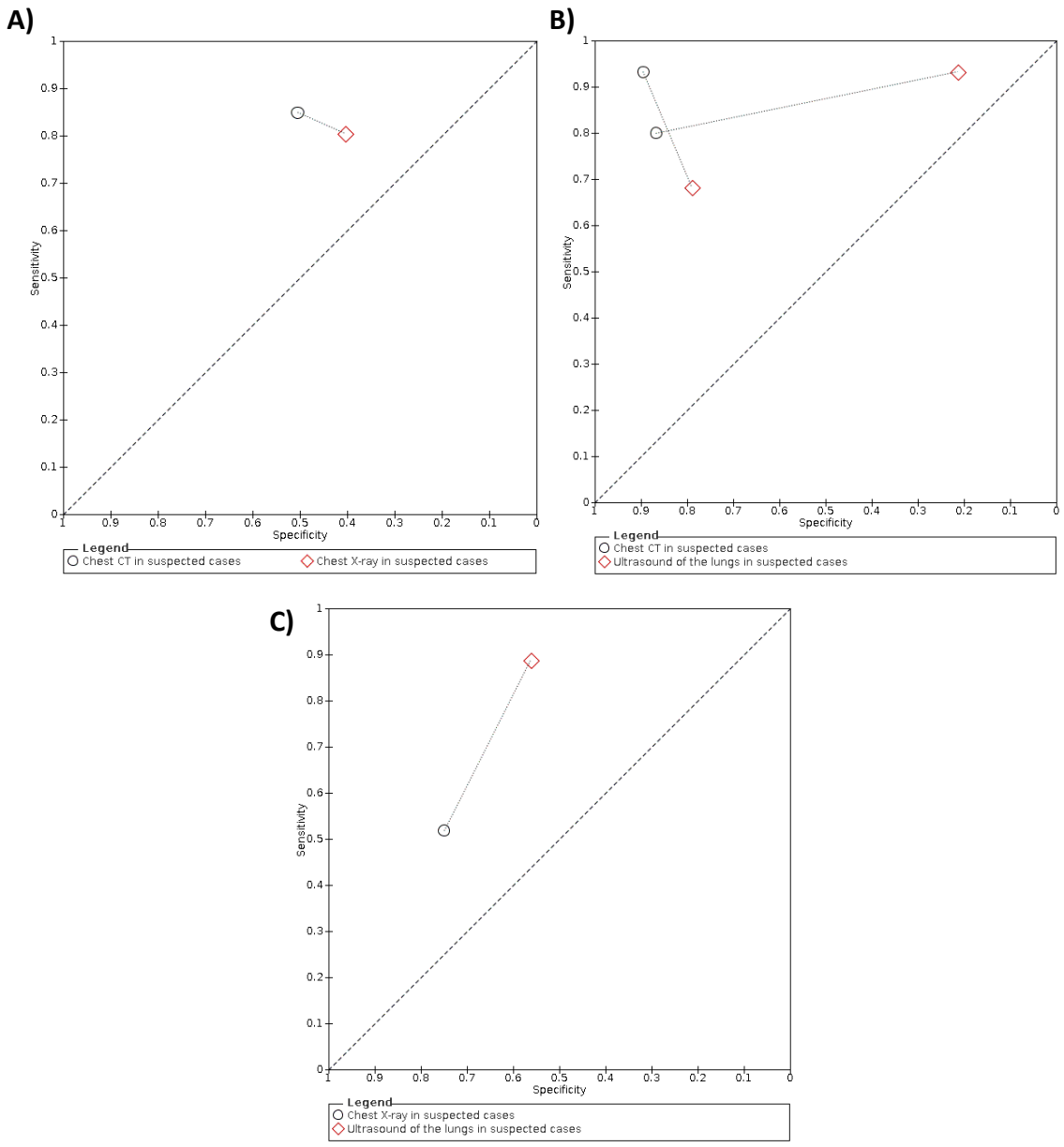


Figure 2.6. Summary ROC plot of comparative studies: A) chest CT versus chest X-ray; B) chest CT versus ultrasound of the lungs; and C) chest X-ray versus ultrasound of the lungs. The dotted line connects index tests pairs evaluated in the same study (i.e. a comparative study).

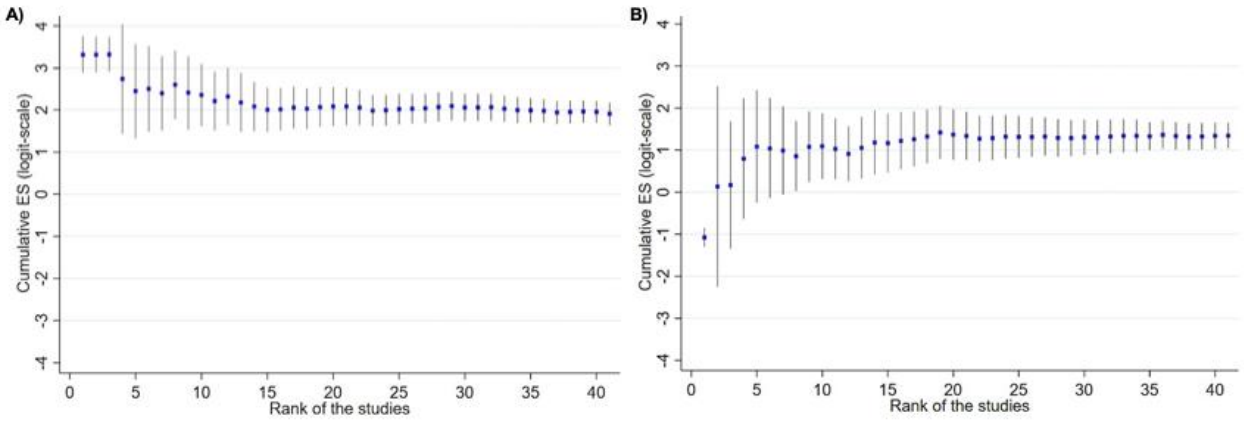


Figure 2.7. Cumulative meta-analysis of chest CT A) logit-sensitivity and B) logit-specificity. Includes 41 studies in the current review. Studies are ranked according to date of publication (1=oldest, 41=most recent).

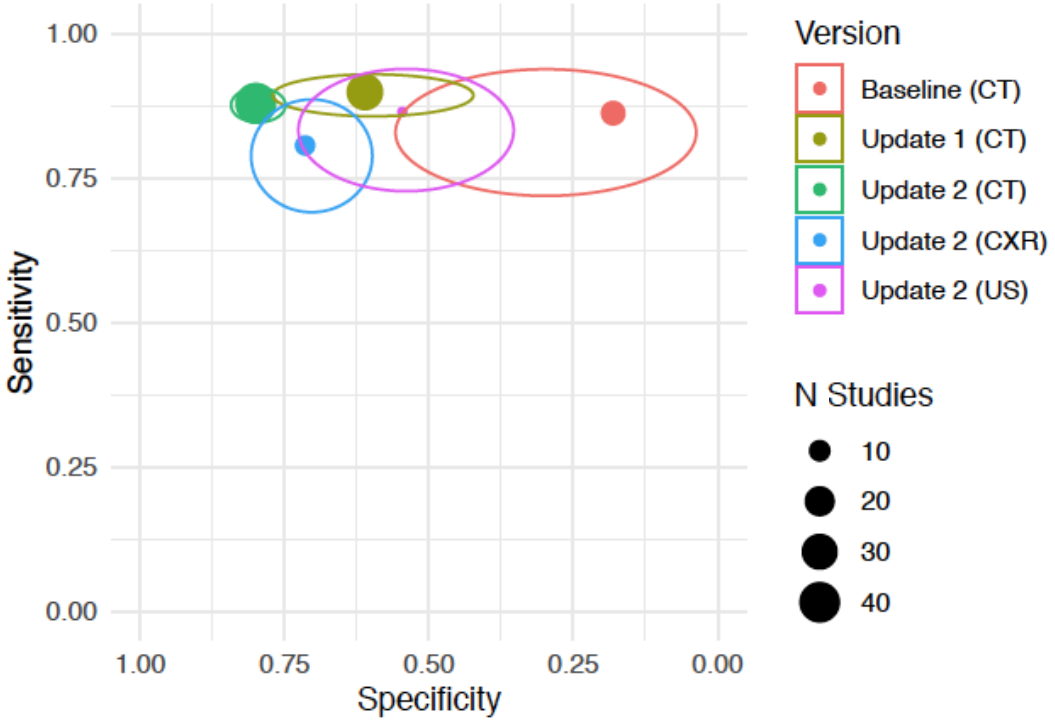


Figure 2.8. Pooled sensitivity and specificity estimates and 95% confidence intervals across all versions of this review.

Tables

Table 2.1. 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 COVID-19 (all symptomatic)	Adults, perhaps also children	Outpatient	Chest CT (non-contrast)	Structured report about the probability of COVID-19 pneumonia: highly suggestive, suggestive, non-suggestive	Radiologist	RT-PCR once; twice in some	0.26	0.9
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

						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

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	Unclear	0.4

								imaging tests)		
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

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 2020	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

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	Radiologist	RT-PCR, no other details provided	Unclear	0.2

						points 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

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

Table 2.2. 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% (78.5-93.8)	71.3% (59.9-80.6)
RT-PCR testing not repeated for all initial negative results	22	11,344 (5779)	86.7% (82.3-90.1)	82.6% (77.8-86.6)
P value			0.74	0.05
Definition for index test positivity				
Radiologist impression	13	7000 (3565)	90.3% (84.5-94.1)	77.2% (67.0-84.9)
Formal scoring system	23	6805 (3333)	85.9% (81.2-89.2)	80.0% (75.0-84.2)
P value			0.15	0.58

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

Table 2.3. 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 (range)	Pooled sensitivity (95% CI)	Specificity (range)	Pooled specificity (95% CI)
5	7	2560 (1042)	41.5% to 77.9%	67.0% (56.4-76.2)	83.5% to 96.2	91.3% (87.6-94.0)
4	7	2560 (1042)	56.3% to 92.9%	83.5% (74.4-89.7)	77.2% to 90.4%	83.6% (80.5-86.4)
3	9	2807 (1139)	65.5% to 98.8%	90.7% (85.2-94.3)	56.6% to 86.9%	69.4% (63.3-74.9)
2	7	2560 (1042)	75.4% to 99.7%	94.3% (88.6-97.2)	26.5% to 57.1%	44.1% (36.5-52.0)
1[a]	-	-	-	-	-	-

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

[a]Meta-analysis for a CO-RADS threshold of 1 was not performed since at this threshold all sensitivity values are equal to 1, and all specificity values are equal to 0.

Table 2.4. 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% (84.6-90.6) 80.0% (74.9-84.3)	80.6% (69.1-88.6) 71.5% (59.8-80.8)	86.4% (72.7-93.9) 54.6% (35.3-72.6)
Chest CT	41 (16,133)	87.9% (84.6-90.6) 80.0% (74.9-84.3)	-	-	-
Chest CT/X-ray	9 (3694)	80.6% (69.1-88.6) 71.5% (59.8-80.8)	P=0.10 P=0.12	-	-
Ultrasound of the lungs	5 (446)	86.4% (72.7-93.9) 54.6% (35.3-72.6)	P=0.77 P=0.0052	P=0.43 P=0.13	-

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

References

References cited in the text

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]

Borakati 2020
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]

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 2021).

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]

Fonsi 2020
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]

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]

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]

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]

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]

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]

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]

Narinx 2020
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.

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]

Pare 2020
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]

Pedersen 2020
Pedersen, TL. Package 'ggforce'. www.ggforce.data-imaginist.com 2020 (accessed 5 January 2021).

Prokop 2020
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.

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]

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]

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.

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

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

References to studies included in this review

Ai 2020a
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
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
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
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
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
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
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
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
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
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
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
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
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
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
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
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
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
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
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
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
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 Interventional Radiology* 2020;26(4):315–22.

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

Kuzan 2020
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 2020
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
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
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
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
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
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
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
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
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
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
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
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
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
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
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
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
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
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.

Chapter 3. Cochrane 'living' systematic review on diagnostic accuracy of imaging for COVID-19: evidence until February 2021

List of Authors

1. Nayaar Islam, BHSc*. School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa. Department of Radiology, University of Ottawa. Clinical Epidemiology Program, Ottawa Hospital Research Institute.
2. Sanam Ebrahimzadeh, PhD*. Department of Radiology, University of Ottawa. Clinical Epidemiology Program, Ottawa Hospital Research Institute.
3. Haben Dawit, BSc. Department of Radiology, University of Ottawa. Clinical Epidemiology Program, Ottawa Hospital Research Institute.
4. Jean-Paul Salameh, MSc. Department of Radiology, University of Ottawa.
5. Sakib Kazi, BSc. Department of Radiology, University of Ottawa.
6. Nicholas Fabiano, BSc. Department of Radiology, University of Ottawa.
7. Lee Treanor, BSc. Department of Radiology, University of Ottawa.
8. Marissa Absi, BSc. Department of Radiology, University of Ottawa.
9. Faraz Ahmad, BSc. Department of Radiology, University of Ottawa.
10. Paul Rooprai, BSc. Department of Radiology, University of Ottawa.
11. Ahmed Al Khalil, BSc. Department of Radiology, University of Ottawa.
12. Kelly Harper, MD. Department of Radiology, University of Ottawa.
13. Neil Kamra, BSc. Department of Radiology, University of Ottawa.
14. Mariska MG Leeflang, PhD. Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam University Medical Centers, University of Amsterdam.
15. Lotty Hooft, PhD. Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University.
16. Christian B van der Pol, MD. Department of Radiology, McMaster University.

17. Ross Prager, MD. Department of Medicine, University of Ottawa.
18. Samanjit S Hare, MD. Department of Radiology, Royal Free London NHS Trust.
19. Carole Dennie, MD. Department of Radiology, University of Ottawa. Department of Medical Imaging, The Ottawa Hospital.
20. René Spijker, PhD. Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam University Medical Centers, University of Amsterdam. Medical Library, Amsterdam UMC, University of Amsterdam, Amsterdam Public Health.
21. Jonathan J Deeks, PhD. Test Evaluation Research Group, Institute of Applied Health Research, University of Birmingham. NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham.
22. Jacqueline Dinnes, PhD. Test Evaluation Research Group, Institute of Applied Health Research, University of Birmingham. NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham.
23. Kevin Jenniskens, PhD. Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam University Medical Centers, University of Amsterdam.
24. Daniël A Korevaar, MD, PhD. Department of Respiratory Medicine, Amsterdam UMC, University of Amsterdam.
25. Jérémie F Cohen, MD, PhD. Obstetrical, Perinatal and Pediatric Epidemiology Research Team (EPOPé), Centre of Research in Epidemiology and StatisticS (CRESS), Inserm UMR1153, Université de Paris.
26. Ann Van den Bruel, MD, PhD. Academic of Primary Care, KU Leuven.

27. Yemisi Takwoingi, PhD. Test Evaluation Research Group, Institute of Applied Health Research, University of Birmingham. NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham.
28. Janneke van de Wijgert, MD, MPH, PhD. Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam University Medical Centers, University of Amsterdam. Institute of Infection, Veterinary, and Ecological Sciences, University of Liverpool
29. Junfeng Wang, PhD. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht.
30. Elena Pena, MD. Department of Radiology, University of Ottawa. Department of Medical Imaging, The Ottawa Hospital.
31. Sandra Sabongui, BSc. Faculty of Medicine, University of Toronto.
32. Matthew DF McInnes, MD, PhD. School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa. Department of Radiology, University of Ottawa. Clinical Epidemiology Program, Ottawa Hospital Research Institute.

*Equal contribution and shared first-authorship for Nayaar Islam and Sanam Ebrahimzadeh

Preface

Objective

To assess the diagnostic accuracy of thoracic imaging (computed tomography (CT), chest X-ray and ultrasound) in the evaluation of people with suspected COVID-19, based on compiled evidence available until 17 February 2021.

Funding

1. Frederick Banting and Charles Best Canada Graduate Scholarship - Master's (Canadian Institutes of Health Research)
2. University of Ottawa Faculty of Medicine COVID-19 Pandemic Response Funding Program
3. Government of Ontario Ministry of Health COVID-19 Rapid Response Research Grant
4. The Cochrane Infectious Diseases Group (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).

Role of funders

None of the listed funding bodies had any role in designing or conducting the study; collecting, managing, analyzing, or interpretating the data; or preparing, reviewing, or approving the manuscript.

Appendices

Four appendices (3.1 to 3.4) are included at the end of this document.

Ethics approval

This study did not require ethics approval.

Contributions of co-authors

- Concept and design: Islam, Ebrahimzadeh, Salameh, McInnes.
- Search strategy: Spijker.
- Acquisition of data: Islam, Ebrahimzadeh, Dawit, Salameh, Kazi, Fabiano, Treanor, Absi, Ahmad, Rooprai, Al Khalil, Harper, Kamra, van der Pol, Prager, Jenniskens, Korevaar, Cohen, Wang, Pena, Sabongui, McInnes.
- Analysis and interpretation of data: Islam, Ebrahimzadeh, Takwoingi.
- Drafting of the manuscript: Islam, Ebrahimzadeh, McInnes.
- Critical revision of the manuscript for important intellectual content: Leeflang, Hooft, van der Pol, Prager, Hare, Dennie, Deeks, Dinnes, Korevaar, Cohen, Van den Bruel, van de Wijgert.
- Administrative, technical, or material support: McInnes.

Citation details

Islam N*, Ebrahimzadeh S*, Dawit H, Salameh J-P, Kazi S, Fabiano N, Treanor L, Absi M, Ahmad F, Rooprai P, Al Khalil A, Harper K, Kamra N, 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, Wang J, Pena E, Sabongui S, McInnes MDF. Thoracic imaging tests for the diagnosis of COVID-19. Cochrane Database of

Systematic Reviews 2022, Issue 5. Art. No.: CD013639. DOI:

10.1002/14651858.CD013639.pub5

Abstract

Background: The previous version of this review showed thoracic imaging (specifically, chest computed tomography (CT)) to be sensitive and moderately specific in diagnosing COVID-19, based on evidence until 30 September 2020. This updated version of the review includes new relevant studies.

Objectives: To evaluate the diagnostic accuracy of thoracic imaging in people with suspected COVID-19; assess the rate of positive imaging in people with an initial reverse transcriptase polymerase chain reaction (RT-PCR) negative result and a positive RT-PCR result on follow-up; and evaluate the accuracy of thoracic imaging for screening COVID-19 in asymptomatic individuals. The secondary objectives were to assess sources of heterogeneity and evaluate threshold effects of index test positivity on accuracy.

Methods: The Bern COVID-19 Living Database, Cochrane COVID-19 Register, CDC Library, and repositories of COVID-19 publications were searched through to 17 February 2021. Language restrictions were not applied. Studies of all designs, except for case-control, that recruited participants of any age group suspected to have COVID-19, assessed chest CT, chest X-ray, or ultrasound of the lungs for the diagnosis of COVID-19, and reported estimates of test accuracy were included. Studies were excluded if they excluded participants with normal index test results, if the definition used for a positive index test was unclear, if the reference standard did not include RT-PCR, or if imaging was a part of the reference standard. The review authors independently and in duplicate screened articles, extracted data and assessed risk of bias and applicability concerns using QUADAS-2. A bivariate meta-analysis model was used where

appropriate to determine pooled estimates of sensitivity and specificity and 95% confidence intervals (CIs).

Results: Ninety-eight studies were included across all objectives; 94 evaluated the diagnostic accuracy of thoracic imaging in people with suspected COVID-19; eight assessed the rate of positive imaging in individuals with initial RT-PCR negative results and positive RT-PCR results on follow-up; and 10 evaluated the accuracy of thoracic imaging for imaging asymptomatic individuals. For all 98 included studies, risk of bias was high or unclear in 52 (53%) for participant selection, 64 (65%) studies for reference standard, 46 (47%) studies for index test, and 48 (49%) studies for flow and timing.

Of the 94 studies evaluating diagnostic accuracy of imaging in suspected individuals, 87 evaluated one imaging modality, and seven evaluated two each. All studies used RT-PCR alone or in combination with other criteria as the reference standard for diagnosing COVID-19. Chest CT (69 studies, 28285 participants, 14342 (51%) cases) had a pooled sensitivity of 86.9% (95%CI 83.6-89.6) and specificity of 78.3% (73.7-82.3). Chest X-ray (17 studies, 8529 participants, 5303 (62%) cases) had a pooled sensitivity of 73.1% (64.1-80.5) and specificity of 73.3% (61.9-82.2). Ultrasound of the lungs (15 studies, 2410 participants, 1158 (48%) cases) had a pooled sensitivity of 88.9% (84.9-92.0) and specificity of 72.2% (58.8-82.5). Based on an indirect comparison using all 94 studies, chest CT and ultrasound gave similar sensitivity estimates (P=0.42), and each gave higher sensitivities than X-ray (P=0.0003 and P=0.001, respectively). All modalities had similar specificities (P≥0.05).

Conclusion: Chest CT and ultrasound of the lungs are sensitive and moderately specific in diagnosing COVID-19. Chest X-ray is moderately sensitive and moderately specific in diagnosing COVID-19. Thus, chest CT and ultrasound may have more utility for ruling out COVID-19 than for differentiating SARS-CoV-2 infection from other causes of respiratory illness. The uncertainty resulting from risk of bias and heterogeneity of included studies limit the confidence of these results.

3.1 Introduction

The value of imaging tests in the diagnostic pathway for COVID-19 has continued to be of interest during the COVID-19 pandemic. While there remains some uncertainties about the role of imaging in COVID-19 patients, research on the topic 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). Also, asymptomatic transmission of COVID-19 is one of its biggest diagnostics challenges, and the World Health Organization (WHO) has recently emphasized the distinction between asymptomatic patients and presymptomatic patients (Walker 2020). The role of imaging in screening asymptomatic patients is currently unknown.

This review from the suite of Cochrane 'living' systematic reviews summarizes evidence on the accuracy of different imaging tests and diagnostic features in participants regardless of their symptoms. Details on the target condition (COVID-19), index tests, and clinical pathway are included in Appendix 3.1.

Summary of previous versions of the review

In the first version of the review (studies until 5 May 2020), thirteen studies that assessed chest CT in participants with suspected COVID-19 demonstrated a sensitivity of 86.2% (95% confidence interval (CI) 71.9-93.8) but a low specificity of 18.1% (95%CI 3.71-55.8) (Salameh 2020a). The second version the review (studies until 22 June 2020) focused on people suspected of having COVID-19 and excluded studies evaluating only confirmed cases of COVID-19 (Islam 2020); 31 studies that evaluated chest CT in suspected participants demonstrated a pooled sensitivity of 89.9% (95%CI 85.7-92.9) and specificity of 61.1% (95%CI 42.3-77.1). Chest X-ray and ultrasound accuracies could not be evaluated in the first two reviews due to limited data.

Compared to the second version of the review, the third version had stricter inclusion criteria, excluding studies of case-control design and those that did not explicitly classify the imaging test as either COVID-19 positive or negative. Chest CT (41 studies) had a pooled sensitivity of 87.9% (95%CI 84.6-90.6) and specificity of 80.0% (95%CI 74.9-84.3). Chest X-ray (nine studies) had a pooled sensitivity of 80.6% (95%CI 69.1-88.6) and specificity of 71.5% (95%CI 59.8-80.8). Ultrasound (five studies) had a pooled sensitivity of 86.4% (95%CI 72.7-93.9) and specificity of 54.6% (95%CI 35.3-72.6). Definition of index test positivity and reference standard conduct were not found to impact chest CT accuracy.

For this current update (fourth version of the review), the inclusion criteria have been further refined, and now exclude studies that used imaging as a reference standard or excluded participants with normal index test results. The impact of definition of index test positivity on the accuracy of X-ray and ultrasound has been formally assessed for the first time, along with chest CT. The rate of positive imaging in people who had an initial RT-PCR negative result and a positive RT-PCR result on follow-up, and the accuracy of imaging for screening for COVID-19 in asymptomatic individuals were also assessed.

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 second version of this review, the value of formal scoring systems was explored but could not be formally analyzed due to limited data (Islam 2020). In the third version of the review, the value of formal scoring systems on accuracy estimates of imaging tests (Irwig 1995) and threshold effects of the CO-RADS scoring system for chest CT studies were evaluated (Islam 2021). Since the third version, more studies have been published which have comparative designs and compare different imaging modalities; evaluate the rate of positive imaging in those with initial RT-PCR negative results and positive RT-PCR results on follow-up; and assess the accuracy of imaging for screening asymptomatic individuals.

3.2 Objectives

The primary objectives are 1) to evaluate the diagnostic accuracy of thoracic imaging (chest CT, chest X-ray and ultrasound) in people with suspected COVID-19; 2) to assess the rate of positive imaging in individuals with initial RT-PCR negative results and positive RT-PCR results on follow-up; and 3) to evaluate the accuracy of thoracic imaging for screening asymptomatic individuals.

The secondary objectives are to evaluate: 1) sources of heterogeneity for accuracy estimates; and 2) 'threshold' effects of index test positivity on accuracy.

3.3 Methods

The protocol for this review has been published in the Cochrane Database (McInnes 2020). The methods of this review version are largely similar to those of the previous version (section 2.3 Methods), with some changes, including refined inclusion criteria. Protocol deviations are outlined in Appendix 3.2.

Criteria for considering studies for this review

Types of studies

Studies of all designs, except for case-control studies, that included participants suspected of having the target condition and produced estimates of test accuracy or provided 2x2 data (true positive (TP), true negative (TN), false positive (FP), false negative (FN)) were included. All settings, languages, and all variations of a test were accepted. If data were not available, study authors were contacted 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

Studies with individuals suspected of having COVID-19 as outlined in the 'Target condition' section in Appendix 3.1; studies could have 'symptomatic populations' or 'mixed populations' (asymptomatic and symptomatic participants). There were no age or gender restrictions. 'Asymptomatic populations' were also included for the objective on imaging of asymptomatic individuals in this review.

To reduce the effect of selection bias, studies were excluded if they excluded participants who had normal index test results.

Index tests

The index tests were chest CT, chest X-ray, or ultrasound of the lungs, meeting the criteria described in the 'Index tests' section in Appendix 3.1. 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.

Only index tests interpreted by humans were included. Studies involving interpretation by an algorithm (machine learning/artificial intelligence (AI)) were included only if they also provided data pertaining to diagnostic accuracy of human interpretation.

Definitions of index test positivity

Inclusion was limited to 'diagnostic test accuracy studies' in which the index test readers either (1) used a radiological scoring system (e.g. CO-RADS), or (2) explicitly classified patients as being positive or negative for COVID-19 based on imaging. Studies that reported an overview of index test findings without explicitly classifying the imaging test as either COVID-19 positive or negative were excluded.

There has been considerable heterogeneity and changes over time in the definitions used for positive imaging findings. Some groups have used constellations of specific findings (such as multiple peripheral ground-glass opacities on CT), some have considered the combined effect of specific findings (a 'gestalt' approach), and some have used formal scoring systems, such as CO-RADS (5 categories; Prokop 2020). As such, inclusion was not limited to predefined definition or threshold for positivity; instead, the definition for positivity used in each study was extracted. However, studies that did not clearly report the definition for a positive index test were excluded.

Target conditions

The target condition was COVID-19 (Appendix 3.1). However, all studies reporting data on COVID-19 or COVID-19 pneumonia that provided data relevant to the review objectives were included.

Reference standards

A positive diagnosis for COVID-19 had to be confirmed by 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), alone or in combination with other criteria as follows.

- positive on WHO criteria for COVID-19
- positive on China CDC criteria for COVID-19
- positive serology for SARS-CoV-2 antibodies in addition to consistent symptomatology
- positive on study-specific list of criteria for COVID-19 (e.g., symptoms, imaging findings, other tests, infected contacts)

A negative diagnosis for COVID-19 had to be confirmed by the following.

- suspected COVID-19 with negative RT-PCR test results, whether tested once or more than once

Studies that did not include RT-PCR as a part of the reference standard were excluded. Studies that used imaging as a part of the reference standard were excluded because of a risk of incorporation bias.

Methodological quality was assessed based on how likely the reference standard definition used in each study correctly classified 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 3.3.

Search methods for identification of studies

Electronic searches through 17 February 2021 were conducted using three main resources:

1. Living search from the University of Bern
2. Cochrane COVID-19 Study Register searches
3. The Stephen B. Thacker CDC Library, COVID-19 Research Articles Downloadable Database

The following repositories of COVID-19 publications were checked against these search results.

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, company and product websites for studies about test accuracy were searched.
4. Companies were contacted to ask for further information about studies.
5. Research groups completing test evaluations (e.g. UK Public Health England-funded studies, Foundation for Innovative New Diagnostics (FIND) studies) were also contacted.

The same search strategies that were previously devised with the help of an experienced Cochrane Information Specialist with diagnostic test accuracy expertise (RSp) and used in the previous version of the review (outlined in Appendix 2.3) was applied with an updated search 'end date' of 17 February 2022. These searches aimed to identify all articles related to COVID-19 and SARS-CoV-2 and were not restricted to those evaluating imaging tests; the searches used no terms that specifically focused on an index test, diagnostic accuracy, or study methodology.

A search classification model using artificial intelligence text analysis was implemented from 25 May 2020 and onwards, due to the increased volume of published and preprint articles (this was also used in the previous version of the review; outlined in Appendix 2.3).

Selection of studies

The review authors screened studies independently, in duplicate. A third, experienced review author resolved disagreements about initial title and abstract screening. Disagreements about eligibility assessments during full-text screening were resolved 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, 2x2 contingency tables (numbers of TP, TN, FP, FN) were extracted. If a study reported accuracy data for more than one index test reader, the average of the data from all readers were taken 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, only the 2x2 contingency table corresponding to the radiologist accuracy data was extracted. If a study used multiple reference standards, but 2x2 contingency tables that included only RT-PCR as the reference standard could be computed, the latter were extracted and analyzed. If a study reported accuracy data for multiple thresholds of index test positivity (e.g. studies that used the CO-RADS scoring system), 2x2 contingency tables for all available thresholds were extracted.

Two of the 11 studies that used the CO-RADS scoring system did not report the 2x2 data for all five CO-RADS thresholds; the corresponding authors were contacted but the data could not be obtained. For these two studies, it was possible to extract data for a CO-RADS threshold of 3 only. One of the five studies that used the RSNA scoring system did not report the 2x2 data for all four RSNA thresholds; the corresponding authors were contacted but the data could not be obtained. For this study, it was possible to extract data for RSNA thresholds 3 and 4 only.

In addition, the following items were extracted.

- 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
- Study design
- Imaging timing relative to disease course
- CT, chest X-ray and ultrasound findings

- Criteria for ‘positive’ diagnosis of COVID-19 on imaging
- Index test technical parameters
- 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).
- Details regarding interpretation of the index test (level of training, number of readers, the inter-observer variability)
- The number of TP, TN, FP, FN or summary statistics from which they can be computed.
- 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 tool (Appendix 3.3) which includes four domains: participant selection, index test(s), reference standard(s), flow and timing. Three review authors resolved any disagreements through discussion.

Statistical analysis and data synthesis

Estimates of sensitivity and specificity were presented using paired forest plots and pooled estimates summarized in tables. Data were analyzed on a participant level.

A bivariate model for meta-analyses, which takes into account the within- and between-study variance, and the correlation between sensitivity and specificity across studies (Chu 2006; Reitsma 2005), was applied using metandi in STATA (Harbord 2009; StataCorp 2019). Meta-analyses were performed when four or more studies evaluated a given modality. Sensitivity

analyses were performed by limiting inclusion in the meta-analysis to studies published in peer-reviewed journals.

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

The same meta-analysis methods were used for all primary and secondary objectives (metandi and meqrlogit in STATA, specifically).

In addition, for studies that evaluated the rate of positive chest CT imaging in repeat RT-PCR positive results, the rates of positive imaging per study were presented using forest plots.

Investigations of heterogeneity

Heterogeneity was assessed by visual inspection of paired forest plots and summary receiver operating characteristics (SROC) plots. Meta-regression was performed using meqrlogit in STATA (StataCorp 2019) 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. The variable of interest was added as a covariate to a bivariate model; using the model parameters, postestimation command was used to compute absolute differences in pooled sensitivity and specificity and 95% CIs were obtained using the delta method. P values were obtained using the Wald test.

The impact of reference standard conduct (RT-PCR performed at least twice in all participants with initial negative results versus RT-PCR not done twice) on accuracy was assessed in chest CT studies. The impact of definition for index test positivity (radiologist

impression versus formal scoring system) on accuracy was assessed in chest CT, chest X-ray and ultrasound studies.

Threshold effects

Meta-analyses were performed 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 and chest CT studies at RSNA thresholds 2, 3 and 4).

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

Indirect test comparisons

This was performed 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. P values were obtained using the Wald test. It should be noted that there were not enough studies to evaluate direct comparisons.

Time trends

A plot displaying meta-analysis results (i.e. pooled sensitivity and specificity estimates) across the first (Salameh 2020a), second (Islam 2020) third (Islam 2021), and current versions of this review were generated using ggplot2 and ggforce in R (Wickham 2016; Pedersen 2020; R Core Team 2021).

Assessment of reporting bias

Tests for publication bias and formal assessments of reporting bias were not performed (Salameh 2020b; Moher 2009).

3.4 Results

Results of the search

The PRISMA flow diagram is shown in Figure 3.1 (Salameh 2020b; Moher 2009). The search identified 7734 results, 965 unique references (published or preprint studies) were screened for inclusion, and 188 records were selected for full-text assessment. A total of 98 studies were included in this review for all objectives. Of these, 94 were included to evaluate the diagnostic accuracy of thoracic imaging in people with suspected COVID-19, of which four have been included since the first review version (Salameh 2020a), and 12 have been included since the second review version (Islam 2020), and 29 have been included since the third review version (Islam 2021). Furthermore, of the 98 studies in this review, eight assessed the rate of positive imaging in individuals with initial RT-PCR negative results and positive RT-PCR results on follow-up, and 10 evaluated the accuracy of thoracic imaging for imagining asymptomatic individuals.

Description of included studies (diagnostic accuracy in suspected participants)

Ninety-four studies (64 CT, 12 X-ray, 11 ultrasounds, three both CT and X-ray, two both CT and ultrasound, and two both X-ray and ultrasound) were included, with a total of 37,631 participants suspected of having COVID-19, of whom 19768 (53%) had a final diagnosis of COVID-19.

The median sample size was 234 (interquartile range (IQR) 101.25 to 478.75). Sixty-five studies were conducted in Europe (Italy 19, the Netherlands 9, France 9, Turkey 6, Belgium 5, Germany 7, UK 4, Switzerland 2, Czech Republic 1, Ireland 1, Spain 1, Denmark 1), 19 were conducted in Asia (China 9, Korea 1, India 4, Iran 2, Japan 1, Pakistan 1, United Arab Emirates 1), and the remaining studies were conducted in North America (USA 6, Canada 1) and South America (Brazil 3).

Index test readings were performed by radiologists in 49 studies (52%), radiology residents in two (2%), both radiologists and residents in three (4%), and both radiographers and radiologists in one (1%); 39 studies (37%) did not clearly report the level of training of readers. Technical parameters regarding the protocol of chest CT used in 69 studies were not clearly reported in 31 (44%) studies, while non-contrast CT was used in 25 (36%), high-resolution chest CT was used in eight (11%), low-dose CT with or without contrast was used in 11 (15%) and CT with IV contrast was used in five (7%).

Manuscripts of three (3%) of the studies were available only as preprints at the time of the search.

Characteristics of the included studies are summarized in Table 3.1.

Participant characteristics (diagnostic accuracy in suspected participants)

Fifty-seven studies included only adult participants (aged 16 years and over), 32 included both children and adults (although in most cases, only a minority of included patients were children), one included only children, one included participants aged 70 years and older, and the remaining three did not clearly report the age range of participants. All participants were suspected of having COVID-19. Seventy (74%) studies involved only symptomatic participants,

20 (21%) involved both symptomatic and asymptomatic participants, and four (4%) did not clearly report participants' symptom status.

All 94 studies used RT-PCR as the reference standard for the diagnosis of COVID-19, with 82 using only RT-PCR as the reference standard and seven using a combination of RT-PCR and other criteria (laboratory tests 2, clinical signs and symptoms 2, clinical signs on follow-up 1, positive contacts 1, and follow-up phone calls 1) as the reference standard. With respect to RT-PCR testing, eight studies tested each participant once, 42 tested some participants with initial negative RT-PCR results at least twice, 19 tested all participants with initial negative RT-PCR results at least twice, and 25 did not report on the frequency of testing per participant.

Seventeen studies included inpatients, 65 included outpatients, one included both in- and outpatients, while the remaining 23 were conducted in unclear settings. Thirty-three (35%) 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.

Description of included studies (positive imaging in repeat RT-PCR positive results)

Eight studies were included for this objective (seven CT, and one ultrasound), with a total of 198 participants suspected of having COVID-19, all of whom had a final diagnosis of COVID-19. All studies were also included for the objective on diagnostic accuracy of imaging in suspected participants.

Seven studies were conducted in Europe (Italy 4, France 2, Belgium 1), and one was conducted in Asia (China). Index test readings were performed by radiologists in five studies (62%), while three (37%) did not clearly report the level of training of readers. Technical

parameters regarding the protocol of chest CT were not clearly reported in two (29%) studies, while non-contrast CT was used in four (57%), and low-dose CT with or without contrast was used in one (14%). Characteristics of the included studies are summarized in Table 3.2.

Participant characteristics (positive imaging in repeat RT-PCR positive results)

Five studies included only adult participants (aged 16 years and over), and three included both children and adults. Most (75%) studies involved only symptomatic participants. All the studies used RT-PCR as the reference standard for the diagnosis of COVID-19; one study tested all participants with initial negative RT-PCR results at least twice, and seven studies tested some participants with initial negative RT-PCR results at least twice. Five studies included outpatients, two included inpatients, and one was conducted in an unclear setting. Three (37%) studies described the co-morbidities of the study population, which included hypertension, cardiovascular disease, diabetes, and asthma. However, the overall presence of co-morbidities in the participant groups of these studies was unclear.

Description of included studies (imaging asymptomatic individuals)

Ten studies were included (seven CT, one X-ray, two ultrasound) with 2007 participants suspected of having COVID-19, of whom 127 (6%) had a final diagnosis of COVID-19. For example, one study included for this objective involved patients who had preoperative chest CT (Gumus 2020). Of these 10 studies, six were also included for the objective on diagnostic accuracy of imaging in suspected participants.

Eight studies were conducted in Europe (Italy 1, UK 2, Belgium 2, the Netherlands 1, Turkey 3), and one was conducted in Korea. Index test readings were performed by radiologists

in three studies (30%), and both radiologist and resident in one (10%); six studies (60%) did not clearly report the level of training of readers. Technical parameters regarding the protocol of chest CT were not clearly reported in six (60%) studies, while non-contrast CT was used in two (20%), low-dose CT with or without contrast was used in one (10%) and high resolution in one (10%). Characteristics of the included studies are summarized in Table 3.3.

Participant characteristics (imaging asymptomatic individuals)

Six studies included only adult participants (aged 16 years and over), three included both children and adults, and one included 70 years of age and older. All the studies used RT-PCR as the reference standard for the diagnosis of COVID-19. With respect to RT-PCR testing, two studies tested each participant once, one study tested all participants with initial negative RT-PCR results at least twice, five studies tested some participants with initial negative RT-PCR results at least twice, and two studies did not report on the frequency of testing per participant. Three studies included outpatients, five included inpatients, and two were conducted in unclear settings. One study (10%) described the co-morbidities of the study population, which included hypertension, kidney disease, heart failure, and diabetes; however, the overall presence of co-morbidities in the participant groups of these studies was unclear

Index tests

With respect to the objective on diagnostic accuracy of imaging in suspected participants (94 studies), 87 evaluated a single imaging modality and seven evaluated two imaging modalities each. In total, the 94 studies reported a total of 101 imaging modality evaluations. Chest CT was

evaluated in 69 studies, chest X-ray was evaluated in 17 studies, and ultrasound of the lungs was evaluated in 15 studies.

For the objective on positive imaging in repeat RT-PCR positive results, all studies evaluated a single imaging modality each. Chest CT was evaluated in seven studies and ultrasound of the lungs was evaluated in one study.

For the objective on accuracy of imaging for asymptomatic individuals, all studies evaluated a single imaging modality. Chest CT was evaluated in seven studies, chest X-ray was evaluated in one study, and ultrasound of the lungs was evaluated in two studies.

Methodological quality of included studies

Figure 3.2 provides a summary of the overall methodological quality assessment using the QUADAS-2 tool for all 98 included studies. Across all 98 included studies, risk of bias because of participant selection was high in 10 (10%) studies due to inappropriate exclusions (n=10). Risk of bias for: chest CT (73 studies) was high in six (8%) studies; chest X-ray (17 studies) was unclear in seven (41%) studies; and ultrasound of the lungs (15 studies) was unclear in six (37%) studies. Risk of bias based on concerns about the reference standard was high in 25 (26%) studies that 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 nine (9%) studies that did not provide the same reference standard to all participants (n=3), or did not have an appropriate time interval between the reference standard and index test (n=6). Figure 3.A1 in Appendix 3.4 displays a study-level quality assessment.

For the objective on rate of positive imaging in repeat RT-PCR positive results (8 studies), there was high risk of bias in 2 (25%) and unclear risk of bias in 6 (75%) of studies with respect to participant selection.

Findings

Pooled estimates in suspected individuals

The sensitivity of CT in 69 studies (14342 (51%) cases in 28285 participants) ranged from 45% to 100%, and the specificity ranged from 10% to 99% (forest plot presented in Figure 3.A2 in Appendix 3.4). The pooled sensitivity for chest CT was 86.9% (95%CI 83.6-89.6), and the pooled specificity was 78.3% (95%CI 73.7-82.3). The scatter of the study points in ROC space on the SROC plot (Figure 3.3) shows substantial variability in sensitivity and specificity.

The sensitivity of chest X-ray in 17 studies (5303 (62%) cases in 8529 participants) ranged from 44% to 94% and the specificity ranged from 24% to 93%. The pooled sensitivity for chest X-ray was 73.1% (95%CI 64.1-80.5) and the pooled specificity was 73.3% (95%CI 61.9-82.2). The scatter of the study points in ROC space on the SROC plot (Figure 3.4) shows substantial variability in sensitivity, and specificity. The sensitivity of ultrasound of the lungs in 15 studies (1158 (49%) cases in 2410 participants) ranged from 73% to 94% and the specificity ranged from 21% to 98%. The pooled sensitivity for ultrasound was 88.9% (95%CI 84.9-92.0), and the pooled specificity was 72.2% (95%CI 58.8-82.5). The scatter of the study points in ROC space on the SROC plot (Figure 3.5) shows substantial variability in sensitivity and specificity for ultrasound of the lungs. The forest plots for chest X-ray and ultrasound of the lungs are presented in Figure 3.A3 in Appendix 3.4.

Sensitivity analyses

For CT studies with suspected participants, excluding three studies published as preprints did not affect summary sensitivity and specificity; studies published in peer-reviewed journals (n=66) had a pooled sensitivity of 87.5% (95%CI 84.3-90.1) and a pooled specificity of 78.0% (95%CI 72.9-82.4). The publication status of studies has been updated as of 17 February 2021.

Investigations of heterogeneity

The results of the investigations of heterogeneity are outlined in Table 3.4. Reference standard conduct (RT-PCR testing at least twice versus not twice for all participants with initial negative results) did not have an impact ($P \geq 0.05$) on accuracy of chest CT. For definition for index test positivity, results defined based on radiologist's impression gave a higher sensitivity $P=0.037$ compared to those based on a formal scoring system among chest CT studies. Definition for index test positivity did not impact ($P \geq 0.05$) the specificity of chest CT, or the accuracies of chest X-ray or ultrasound.

Threshold effects

Eleven studies that evaluated CT used the CO-RADS scoring system (1769 (40%) cases amongst 4416 participants) to define index test positivity. For nine studies, 2x2 data at all five CO-RADS thresholds were obtained; two studies only reported 2x2 data at a CO-RADS threshold of 3, and the authors could not provide additional data. Table 3.5 summarizes the pooled sensitivity and specificity at each CO-RADS threshold.

Five studies that evaluated CT used the RSNA scoring system (601 (52%) cases amongst 1162 participants) to define index test positivity. For four studies, 2x2 data at all four RSNA

thresholds were obtained; one study did not report 2x2 data at a RSNA threshold of 1 or 2, and the authors could not provide additional data. Table 3.6 summarizes the pooled sensitivity and specificity at each RSNA threshold.

The forest plots of chest CT studies that used CO-RADS and RSNA are presented by threshold in Appendix 2.4 Figures 3.A4 and 3.A5, respectively.

Indirect test comparisons in suspected individuals

Indirect comparisons of modalities evaluated across all 94 studies in suspected participants indicated that chest CT and ultrasound gave similar sensitivity estimates (P=0.42) compared to each other; chest CT (69 studies) and ultrasound (15 studies) gave higher sensitivities than X-ray (P=0.0003 and P=0.001, respectively). All modalities had similar specificities (CT versus X-ray P=0.36; CT versus ultrasound P=0.32; X-ray versus ultrasound P=0.89).

Pooled rates of positive imaging in individuals with initial RT-PCR negative results

For chest CT (7 studies, 177 participants), the rate of positive imaging in people with an initial RT-PCR negative result and RT-PCR positive result on follow-up ranged from 21% to 100%. The pooled rate was 75.8% (95%CI 45.3-92.2). For ultrasound of the lungs (one study, 21 participants), the pooled rate was 90.4%. The forest plot of chest CT studies showing the rates of positive imaging per study is presented in Figure 3.A6 in Appendix 3.4.

Pooled estimates in asymptomatic individuals

For chest CT (7 studies, 3134 participants, 315 (10%) cases) used to screen asymptomatic individuals, the sensitivity ranged from 20.7% to 80%, and specificity ranged from 68.4% to 100%. The pooled sensitivity of chest CT was 55.7% (95%CI 35.4-74.3) and the pooled specificity was 91.1% (95%CI 82.6-95.7). For chest X-ray (one study, 85 participants, 4 cases) the sensitivity was 75.0% and the specificity was 74.0%. For ultrasound of the lungs (2 studies, 329 participants, 45 cases) the sensitivities were 50.0% and 69.7%, and specificities were 98.8% and 68.0%, respectively. The SROC of chest CT studies for asymptomatic screening are presented in Figure 3.6; the forest plot is presented in Figure 3.A7 in Appendix 3.4.

Changes across review versions

Figure 3.7 displays the pooled sensitivity and specificity estimates with 95% CIs from all four versions of this review (i.e. the initial review published in September 2020 (Salameh 2020a), the second version published in November 2020 (Islam 2020), the third version published in March 2021 (Islam 2021), and this current version). The sensitivity estimates of chest CT appear to be similar across all versions, while the specificity estimates of chest CT appear to increase from version 1 to 3, and then remain similar between version 3 and current version.

For chest X-ray, which was evaluated only in the third version (Islam 2021) and the current version, the specificities appear to be similar, while the sensitivity appears to slightly increase in the current version. For ultrasound of the lungs, which was evaluated only in the third version (Islam 2021) and current version, the sensitivities appear to be similar, while the specificity appears to increase in the current version.

3.5 Discussion

This is the fourth version of a Cochrane living systematic review evaluating the diagnostic accuracy of thoracic imaging in people with suspected COVID-19.

Summary of main results

In the current version of the review, chest CT (69 studies) demonstrated a sensitivity of 86.9% (95%CI 83.6 to 89.6), and a specificity of 78.3% (95%CI 73.7 to 82.3) for the diagnosis of COVID-19 in suspected participants. Compared with the findings of the previous review version (Islam 2021), the current version demonstrates similar sensitivity and specificity of chest CT. The refined inclusion criteria used in this review mean that while the sensitivity and specificity are not vastly different, the confidence in these results has improved since the prior version.

There was no statistical evidence indicating an effect of reference standard conduct on the chest CT accuracy; 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 the previous review versions (Salameh 2020a, Islam 2020 and Islam 2021).

The definition used for index test positivity in chest CT studies appeared to impact sensitivity but not specificity, as studies that used radiologists' impressions showed higher sensitivities than those that used formal scoring systems. A possible explanation is that a 'threshold effect' seems to apply to the different definitions for index test positivity, with radiologists’ impression offering higher sensitivity and lower specificity, and formal scoring

systems offering lower sensitivity and higher specificity (when the threshold with the highest YI is considered).

Chest X-ray (17 studies) demonstrated a sensitivity of 73.1% (95%CI 64.1 to 80.5), and a specificity of 73.3% (95%CI 61.9 to 82.2) for diagnosing COVID-19 in suspected participants. Compared to the previous review version (Islam 2021), the specificities appear to be similar, while the sensitivity appears to slightly increase in the current version. Ultrasound of the lungs (15 studies) demonstrated a sensitivity of 88.9% (95% CI 84.9 to 92.0), and a specificity of 72.2% (95% CI 58.8 to 82.5). Compared to the previous review version (Islam 2021), the sensitivities appear to be similar, while the specificity appears to increase in the current version.

Threshold effects (CO-RADS and RSNA)

In chest CT studies that used the CO-RADS scoring system to define index test positivity (11 studies), when the threshold for index test positivity increased (i.e. from 2 to 5), sensitivity decreased and specificity increased, as expected. The same pattern was seen for studies that used the RSNA scoring system (5 studies) when the threshold for index test positivity increased (i.e. from 2 to 4). These results align with those in the previous review version (Islam 2021).

Indirect test comparisons

Based on indirect comparisons of all included studies, both chest CT and ultrasound gave higher sensitivity estimates than X-ray. Chest CT and ultrasound gave similar sensitivities. All modalities had similar specificities.

Rate of positive imaging in individuals with initial RT-PCR negative results

The pooled rate of positive chest CT imaging (7 studies) in repeat RT-PCR positive results where initial RT-PCR was negative, was 75.8% (95%CI 45.3 to 92.2). Pooled rates for X-ray and ultrasound could not be derived due to insufficient available data.

Asymptomatic screening

Chest CT (8 studies) demonstrated a sensitivity of 55.7% (95%CI 35.4 to 74.3), and a specificity of 91.1% (95%CI 82.6 to 95.7) for detecting COVID-19 in asymptomatic participants. These findings suggest that imaging is not useful for screening asymptomatic patients. Pooled accuracy estimates for screening with X-ray and ultrasound could not be derived due to insufficient available data.

Changes across review versions

Based on the visual assessments of the ggplot graphs, chest CT sensitivity appears to remain similar across all versions of the review (Salameh 2020a, Islam 2020, Islam 2021, and this current version). Chest CT specificity appears to increase from the first to the second version, and from the second to the third version, and then remains similar between the third and current versions. Given the large number of chest CT studies included in the prior review, which provided sensitivity and specificity estimates with narrow confidence intervals, the review authors had expected that sensitivity and specificity estimates of chest CT would not notably differ in future updates of this review. The results of the current review align with this expectation. For chest X-ray, the specificities between the third (Islam 2021) and current versions appear to be similar, while the sensitivity appears to have slightly increased in the current version. For ultrasound of the lungs, the sensitivities between the third (Islam 2021) and

current versions appear to be similar, while the specificity appears to have increased in the current version.

Strengths and weaknesses of the review

The search strategy was broad and allowed for identification of a wide range of articles about COVID-19 diagnosis. The eligibility criteria was inclusive, and although screening and exclusion of articles that had been retracted or published in predatory journals were not conducted in this review, risk of bias assessment was performed for each included study to judge the quality of the study design and methods. The review authors screened records, extracted data, and assessed study methodology independently and in duplicate. Furthermore, compared to the previous versions of the review (Salameh 2020a; Islam 2020; Islam 2021) this current update includes a greater number of studies that evaluated accuracy of imaging tests in the diagnosis of suspected COVID-19 participants.

For the primary objective on accuracy of imaging for diagnosing suspected participants, studies with only symptomatic participants, and studies with mixed populations were included. 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 included studies clearly reported information on participants' symptoms.

Reference standard conduct was not identified as a source of variability for chest CT accuracy. Index test positivity was found to impact chest CT sensitivity but did not impact CT specificity, or the accuracy of any other modality. These findings may suggest that the factors investigated did not notably contribute to variability; alternatively, there may be unmeasured confounding factors obscuring the analyses. Due to insufficient granularity of data, it was not

possible to investigate additional potential sources of variability, particularly participant setting (inpatient versus outpatient). These analyses will be performed in future versions of the review when sufficient data become available.

This was the first version that addressed the additional objectives of evaluating the rate of positive imaging in repeat RT-PCR positive results where initial RT-PCR was negative, and evaluating the accuracy of imaging in asymptomatic individuals.

Indirect comparisons of chest CT, chest X-ray and ultrasound of the lungs were formally assessed. Direct comparisons could not be evaluated due to the limited number of studies that evaluated multiple imaging modalities in the same population. Formal analyses of direct comparisons of imaging tests are planned for future versions of the review, as more studies with comparative designs become available.

It was not possible 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, this will be considered this in future versions of the review.

It was not possible to evaluate the planned additional secondary objective on possible associations between findings on thoracic imaging for patients with COVID-19, the number of days after symptom onset, symptom severity and subsequent hospitalization (McInnes 2020) due to insufficient data. Following discussion with clinicians and systematic review methods experts, the review authors have planned to address this objective in a separate systematic review within the Cochrane suite of 'living' systematic reviews, which will evaluate the prognostic value of thoracic imaging for patients with COVID-19 (Chapter 4).

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. In this version, half of all studies seemed to have low risk of bias data, while, in the previous version (Islam 2021), the majority were high or unclear.

Of the studies that did report recruitment methods, most reported including 'consecutive' participants. However, 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 selection bias.

In all included studies, RT-PCR was used alone or combination with other criteria to compose 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. As such, the accuracy estimates reported in this review should be interpreted with caution. However, the investigations of heterogeneity for chest CT studies in this review did not identify differing accuracy estimates between studies that used at least two RT-PCR test results to define disease status versus studies that did not have repeat RT-PCR testing. At this stage, despite its limitations, RT-PCR remains the best tool for diagnosing COVID-19. However, the best reference standard may vary across clinical questions, settings, and populations (Korevaar 2020).

Three out of 98 included studies (3%) were only available as preprints at the time of the search. In future versions of this review, data extracted from these studies will be updated as they become published in peer-reviewed journals.

Applicability of findings to the review question

The review findings are applicable to individuals suspected to have COVID-19. The search did not identify many studies in pediatric populations and most participants evaluated across all included studies were adults. The results of this review are applicable to adults, as the diagnostic accuracy of these modalities in children is not as well-established. It should be noted that the results apply mostly to imaging interpreted by radiologists. Considering country of study origin as a proxy for participant nationality and race, the results of this review are likely most generalizable to individuals in Europe who are Caucasian. In addition, the lack of data available in the included studies pertaining to signs and symptoms of presenting cases, severity of symptoms, and timing of symptom onset adds complexity to the interpretation of the findings in this review.

Conclusion

Chest CT and ultrasound of the lungs are sensitive and moderately specific in diagnosing COVID-19. Chest X-ray is moderately sensitive and moderately specific in diagnosing COVID-19. Thus, chest CT and ultrasound may have more utility for ruling out COVID-19 than for differentiating SARS-CoV-2 infection from other causes of respiratory illness. The uncertainty resulting from risk of bias and heterogeneity of included studies limit the confidence in these results.

Figures

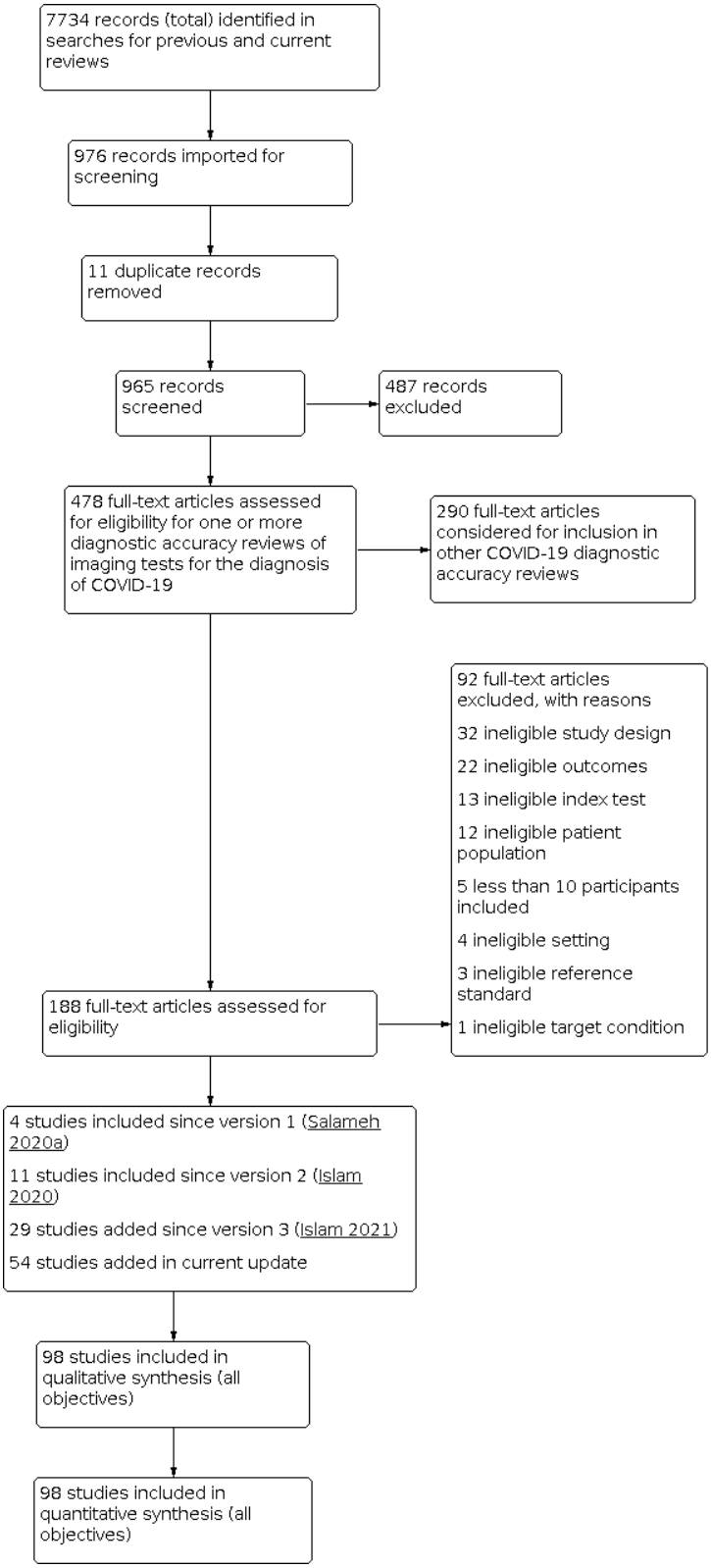


Figure 3.1. Study flow diagram

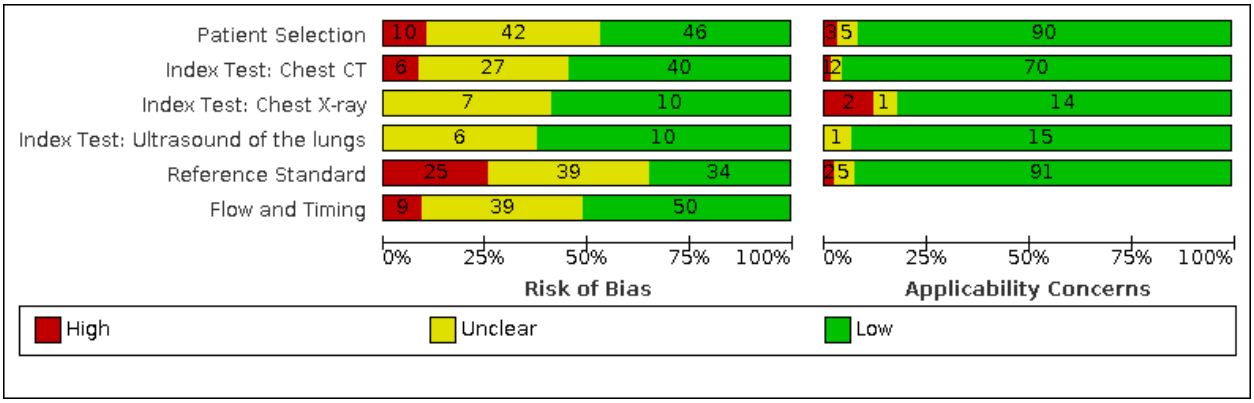


Figure 3.2. Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies (n=98).

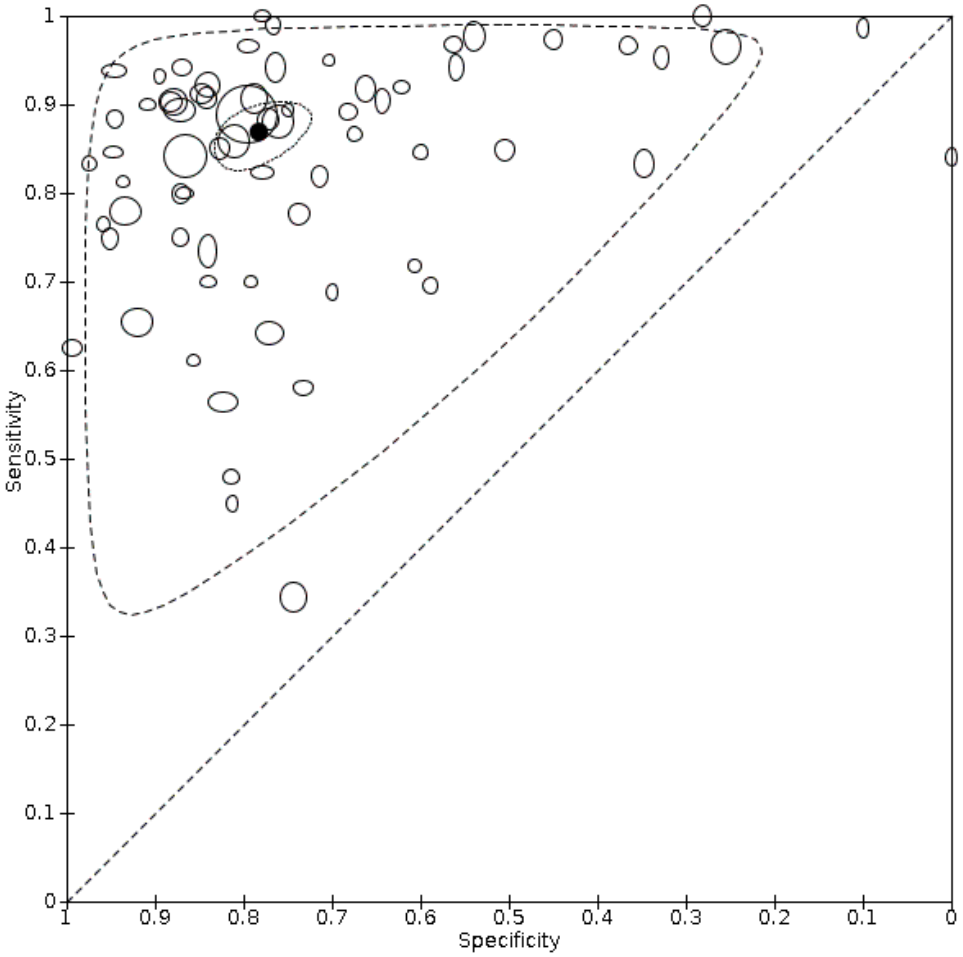


Figure 3.3. Summary ROC plot of chest CT in suspected cases. The summary point is indicated by the solid black circle, individual studies are indicated by outlined circles (scale=study sample size). The dotted border and the dashed border represent 95% confidence regions and 95% prediction regions, respectively.

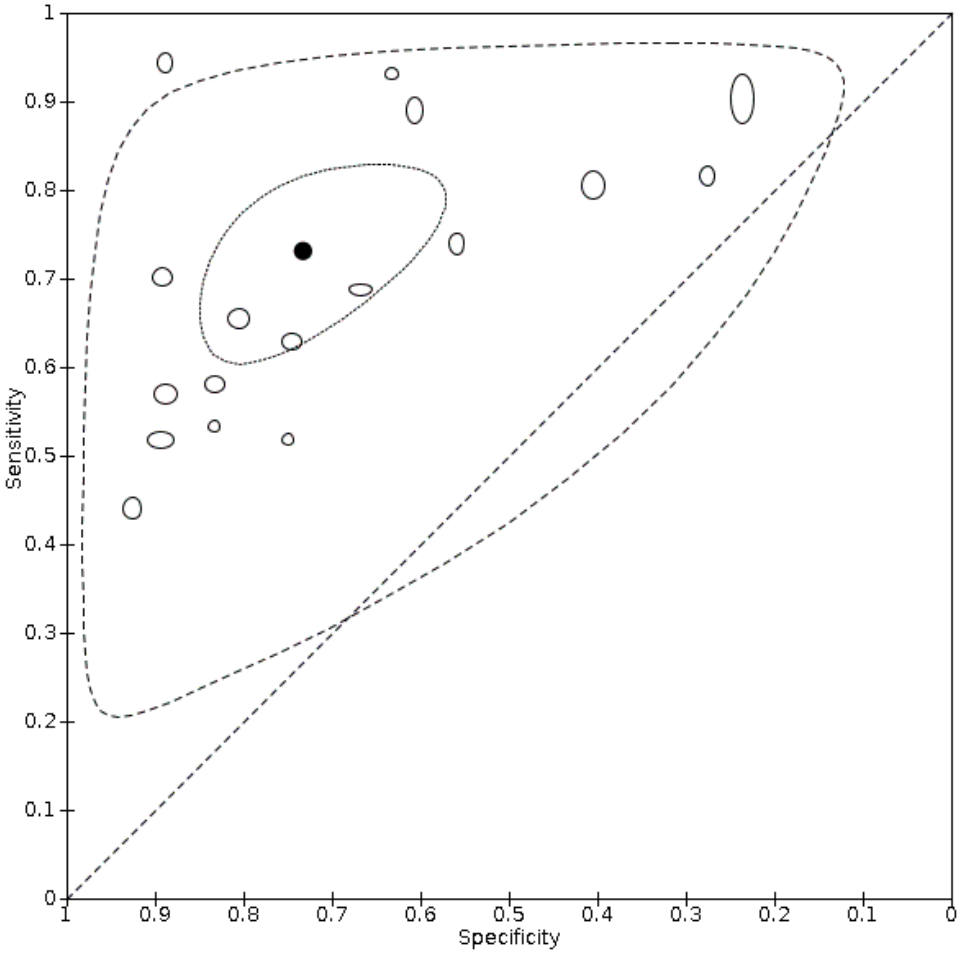


Figure 3.4. Summary ROC plot of chest X-ray in suspected cases. The summary point is indicated by the solid black circle, individual studies are indicated by outlined circles (scale=study sample size). The dotted border and the dashed border represent 95% confidence regions and 95% prediction regions, respectively.

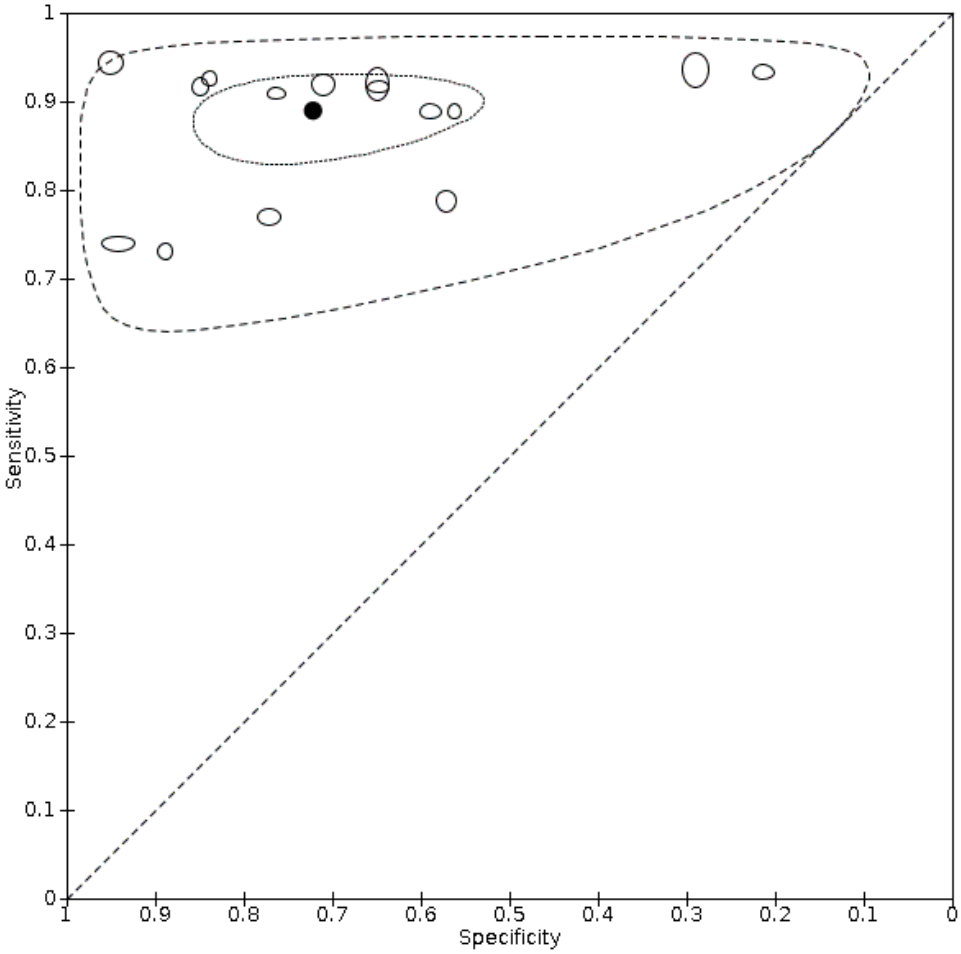


Figure 3.5. Summary ROC plot of ultrasound of the lungs in suspected cases. The summary point is indicated by the solid black circle, individual studies are indicated by outlined circles (scale=study sample size). The dotted border and the dashed border represent 95% confidence regions and 95% prediction regions, respectively.

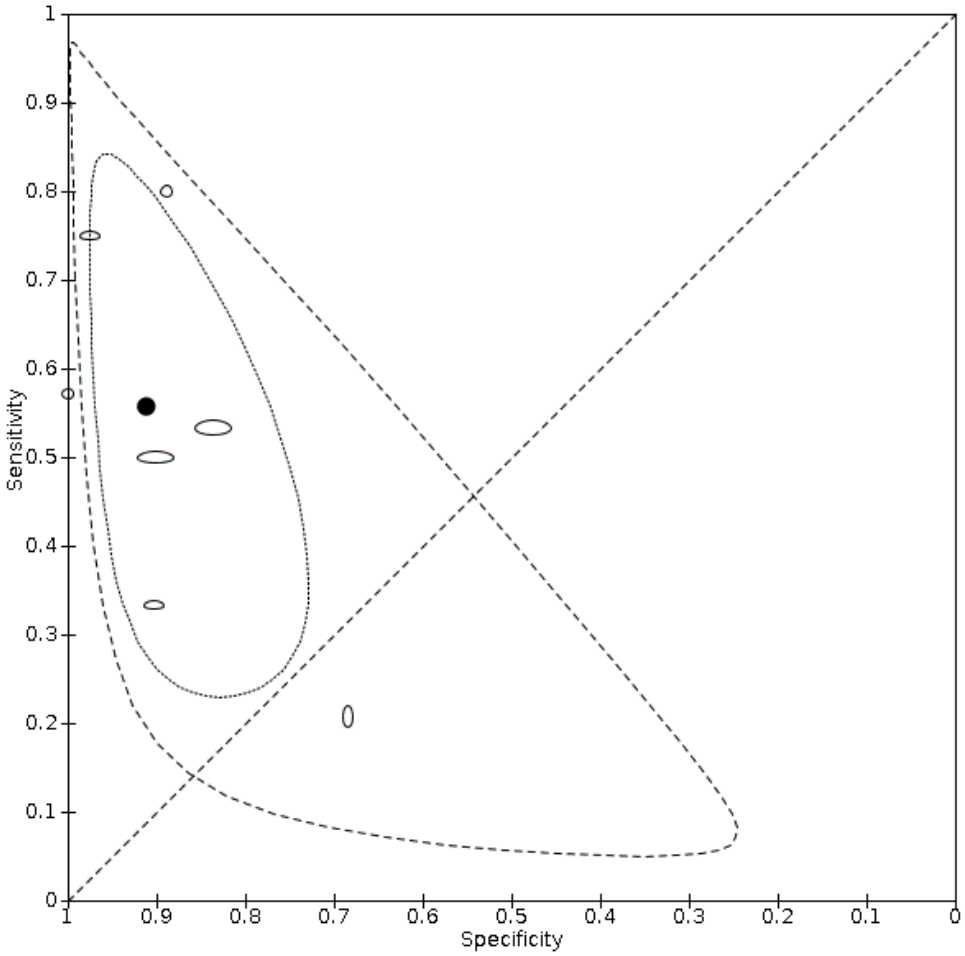


Figure 3.6. Summary ROC plot of chest CT in asymptomatic cases. The summary point is indicated by the solid black circle, individual studies are indicated by outlined circles (scale=study sample size). The dotted border and the dashed border represent 95% confidence regions and 95% prediction regions, respectively.

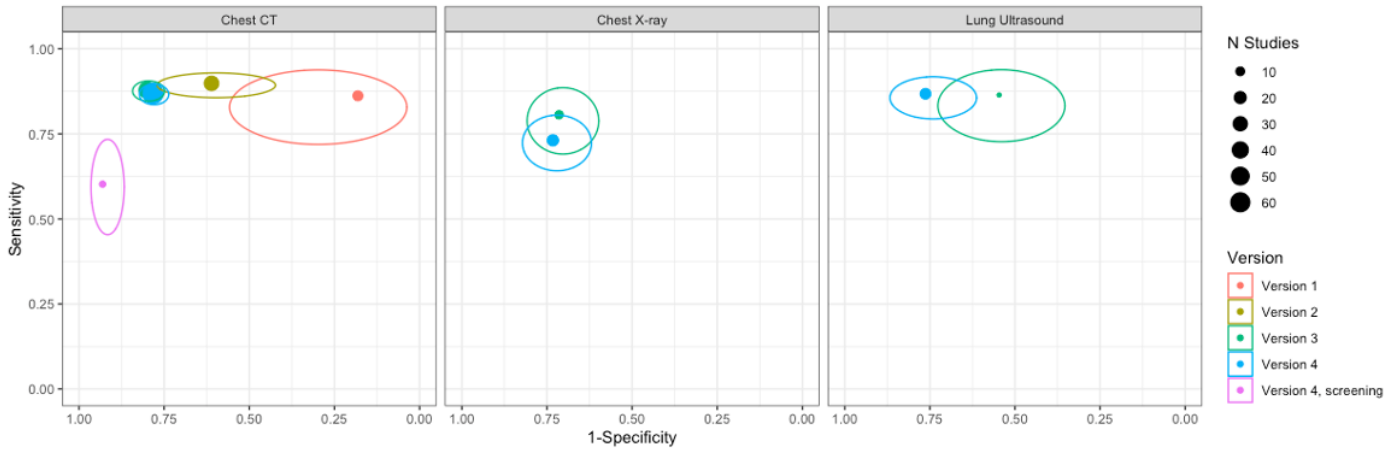


Figure 3.7. Pooled sensitivity and specificity estimate and 95% confidence intervals across all review versions (Salameh 2020a (Version 1); Islam 2020 (Version 2); Islam 2021 (Version 3); and this review update version (Version 4)) for chest CT, chest X-ray and ultrasound of the lungs.

Tables

Table 3.1. Summary of included studies for diagnostic accuracy in suspected participants

Study ID	Country of corresponding author	Study design	Age group	Setting	Index test(s)	Definition for index test positivity	Level of training of readers	Reference standard	Prevalence
Ai 2020a	China	Suspected patients (unclear)	Adults only	Inpatient	Chest CT	Unclear	Radiologist	RT-PCR, no other details provided	0.6
Aslan 2020	Turkey	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast, low dose)	Pneumonia appeared to be radiologist's impression	Radiologist	RT-PCR twice, in all with initial negative results	0.8
Bahrami-Motlagh 2020	Iran	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (low dose)	They reported negative or positive CT, according to previous reports on typical and atypical CT findings of COVID-19 pneumonia.	Unclear	RT-PCR, no other details provided	0.5
Barbosa 2020	Brazil	Suspected patients (all symptomatic)	Adults only	Unclear	Chest CT	RSNA classification	Radiologist	RT-PCR, no other details provided	0.3
Bellini 2020	Italy	Suspected patients (all symptomatic)	Children and adults	Unclear	Chest CT (non-contrast)	CO-RADS classification	Radiologist	RT-PCR twice, in some with initial negative results	0.2
Besutti 2020	Italy	Suspected patients (all symptomatic)	Adults, perhaps also children	Outpatient	Chest CT (non-contrast)	A structured report about the probability of COVID-19 pneumonia	Radiologist	RT-PCR twice, in some with initial negative results	0.8

Bock 2021	Denmark	Suspected patients (all symptomatic)	Adults only	Outpatient	Ultrasound of the lungs (POCUS)	LUS was performed to determine the presence of the following predefined conditions: focal B-lines, interstitial syndrome, lung consolidation, pleural effusion and pneumothorax. In all 14 zones, it was noted whether lung sliding, lung pulse, lung point, multiple B-lines (≥ 3 per intercostal space), or thickened or fragmented visceral pleura were present. A normal LUS was defined as sufficient LUS investigation with none of the above-mentioned findings.	Unclear	RT-PCR, no other details provided	0.4
Bollineni 2021	Belgium	Suspected patients (all symptomatic)	Mix of children and adults	Outpatient	Chest CT (non-contrast, low dose)	Unclear	Unclear	RT-PCR twice, in all with initial negative results	0.6
Borakati 2020	UK	Suspected patients (symptomatic or asymptomatic)	Adults, perhaps also children	Outpatient	Chest CT (non-contrast, IV contrast)/ chest radiographs	BSTI classification	Radiologist	RT-PCR twice, in some with initial negative results	0.6
Bosso 2021	Italy	Suspected patients (all symptomatic)	Adults, perhaps also children	Outpatient	Ultrasound of the lungs (POCUS)	Unclear	Unclear	RT-PCR twice, in some with initial negative results	0.4
Boussouar 2020	France	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast)	The conclusion was therefore one of the following: 1) imaging patterns suggesting the presence of COVID-19; 2) imaging patterns suggesting an alternative diagnosis; 3) imaging patterns suggesting a combination of COVID-19 with underlying lung disease; 4) CT considered normal	Radiologists	RT-PCR twice, in all with initial negative results	0.5

Brun 2021	France	Suspected patients (all symptomatic)	Adults, perhaps also children	Outpatient	Chest CT (low dose)	Highly probable, probable, and less probable of COVID-19 pneumonia, alternative diagnosis, or normal. They established their diagnosis based on recent publications from China illustrating typical and atypical patterns in patients with COVID-19 pneumonia (Pan 2020; Li 2020; Ye 2020; Kanne 2020, Zhao 2020, Wang 2020a; Salehi 2020) and according to the Radiological Society of North America expert consensus statement (Zhou 2020)	Unclear	RT-PCR, no other details provided	0.6
Caruso 2020	Italy	Suspected patients(all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast)	Pneumonia	Radiologist	RT-PCR twice, in all with initial negative results	0.4
Cengel 2021	Turkey	Suspected patients (symptomatic or asymptomatic)	Adults, perhaps also children	Outpatient	Chest CT (non-contrast)	RSNA classification	Unclear	RT-PCR twice, in some with initial negative results	0.7
Colombi 2020a	Italy	Suspected patients (all symptomatic)	Adults, perhaps also children	Outpatient	Chest CT (low dose)/ultrasound of lungs	RSNA classification	Unclear	RT-PCR twice, in some with initial negative results	0.7
Cozzi 2020	Italy	Suspected patients (symptomatic or asymptomatic)	Unclear	Outpatient	Chest radiographs/ Chest X-rays	The presence of interstitial infiltrates with predominantly bilateral and basal distribution	Radiologist	RT-PCR, no other details provided	0.8
De Smet 2020	Belgium	Suspected patients (all symptomatic)	Children and adults	Inpatient	Chest CT	CO-RADS classification	Unclear	RT-PCR, no other	0.4

								details provided	
Debray 2020	France	Suspected patients (unclear)	Adults only	Inpatient	Chest CT (non-contrast)	"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	Radiologist	RT-PCR twice, in some with initial negative results	0.6
Deng 2020	China	Suspected patients (all symptomatic)	Children and adults	Inpatient	Chest CT (high resolution)	Any one of the following: a) Single, multiple, or diffuse ground-glass opacity, with thickened blood vessels and thickened bronchial shadows passing through, with or without localized lobular septal grid thickening; b) Single or multiple real shadows, (2) Reexamination 3 to 5 days later showed that the original ground-glass opacity or consolidation range increased, the number increased, or accompanied by pleural effusion on one or both sides	Radiologist	RT-PCR once	0.7
Dimeglio 2021	France	Suspected patients (all symptomatic)	Unclear	Outpatient	Chest CT	Following the recommendation of the French Society of Radiology	Unclear	RT-PCR once	0.4
Dini 2020	Italy	Suspected patients (symptomatic or asymptomatic)	70 years of age and older	Outpatient(LTC)	Ultrasound of lungs(POCUS)	Scoring system: non-coalescent B-lines, coalescent and with hyperdensed non-consolidated state.	Unclear	RT-PCR once	0.6
Djangang 2020	Belgium	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT	CT-scan was suggestive or not for COVID-19 (i.e., ground-glass opacities, consolidation or crazy-paving patterns) (Ai 2020a; Zhang 2020)	Radiologist	RT-PCR twice, in some with initial negative results	0.5
Dofferhoff 2020	The Netherlands	Suspected patients	Adults only	Inpatient	Chest CT (low dose)	CO-RADS classification; threshold not pre-specified	Unclear	RT-PCR twice, in	0.5

		(symptomatic or asymptomatic)						some with initial negative results	
Dogan 2020	Turkey	Suspected patients (symptomatic or asymptomatic)	Adults only	Outpatient	Chest CT (non-contrast)	RSNA criteria: typical, indeterminate, atypical, negative	Radiologist	RT-PCR twice, in all with initial negative results	0.5
Ducray 2020	France	Suspected patients (symptomatic or asymptomatic)	Adults only	Outpatient	Chest CT (non-contrast, IV contrast)	On the final report, patients were rated as “Surely COVID+” when presenting with peripheral, bilateral, or multifocal GGO of rounded morphology ± consolidation or crazy paving, reversed halo sign, or sub-pleural bands of consolidations. Patients were rated as “Possible COVID+” when presenting with multifocal, diffuse, peripheral, or unilateral GGO ± consolidation lacking a specific distribution and non-rounded or non-peripheral or with only few very small GGO with a non-rounded and non-peripheral distribution or with atypical findings: large pleural effusion, major lymph node size increase, or bronchiolitis pattern. Patients were rated as “COVID−” when the chest CT was normal or demonstrating another pathology	Radiologist	RT-PCR twice, in some with initial negative results	0.4
Erxleben 2021	Germany	Suspected patients (all symptomatic)	Adults, perhaps also children	Outpatient	Chest CT (low dose)	Unclear: "All CT images were evaluated manually and data on presence/absence of COVID-19 was assessed"	Unclear	RT-PCR twice, in some with initial negative results	0.1

Falaschi 2020	Italy	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast)	RSNA classification	Radiologist	RT-PCR twice, in some with initial negative results	0.6
Ferda 2020	Czech Republic	Suspected patients (all symptomatic)	Mix of children and adults	Outpatient	Chest CT(IV contrast)	Ground glass opacities, mixed ground-glass opacities, thickening of intra-lobular septa, negative bronchogram, reverse halo sign, and dilatation of the vascular structures. Predominant peripheral, bilateral and caudal distributions were suspected to be COVID-19 pneumonia.	Radiologist	RT-PCR twice, in some with initial negative results	0.1
Fink 2021	Germany	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (High resolution)/ Chest X-rays	CT scans were classified according to two different reading scores: 1) presence of pneumonic features (0 – absent, 1 – present) and 2) presence of COVID-19 typical features (0 – not typical, 1 – possible, 2 – highly suspicious). According to the current literature, COVID-19 typical features were defined as ground glass opacities (GGO) with or without “crazy paving” and/or consolidations with peripheral emphasis.	Radiologist	RT-PCR twice, in some with initial negative results	0.3
Fonsi 2020	Italy	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast)	Ground glass opacities (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	Radiologist	RT-PCR once	0.7

						enlargement ; pleural effusion; and pericardial effusion.			
Fujioka 2020	Japan	Suspected patients (all symptomatic)	Adults only	Unclear	Chest CT	CO-RADS classification	Radiologist	RT-PCR once	0.5
Gaia 2020	Italy	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT	Simpson 2020	Radiologist	RT-PCR once	0.5
Giannitto 2020	Italy	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast)	Unclear	Radiologist	RT-PCR twice, in all with initial negative results	0.3
Gietema 2020	The Netherlands	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast)	Reporting scheme	Resident	RT-PCR twice, in some with initial negative results	0.4
Gil-Rodrigo 2020	Spain	Suspected patients (all symptomatic)	Adults only	Outpatient	Ultrasound of the lungs (POCUS)	Scoring system by Soldati 2020	Unclear	RT-PCR once	0.4
Grando 2020	Brazil	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast)	CT features were classified as "typical," "indeterminate," "atypical," and "negative" for COVID-19 pneumonia", according to RSNA expert consensus	Radiologist.	RT-PCR twice, in some with initial negative results	0.5
Gross 2021	Germany	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (low dose)	CO-RADS classification	Radiologists	RT-PCR twice, in all with initial negative results	0.2
Guillo 2020	France	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast, IV contrast)	A structured report about the probability of COVID-19 pneumonia	Resident	RT-PCR twice, in some with	0.6

								initial negative results	
Haak 2021	The Netherlands	Suspected patients (all symptomatic)	Adults only	Outpatient	Ultrasound of the lungs (POCUS)	Score of >/= 2 based on (Peng 2020a; 4 Lung ultrasound in COVID-19 2020; Focus met POCUS op COVID-19 2020)	Unclear	RT-PCR twice, in all with initial negative results	0.3
Hanif 2021	Pakistan	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (high resolution)	Positive findings for COVID-19 defined as bilateral, multifocal, multilobar ground glass opacities with or without sub-segmental consolidations or crazy paving pattern in a peripheral distribution (Han 2020; Lee 2020; Simpson 2020) Negative findings defined as presence of isolated lobar consolidation, pleural effusion, nodularity and absence of the positive findings of COVID-19. Indeterminate cases defined as having multilobar ground glass opacities or consolidation with central or diffuse distribution lacking subpleural pattern or unilateral ground glass opacities; these were further categorized as positive or negative for COVID-19 on the basis of clinical history, mutual consensus and RT-PCR results, if available.	Radiologists	RT-PCR twice, in some with initial negative results	0.8
He 2020	China	Suspected patients (unclear)	Children and adults	Inpatient	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 twice, in some with initial negative results	0.4
Hermans 2020	The Netherlands	Suspected patients (symptomatic	Adults only	Outpatient	Chest CT	CO-RADS classification	Radiologist	RT-PCR once	0.4

		or asymptomatic)							
Hernigou 2020	Belgium	Suspected patients (symptomatic or asymptomatic)	Adults only	Inpatient	Chest CT (low dose)	Unclear	Radiologist	RT-PCR twice, in some with initial negative results	0.3
Herpe 2020	France	Suspected patients (all symptomatic)	Children and adults	Unclear	Chest CT	Bilateral ground glass opacities with peripheral distribution, bilateral crazy paving appearance with intralobular thickening, reverse halo sign, or other signs compatible with organizing pneumonia.	Radiologist	RT-PCR twice, in some with initial negative results	0.1
Hwang 2020	Korea	Suspected patients (symptomatic or asymptomatic)	Adults, perhaps also children	Unclear	Chest radiographs / chest X-rays	Abnormality suggesting pneumonia	Radiologists and Resident	RT-PCR, no other details provided	0.05
Ippolito 2020	Italy	Suspected patients (all symptomatic)	Children and adults	Inpatient	Chest radiographs / chest X-rays	Reticulations, alveolar opacities, or both	Radiologist	RT-PCR, no other details provided	0.4
Jalil 2020	USA	Suspected patients (all symptomatic)	Adults, perhaps also children	Outpatient	Ultrasound of the lungs (POCUS)	Unclear	Unclear	RT-PCR twice, in all with initial negative results	0.5
Krdzalic 2020	The Netherlands	Suspected patients (all symptomatic)	Adults only	Unclear	Chest CT	CO-RADS classification	Radiologist	RT-PCR twice, in some with initial negative results	0.5
Kuzan 2020	Turkey	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast)	BSTI classification	Radiologist	RT-PCR twice, in some with initial	0.6

								negative results	
Lieveld 2021a	The Netherlands	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT	CO-RADS classification	Radiologists	RT-PCR twice, in all with initial negative results	0.3
Lieveld 2021b	The Netherlands	Suspected patients (all symptomatic)	Adults only	Outpatient	Ultrasound of the lungs (POCUS)	CO-RADS classification	Unclear	RT-PCR twice, in some with initial negative results	0.4
Luo 2020a	China	Suspected patients (all symptomatic)	Children and adults	Inpatient	Chest CT	Scoring system was developed; threshold not pre-specified	Radiologist	RT-PCR twice, in all with initial negative results	0.4
Majeed 2020	UK	Suspected patients (symptomatic or asymptomatic)	Adults only	Outpatient	Chest CT	BSTI classification and RSNA classification	unclear	RT-PCR twice, in some with initial negative results	0.3
Mei 2020	USA	Suspected patients (symptomatic or asymptomatic)	Children and adults	Unclear	Chest CT	Unclear	Radiologist	RT-PCR twice, in all with initial negative results	0.5
Miranda Magalhaes Santos 2020	Brazil	Suspected patients (all symptomatic)	Children and adults	Outpatient	Chest CT	RSNA classification	Radiologist	RT-PCR, no other details provided	0.5
Moroni 2021	Italy	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest radiographs / Chest X-rays	Unclear	Unclear	RT-PCR, no other details provided	0.3

Murphy 2020	The Netherlands	Suspected patients (all symptomatic)	Children and adults	Outpatient	Chest radiographs / Chest X-rays	Readers assigned each image a category, sensitivities matched to AI reading	Radiologist	RT-PCR, no other details provided	0.5
Narinx 2020	Belgium	Suspected patients (all symptomatic)	Adults, perhaps also children	Outpatient	Chest CT (low dose, with or without contrast)/ultrasound of lungs (POCUS)	For Ultrasound: POCUS lung positive if one or more BLUE points showed a positive B-line parameter. For chest CT: Scored as suggestive for or inconsistent with COVID-19 infection based on the presence of clinical manifestations as presented by Ng 2020 and Shi 2020	Radiologist	RT-PCR, no other details provided	0.2
Nivet 2021	France	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast)	Each reading was categorized using a five-point score, adapted from the recommendations of the Société Française de Radiologie (SFR). (1) normal; (2) non-infectious findings; (3) infectious findings but not consistent with COVID-19 infection; (4) consistent with COVID-19 infection; (5) typical appearance of COVID-19 infection.	Residents and radiologist	RT-PCR twice, in some with initial negative results	0.4
Ohana 2021	France	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast)	CT with typical COVID-19 appearance, i.e. bilateral and predominantly peripheral and sub-pleural ground glass opacities and/or alveolar consolidations, were classified as positive AB65	Radiologists	RT-PCR twice, in some with initial negative results	0.5
O'Neill 2020	Canada	Suspected patients (symptomatic or asymptomatic)	Adults only	Outpatient	Chest CT	RSNA classification and CO-RADS classification	Radiologists	RT-PCR twice, in all with initial negative results	0.7
Pagano 2021	USA	Suspected patients (symptomatic	Adults only	Outpatient	Chest radiographs/chest X-rays	Unclear	Unclear	RT-PCR, no other	0.8

		or asymptomatic)						details provided	
Palmisano 2021	Italy	Suspected patients (all symptomatic)	Adults, perhaps also children	Outpatient	Chest CT (non-contrast)	RSNA classification	Unclear	RT-PCR twice, in some with initial negative results	0.6
Pare 2020	USA	Suspected patients (all symptomatic)	Adults, perhaps also children	Outpatient	Chest radiographs /chest X-rays/Ultrasound of lungs (POCUS)	Classified CXRs as positive if the report included infection in the differential.	Unclear	RT-PCR twice, in some with initial negative results	0.8
Patel 2020	USA	suspected patients (symptomatic or asymptomatic)	Adults, perhaps also children	Outpatient	Chest CT (high resolution)	Category 1 – consistent with multifocal pneumonia; Category 2 – indeterminate for multifocal pneumonia; Category 3 – not consistent with multifocal pneumonia	Unclear	RT-PCR, no other details provided	0.5
Patrucco 2021	Italy	Suspected patients (symptomatic or asymptomatic)	Adults, perhaps also children	Outpatient	Chest CT	RSNA classification and CO-RADS classification	Unclear	RT-PCR, no other details provided	0.4
Peng 2020a	China	Suspected patients (symptomatic or asymptomatic)	Children only	Inpatient	Chest CT	Ground glass opacity, consolidations with surrounding halo sign, nodules, residual fibre strips, lymphadenopathy	Radiologist	RT-PCR, no other details provided; other (positive contacts)	0.5
Pivetta 2021	Italy	Suspected patients (all symptomatic)	Adults only	Outpatient	Ultrasound of the lungs (POCUS)	Unclear	Unclear	RT-PCR twice, in some with initial negative results	0.4

Ravikanth 2021	India	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (CT thorax with IV contrast)	Dichotomous - suspicious or not suspicious for COVID-19.	Resident and radiologist	RT-PCR twice, in some with initial negative results	0.8
Reginelli 2021	Italy	Suspected patients (symptomatic or asymptomatic)	Adults only	Outpatient	Chest CT	Radiologists observed according to localization and distribution of GGO and consolidations, crazy paving pattern, and presence of nodules	Radiologist	RT-PCR twice, in some with initial negative results	0.8
Rona 2021	Turkey	Suspected patients (all symptomatic)	Children and young adults only	Outpatient	Chest CT (non-contrast)	CT images were divided into 3 groups: normal, consistent with COVID-19, and inconsistent with COVID-19. Multifocal consolidation, ground-glass opacity, and reversed halo sign on CT were considered to be consistent with COVID-19.	Unclear	RT-PCR twice, in some with initial negative results	0.4
Roy Choudhury 2020	India	Suspected patients (all symptomatic)	Unclear	Inpatient	Chest radiographs/chest X-rays	Simpson 2020	Unclear	RT-PCR, no other details provided	0.3
Saeed 2020	United Arab Emirates	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (high resolution)	RSNA classification	radiologists	RT-PCR twice, in all with initial negative results	0.7
Salehi-Pourmehr 2020	Iran	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT	Unclear	Unclear	RT-PCR, no other details provided	0.3
Schalekamp 2020	The Netherlands	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast)	CO-RADS classification	radiologists	RT-PCR twice, in some with initial negative results	0.5

Schmid 2020	Germany	Suspected patients (all symptomatic)	Adults only	Inpatient	Ultrasound of the lungs (POCUS)	Unclear	unclear	RT-PCR, no other details provided	0.3
Schulze-hagen 2020	Germany	Suspected patients (all symptomatic)	Adults only	Unclear	Chest CT (non-contrast, Low dose)	COV-Rads classification	Radiologist	RT-PCR twice, in some with initial negative results	0.4
Shah 2021	India	Suspected patients (all symptomatic)	Not Reported	Outpatient	Chest CT (high resolution)	Evaluated for ground-glass opacities (GGOs), reticular thickening, focal consolidations, fibrosis, pleural effusion, nodules, and hilar lymphadenopathy	Radiologists	RT-PCR twice, in some with initial negative results	0.9
Skalidis 2020	Switzerland	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (low dose)	Each specialist classified the abnormal CT according to GGO distribution of the affected lung parenchyma graded on a 3-point scale: 1=light <30%, 2=moderate 30–60%, 3=severe >60%. Finally, the results of the classification were merged by consensus and the specialists classified the CT on positive or negative for COVID-19.	Unclear	RT-PCR twice, in some with initial negative results	0.4
Song 2020a	China	Suspected patients (all symptomatic)	Adults only	Inpatient	Chest CT	Viral pneumonia according to: multiple bilateral, ill-defined ground glass opacities (GGOs) or mixed consolidation with diffuse peripheral distribution or bilateral pulmonary consolidation	Radiologist	RT-PCR twice, in all with initial negative results	0.5
Sorlini 2021	Italy	Suspected patients (all symptomatic)	Adults, perhaps also children	Outpatient	Chest X-rays/ Ultrasound of the lungs (POCUS)	Interstitial lung syndrome: two or more positive regions bilaterally with irregular pleural line. • Interstitial lung pattern: two or more positive regions with irregular pleural line, with focal/unilateral distribution.	Unclear	RT-PCR twice, in some with initial negative results	0.7

						• White lung (coalescent B lines) in two or more zones. • Subpleural consolidations.			
Speidel 2021	Switzerland	Suspected patients (all symptomatic)	Adults only	Inpatient	Ultrasound of the lungs (POCUS)	Unclear	Unclear	RT-PCR, no other details provided	0.2
Steuwe 2020	Germany	Suspected patients (all symptomatic)	Adults only	Unclear	Chest CT (Non-contrast, Low dose)	Unclear	Unclear	RT-PCR twice, in some with initial negative results	0.2
Stevens 2020	UK	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest radiographs/ Chest X-rays	BSTI classification	Radiographer and Radiologist	RT-PCR twice, in some with initial negative results	0.8
Sukhija 2021	India	Suspected patients (all symptomatic)	Adults only	Unclear	Chest X-rays	Unclear	Unclear	RT-PCR, no other details provided	0.6
Sverzellati Nicola 2021	Italy	Suspected patients (all symptomatic)	Adults only	Inpatient	Chest CT (High resolution)/ Chest X-rays	4 CT categories: normal, alternative diagnosis, indeterminate, or typical for COVID-19 pneumonia. Visual analysis: extent of combined GGO and consolidation was visually scored at the nearest 5% on the whole lungs. Distribution of findings, bilateral or unilateral involvement also considered in scoring.	Radiologist	RT-PCR twice, in all with initial negative results	0.7
Teichgraber 2021	Germany	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (Low dose)	RSNA classification	Unclear	RT-PCR twice, in all with initial negative results	0.1

Tsakok 2020	UK	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest X-rays	Unclear	Unclear	RT-PCR, no other details provided	0.4
Wang 2020a	China	Suspected patients (symptomatic or asymptomatic)	Children and adults	Unclear	Chest CT	Standardized imaging reporting system	Unclear	RT-PCR twice, in all with initial negative results	0.1
Wehbe 2021	USA	Suspected patients (all symptomatic)	Adults only	Mixed	Chest X-rays	Point scoring system based on overall impression of "positive for COVID-19" or "negative for COVID-19"	radiologist	RT-PCR twice, in some with initial negative results	0.4
Xiaocheng 2020	China	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT	Unclear	Unclear	RT-PCR, no other details provided	0.1
Xiong 2020	China	Suspected patients (unclear)	Children and adults	Inpatient	Chest CT	Subpleural ground glass opacity without pleural effusion, bronchial changes, or lymphadenopathy	Radiologist	RT-PCR, no other details provided	0.4
Yassa 2020	Turkey	Suspected patients (symptomatic or asymptomatic)	Adults only	Inpatient	Ultrasound of the lungs (POCUS)	4 categories: characteristic changes, ordinary inflammation, other changes, normal	Unclear	RT-PCR twice, in some with initial negative results	0.08
Yates 2021	Ireland	Suspected patients (all symptomatic)	Adults, perhaps also children	Outpatient	Chest X-rays	Unclear	Unclear	RT-PCR twice, in all with initial negative results	0.2

CO-RADS: COVID-19 Reporting and Data System; **CT:** computed tomography; **RSNA:** Radiological Society of North America; **RT-PCR:** reverse transcriptase polymerase chain

Table 3.2. Characteristics of the included studies for rate of positive imaging in repeat RT-PCR positive results

Study ID	Country of corresponding author	Study design	Age group	Setting	Index test(s)	Definition for index test positivity	Level of training of readers	Reference standard	Prevalence
Besutti 2020	Italy	Suspected patients (all symptomatic)	Adults, perhaps also children	Outpatient	Chest CT (non-contrast)	A structured report about the probability of COVID-19 pneumonia	Radiologist	RT-PCR once; twice in some	0.8
Bollineni 2021	Belgium	Suspected patients (all symptomatic)	Mix of children and adults	Outpatient	Chest CT (non-contrast, Low dose)	Unclear	Unclear	RT-PCR twice, in all with initial negative results	0.6
Debray 2020	France	Suspected patients (unclear)	Adults only	Inpatient	Chest CT (non-contrast)	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	Radiologist	RT-PCR once; twice in some	0.7
Giannitto 2020	Italy	Suspected patients (all symptomatic)	Adults only	Outpatient	Chest CT (non-contrast)	Unclear	Radiologist	RT-PCR twice, if necessary	0.3
Herpe 2020	France	Suspected patients (all symptomatic)	Children and adults	Unclear	Chest CT	Bilateral ground glass opacities 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	0.1
Pivetta 2021	Italy	Suspected patients (all symptomatic)	Adults only	Outpatient	Ultrasound of the lungs (POCUS)	Presence of focal or diffuse interstitial syndrome associated with spared areas, subpleural consolidations, and irregular or thickened pleural line was considered suggestive of SARS-CoV2–related pneumonia	Unclear	RT-PCR twice, in some with initial negative results	0.4

Reginelli 2021	Italy	Suspected patients (symptomatic or asymptomatic)	Adults only	Outpatient	Chest CT	Radiologists observed according to localization and distribution of GGO and consolidations, crazy paving pattern, and presence of nodules	Unclear	RT-PCR twice, in some with initial negative results	0.8
Song 2020a	China	Suspected patients (all symptomatic)	Adults only	Inpatient	Chest CT	Viral pneumonia according to: multiple bilateral, ill-defined ground glass opacities (GGOs) or mixed consolidation with diffuse peripheral distribution or bilateral pulmonary consolidation	Radiologist	RT-PCR twice, if necessary	0.5

Abbreviations: **CT:** computed tomography; **GGO:** ground glass opacity; **POCUS:** Point-of-Care Ultrasound; **RSNA:** Radiological Society of North America; **RT-PCR:** reverse transcriptase polymerase chain reaction

Table 3.3. Characteristics of the included studies summarized for asymptomatic studies

Study ID	Country of corresponding author	Study design	Age group	Reason for screening asymptomatic patients	Setting	Index test(s)	Definition for index test positivity	Level of training of readers	Reference standard	Prevalence
Dafydd 2021	UK	Asymptomatic participants	Adults only	Asymptomatic patients referred for elective oncological surgery underwent chest CT within 2 days of surgery in high-risk surgical cases.	Inpatient	Chest CT (High resolution)	Unclear	Radiologist	RT-PCR twice, in some with initial negative results	0.02
De Smet 2020	Belgium	Asymptomatic participants	Children and adults	Asymptomatic patients admitted for COVID-19-unrelated urgent medical needs were screened by chest CT	Inpatient	Chest CT	CO-RADS classification	Unclear	RT-PCR, no other details provided	0.05
Dogan 2020	Turkey	Asymptomatic participants	Adults only	Asymptomatic individuals who were suspected to have COVID-19 based on suspected contact underwent CT chest.	Unclear	Chest CT (non-contrast CT thorax)	RSNA classification	Radiologist	RT-PCR twice, in all with initial negative results	0.3
Dini 2020	Italy	Asymptomatic participants	70 years of age and older	Asymptomatic patients institutionalized in residential age care facilities who were exposed to the infection underwent chest imaging.	Outpatient (LTC)	Ultrasound of lungs (POCUS)	Scoring system: non-coalescent B-lines, coalescent and with hyperdensed non-consolidated state.	Unclear	RT-PCR once	0.6

Gumus 2020	Turkey	Asymptomatic participants	Adults only	Asymptomatic patients scheduled for any surgery were eligible for preoperative chest CT	Inpatient	Chest CT (non-contrast)	RSNA classification	Unclear	RT-PCR twice, in some with initial negative results	0.01
Hernigou 2020	Belgium	Asymptomatic participants	Adults only	Asymptomatic patients institutionalized in residential age care facilities who were exposed to the infection underwent chest imaging	Inpatient	Chest CT (low dose)	Unclear	Radiologist	RT-PCR twice, in some with initial negative results	0.3
Hwang 2020	Korea	Asymptomatic participants	Adults, perhaps also children	Unclear	Unclear	Chest radiographs/Chest X-rays	Abnormality suggesting pneumonia	Radiologists and Resident	RT-PCR, no other details provided	0.05
Ooi 2021	UK	Asymptomatic participants	Adults, perhaps also children	Asymptomatic patients scheduled for elective surgery were eligible for preoperative chest CT.	Outpatient	Chest CT	Each area was given a score between 0 and 3	Unclear	RT-PCR twice, in some with initial negative results	0.1
Puylaert 2020	The Netherlands	Asymptomatic participants	Adults only	Asymptomatic patients scheduled for an elective or emergency surgery or interventional procedure under general anaesthesia were eligible for preoperative chest CT	Inpatient	Chest CT (low dose)	CO-RADS classification	Unclear	RT-PCR once	0.01

Yassa 2020	Turkey	Asymptomatic participants	Adults only	Asymptomatic pregnant women admitted to the hospital underwent radiologic imaging	Inpatient	Ultrasound of the lungs (POCUS)	Unclear	Unclear	RT-PCR twice, in some with initial negative results	0.04

Abbreviations: **CO-RADS:** COVID-19 Reporting and Data System; **CT:** computed tomography; **LTC:** long-term care; **POCUS:** point-of-care Ultrasound; **RSNA:** Radiological Society of North America; **RT-PCR:** reverse transcriptase polymerase chain reaction.

Table 3.4. Meta-regression analyses for chest CT, X-ray, and US of suspected cases

Test, analysis group	Studies (n)	Number of participants (cases)	Sensitivity (95% CI)	Specificity (95% CI)
Reference standard conduct (chest CT)				
RT-PCR testing at least twice for all initial negative results	17	5515 (2665)	88.4% (95% CI 79.4 to 93.8)	72.7% (95% CI 62.0 to 81.3)
RT-PCR testing not done twice for all initial negatives	39	19102 (9909)	86.9% (95% CI 82.9 to 90.2)	81.2% (95% CI 75.8 to 85.6)
P value			0.71	0.13
Definition for index test positivity (chest CT)				
Radiologist impression	27	14266 (7307)	90.4% (95% CI 84.9 to 94.0)	72.4% (95% CI 62.8 to 80.3)
Formal scoring system	42	14019 (7035)	84.3% (95% CI 80.3 to 87.5)	81.5% (95% CI 76.8 to 85.4)
P value			0.037	0.070
Definition for index test positivity (chest X-ray)				
Radiologist impression	6	4489 (3246)	76.2% (62.5 to 85.9)	64.5% (44.0 to 80.8)
Formal scoring system	11	4040 (2057)	71.8% (59.7 to 81.4)	77.7% (65.0 to 86.7)
P value			0.60	0.24
Definition for index test positivity (chest US)				
Radiologist impression	9	1704 (974)	88.6% (95% CI 77.9 to 94.4)	73.8% (95% CI 49.0 to 89.1)
Formal scoring system	6	706 (208)	80.7% (95% CI 74.3 to 85.9)	79.9% (95% CI 64.8 to 89.6)
P value			0.12	0.62

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

Table 3.5. 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 (range)**	**Sensitivity (95% CI)**	**Specificity (range)**	**Specificity (95% CI)**
5	9	4169 (1672)	42% to 80%	67.3% (95% CI 57.9-75.6)	84% to 99%	92.2% (95% CI 89.3-94.3)
4	9	4169 (1672)	56% to 90%	83.3% (95% CI 76.1-88.7)	68% to 91%	84.0% (95% CI 81.3-86.4)
3	11	4416 (1769)	65% to 95%	90.3% (95% CI 85.9-93.5)	54 % to 87%	69.7% (95% CI 64.3-74.6)
2	9	4169 (1672)	75% to 100%	94.0% (95% CI 89.8-96.6)	11% to 57%	45.4% (95% CI 38.4-52.5)
1[a]	-	-		-		-

Abbreviations: **CI:** confidence interval; **CT:** computed tomography.
[a]Meta-analysis was not performed for a CO-RADS threshold of 1 since at this threshold all sensitivity values are equal to one, and all specificity values are equal to zero.

Table 3.6. Analyses of 'threshold' effects for chest CT studies of suspected cases that used the RSNA Reporting and Data System

RSNA threshold	Studies (n)	Number of participants (cases)	Sensitivity (range)	Sensitivity (95% CI)	Specificity (range)	Specificity (95% CI)
4	5	1162 (601)	34% to 88%	68.9% (47.1-84.7)	74% to 97%	90.1% (79.4-94.4)
3	5	1162 (601)	50% to 97%	87.6% (69.4-95.7)	57% to 80%	63.4% (57.1-69.2)
2	4	1071 (576)	55% to 100%	91.6% (67.1-98.3)	10.7% to 43.6%	27.9% (17.0-42.1)
1[a]	-	-		-		-

Abbreviations: **CI:** confidence interval; **CT:** computed tomography.
[a]Meta-analysis was not performed for a RSNA threshold of 1 since at this threshold all sensitivity values are equal to one, and all specificity values are equal to zero.

References

References cited in the text

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 2021).

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]

Han 2020
Han R, Huang L, Jiang H, Dong J, Peng H, Zhang D. Early clinical and CT manifestations of coronavirus disease 2019 (COVID-19) pneumonia. *American Journal of Roentgenology* 2020 Aug;215(2):338-43. [DOI: 10.2214/AJR.20.22961]

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

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]

Islam 2021
Islam N, Ebrahimzadeh S, Salameh J-P, Kazi S, Fabiano N, Treanor L, et al. 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]

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]

Kanne 2020
Kanne J. Chest CT findings in 2019 novel Coronavirus (2019-nCoV) infections from Wuhan, China: key points for the radiologist. *Radiology* 2020 Apr;295(1):16-7. [DOI: 10.1148/radiol.2020200241]

Korevaar 2020
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. [DOI: 10.1183/13993003.01377-2020]

Lee 2020
Lee EY, Ng MY, Khong PL.COVID-19 pneumonia: what has CT taught us? *Lancet Infectious Diseases* 2020 Apr;20(4):384-5. [DOI: 10.1016/S1473-3099(20)30134-1]

Li 2020
Li Y, Xia L. Coronavirus disease 2019 (COVID-19): role of chest CT in diagnosis and management. *American Journal of Roentgenology* 2020;214(6):1280-6. [DOI: 10.2214/AJR.20.22954]

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: 10.1016/j.ejrad.2017.05.032]

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]

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 EY, 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]

Pan 2020
Pan Y, Guan H, Zhou S, Wang Y, Li Q, Zhu T et al. Initial CT findings and temporal changes in patients with the novel coronavirus pneumonia (2019-nCoV): a study of 63 patients in Wuhan, China. *European Radiology* 2020 Jun;30(6):3306-9. [DOI: 10.1007/s00330-020-06731-x]

Pedersen 2020
Pedersen, TL. Package 'ggforce'. www.ggforce.data-imaginist.com 2020 (accessed 5 January 2021).

Peng 2020a
Peng QY, Wang XT, Zhang LN. Findings of lung ultrasonography of novel corona virus pneumonia during the 2019-2020 epidemic. *Intensive Care Medicine* 2020;46(5):849-50. [DOI: 10.1007/s00134-020-05996-6]

Prokop 2020
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):E97-104. [DOI: 10.1148/radiol.2020201473]

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]

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]

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;20(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]

Soldati 2020
Soldati G, Smargiassi A, Inchingolo R, Buonsenso D, Perrone T, Briganti DF, et al. Proposal for international Standardization of the use of lung ultrasound for patients with COVID-19: a simple, quantitative, reproducible method. *Journal of Ultrasound in Medicine* 2020 Jul;39(7):1413-9. [DOI: 10.1002/jum.15285]

StataCorp 2019 [Computer program]
Stata Statistical Software. StataCorp, Version Release 16. College Station, TX: StataCorp LLC, 2019.

Walker 2020
Walker M. What does asymptomatic COVID-19 look like under the surface? www.medpagetoday.com/infectiousdisease/covid19/87168 2020 (accessed 25 July 2021).

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

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

Ye 2020
Ye Z, Zhang Y, Wang Y, Huang Z, Song B. Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review. *European Radiology* 2020;30(8):4381-9. [DOI: 10.1007/s00330-020-06801-0]

Zhang 2020
Zhang JJ, Dong X, Cao YY, Yuan YD, Yang YB, Yan YQ, et al. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. *Allergy* 2020 Jul;75(7):1730-41. [DOI: 10.1111/all.14238]

Zhao 2020
Zhao W, Zhong Z, Xie X, Yu Q, Liu J. Relation between chest CT findings and clinical conditions of coronavirus disease (COVID-19) pneumonia: a multicenter study. *American Journal of Roentgenology* 2020 May;214(5):1072-7. [DOI: 10.2214/AJR.20.22976]

Zhou 2020
Zhou Z, Guo D, Li C, Fang Z, Chen L, Yang R, et al. Coronavirus disease 2019: initial chest CT findings. *European Radiology* 2020 Aug;30(8):4398-406. [DOI: 10.1007/s00330-020-06816-7]

References to studies included in this review

Ai 2020a
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):E32-40. [DOI: 10.1148/radiol.2020200642]

Aslan 2020
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 Mar;27(2):181-7. [DOI: 10.5152/dir.2020.20350]

Bahrami-Motlagh 2020
Bahrami-Motlagh H, Abbasi S, Haghighimorad M, Salevatipour B, Alavi Darazam I, Sanei Taheri M, et al. Performance of Low-Dose Chest CT Scan for Initial Triage of COVID-19. *Iranian Journal of Radiology* 2020 Oct 11;17(4):e104950. [DOI: 10.5812/iranjradiol.104950]

Barbosa 2020
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. [DOI: 10.1590/0100-3984.2020.0040]

Bellini 2020
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 Sep 23;31(4):1932-40. [DOI: 10.1007/s00330-020-07273-y]

Besutti 2020
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. [DOI: 10.1007/s00330-020-07050-x]

Bock 2021
Bock A, Lassen AT, Laursen CB, Posth S. Lung ultrasound as a prognostic tool in emergency patients clinically suspected of COVID-19. *Danish Medical Journal* 2021 Jan 7;68(2):A07200551.

Bollineni 2021
Bollineni VR, Nieboer KH, Döring S, Buls N, de Mey J. The role of CT imaging for management of COVID-19 in epidemic area: early experience from a University Hospital. *Insights into Imaging* 2021 Jan 29;12(1):10. [DOI: 10.1186/s13244-020-00957-5]

Borakati 2020
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: 10.1101/2020.07.07.20147934]

Bosso 2021
Bosso G, Allegorico E, Pagano A, Porta G, Serra C, Minerva V, et al. Lung ultrasound as diagnostic tool for SARS-CoV-2 infection. *Internal and Emergency Medicine* 2021 Oct 03;16(2):471-6. [DOI: 10.1007/s11739-020-02512-y]

Boussouar 2020
Boussouar S, Wagner M, Donciu V, Pasi N, Salem JE, Renard-Penna R, et al. Diagnostic performance of chest computed tomography during the epidemic wave of COVID-19 varied as a function of time since the beginning of the confinement in France. *PLOS One* 2020 Nov 23;15(11):e0242840. [DOI: 10.1371/journal.pone.0242840]

Brun 2021
Brun AL, Gence-Breney A, Trichereau J, Ballester MC, Vasse M, Chabi ML, et al. COVID-19 pneumonia: high diagnostic accuracy of chest CT in patients with intermediate clinical probability. *European Radiology* 2021 Apr;31(4):1969-77. [DOI: 10.1007/s00330-020-07346-y]

Caruso 2020
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;296(2):E79-E85. [DOI: 10.1148/radiol.2020201237]

Cengel 2021
Cengel F, Gurkan O, Calik M, Demirkol MA, Sargin Altunok E, Kaya MF, et al. Diagnosis of the coronavirus disease 2019 with chest computed tomography: a retrospective inter-observer agreement study between radiologists and clinicians. *Hong Kong Journal of Emergency Medicine* 2021 Jan 1;28(1):15-21. [DOI: 10.1177/1024907920968648]

Colombi 2020a
Colombi D, Petrini M, Maffi G, Villani GD, Bodini FC, Morelli N, et al. Comparison of admission chest computed tomography and lung ultrasound performance for diagnosis of COVID-19 pneumonia in populations with different disease prevalence. *European Journal of Radiology* 2020 Dec 1;133:109344. [DOI: 10.1016/j.ejrad.2020.109344]

Cozzi 2020
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. [DOI: 10.1016/j.ejrad.2020.109272]

Dafydd 2021
Ap Dafydd D, O'Mahony M, Jhanji S, Devaraj A, Allum W, Nicol D, et al. The role of CT chest in screening for asymptomatic COVID-19 infection in self-isolating patients prior to elective oncological surgery: findings from a UK Cancer Hub. *British Institute of Radiology* 2021 Jan 1;94(1117):20200994. [DOI: 10.1259/bjr.20200994]

Debray 2020
Debray M-P, Tarabay H, Males L, Chalhoub N, Mahdjoub E, Pavlovsky T, et al. Observer agreement and clinical significance of chest CT reporting in patients suspected of COVID-19. *European Society of Radiology* 2020;31(2):1081–9. [DOI: 10.1007/s00330-020-07126-8]

Deng 2020
Deng Z, Zhang X, Li Y, Xu H, Gang Y, Wang H, et al.[The value of chest CT screening in the early stage of a new coronavirus pneumonia outbreak]. *Chinese Journal of Radiology* 2020 Mar 01;5(54):430-4. [DOI: 10.3760/cma.j.cn112149-20200218-00187]

De Smet 2020
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: 10.1101/2020.05.18.20097444]

Dimeglio 2021
Dimeglio C, Collot S, Abravanel F, Sauné K, Lhomme S, Faruch M, et al. Diagnosis options in patients suffering from COVID-19-like symptoms. *Journal of Medical Virology* 2021;93(7):4076-7. [DOI: 10.1002/jmv.26724]

Dini 2020
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):523-7. [DOI: 10.4081/monaldi.2020.1446]

Djangang 2020
Ndieugnou Djangang N, Peluso L, Talamonti M, Izzi A, Gevenois PA, Garufi A, et al. Eosinopenia in COVID-19 Patients: A Retrospective Analysis. *Microorganisms* 2020 Dec;8(12):1929. [DOI: 10.3390/microorganisms8121929]

Dofferhoff 2020
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.

Dogan 2020
Doğan D, Cüce F, Akay S, Ayaz T, Kinik D, Savaşçi Ü et al. Normal chest CT prevalence in coronavirus disease 2019 (covid-19) patients: A report of 791 cases. *Acta Medica Mediterranea* 2020 Sep 21;36(5):2917-21. [DOI: 10.19193/0393-6384_2020_5_447]

Ducray 2020
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]

Erxleben 2021
Erxleben C, Adams LC, Albrecht J, Petersen A, Vahldiek JL, Thieß H-M, et al. Improving CT accuracy in the diagnosis of COVID-19 in a hospital setting. *Clinical Imaging* 2021 Aug;76:1-5. [DOI: 10.1016/j.clinimag.2021.01.026]

Falaschi 2020
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. [DOI: 10.1016/j.ejrad.2020.109192]

Ferda 2020
Ferda J, Vítovec M, Baxa J, Sedláček D, Havel D, Topolčan O, et al. Chest ct: A valuable tool in discrimination of covid-19 pneumonia, community acquired pneumonia and the other pathologies in slow epidemic phase. *Ceska Radiologie* 2020;74(3):171-9.

Fink 2021
Fink N, Rueckel J, Kaestle S, Schwarze V, Gresser E, Hoppe B, et al. Evaluation of patients with respiratory infections during the first pandemic wave in Germany: characteristics of COVID-19 versus non-COVID-19 patients. *BMC Infectious Diseases* 2021 Feb 10;21(1):167. [DOI: 10.1186/s12879-021-05829-x]

Fonsi 2020
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 Sep 07;40(6):1113-23. [DOI: 10.1002/jum.15487]

Fujioka 2020
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. [DOI: 10.3390/diagnostics10090608]

Gaia 2020
Gaia C, Maria Chiara C, Silvia L, Chiara A, Maria Luisa DC, Giulia B, 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. [DOI: 0.1007/s11547-020-01256-1]

Giannitto 2020
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(12):1260-70. [DOI: 10.1007/s11547-020-01269-w]

Gietema 2020
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]

Gil-Rodrigo 2020
Gil-Rodrigo A, Llorens P, Martínez-Buendía C, Luque-Hernández M-J, Espinosa B, Ramos-Rincón JM. Diagnostic yield of point-of-care ultrasound imaging of the lung in patients with COVID-19. *Emergencias* 2020 Sep;32(5):340-4.

Grando 2020
Grando RD, Brentano VB, Zanardo AP, Hertz FT, Júnior LC, Prietto dos Santos JF, et al. Clinical usefulness of tomographic standards for COVID-19 pneumonia diagnosis: Experience from a Brazilian reference center. *Brazilian Journal of Infectious Diseases* 2020 Nov;24(6):524-33. [DOI: 10.1016/j.bjid.2020.10.002]

Gross 2021
Gross A, Heine G, Schwarz M, Thiemig D, Gläser S, Albrecht T. Structured reporting of chest CT provides high sensitivity and specificity for early diagnosis of COVID-19 in a clinical routine setting. *British Journal of Radiology* 2021 Jan 1;94(1117):20200574. [DOI: 10.1259/bjr.20200574]

Guillo 2020
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. [DOI: 10.1016/j.ejrad.2020.109209]

Gumus 2020
Gümüs T, Kabaoglu ZU, Coskan B, Kartal F, Artukoglu F, Atasoy KC. Preoperative computerized tomography screening for COVID-19 pneumonia in asymptomatic patients: experiences from two centers. *Japanese Journal of Radiology* 2021 Mar;39(3):240-5. [DOI: 10.1007/s11604-020-01061-w]

Haak 2021
Haak SL, Renken IJ, Jager LC, Lameijer H, van der Kolk BY. Diagnostic accuracy of point-of-care lung ultrasound in COVID-19. *Emergency Medicine Journal* 2021 Feb 1;38(2):94-9. [DOI: 10.1136/emermed-2020-210125]

Hanif 2021
Hanif N, Rubi G, Irshad N, Ameer S, Habib U, Zaidi SRH. Comparison of HRCT Chest and RT-PCR in Diagnosis of COVID-19. *Journal of the College of Physicians and Surgeons Pakistan* 2021 Jan;30(1):S1-6. [DOI: 10.29271/jcpsp.2021.01.S1]

He 2020
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 Jul;168:105980. [DOI: 10.1016/j.rmed.2020.105980]

Hermans 2020
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. [DOI: 10.1007/s10140-020-01821-1]

Hernigou 2020
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 Aug;44(8):1571–80. [DOI: 10.1007/s00264-020-04651-5]

Herpe 2020
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 Sep 1;298(2):E81-E87. [DOI: 10.1148/radiol.2020202568]

Hwang 2020
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 Oct;21(10):1150-60. [DOI: 10.3348/kjr.2020.0536]

Ippolito 2020
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. [DOI: 10.1016/j.ejrad.2020.109092]

Jalil 2020
Jalil BA, Khan A, Kugasia IR, Ijaz M. Lung ultrasound in early SARS-CoV-2 pneumonia and the LUS-CoV criteria. *Baylor University Medical Center Proceedings* 2020 Oct 26;34(1):1-4. [DOI: 10.1080/08998280.2020.1834658]

Krdzalic 2020
Krdzalic J, De Jaegere TM, Kwee RM. Diagnostic performance of chest CT in screening patients with suspected COVID-19 infection in a Western population. *British Journal of Radiology* 2020 Sep 1;93(1113):20200643. [DOI: 10.1259/bjr.20200643]

Kuzan 2020
Kuzan TY, Murzoğlu Altıntoprak K, Çiftçi HÖ, Ergül U, Ünal Özdemir NB, Bulut M, 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* 2020 May;27(3):336-43. [DOI: 10.5152/dir.2020.20270]

Lieveld 2021a
Lieveld AW, Azijli K, Teunissen BP, van Haaften RM, Kootte RS, van den Berk IA, et al. Chest CT in COVID-19 at the ED: validation of the COVID-19 Reporting and Data System (CO-RADS) and CT Severity Score: a prospective, multicenter, observational study. *Chest* 2021 Mar 1;159(3):1126-35. [DOI: 10.1016/j.chest.2020.11.026]

Lieveld 2021b
Lieveld AW, Kok B, Schuit FH, Azijli K, Heijmans J, van Laarhoven A, et al. Diagnosing COVID-19 pneumonia in a pandemic setting: Lung Ultrasound versus CT (LUVCT) – a multicentre, prospective, observational study. *ERJ Open Research* 2020 Oct;6(4):00539-2020. [DOI: 10.1183/23120541.00539-2020]

Luo 2020a
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 May 07;20(1):129. [DOI: 10.1186/s12890-020-1170-6]

Majeed 2020
Majeed T, Ali RS, Solomon J, Mesri M, Sharma S, Shamim S, et al. The role of the computed tomography (CT) thorax in the diagnosis of COVID-19 for patients presenting with acute surgical emergencies. a single institute experience. *Indian Journal of Surgery* 2020 Oct 20;82(6):1005-10. [DOI: 10.1007/s12262-020-02626-9]

Mei 2020
Mei X, Lee H-C, Diao K, Huang M, Lin B, Liu C, et al. Artificial intelligence-enabled rapid diagnosis of COVID-19 patients. Nature Medicine 2020 Aug;26(8):1224-8. [DOI: 10.1038/s41591-020-0931-3]

Miranda Magalhaes Santos 2020
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. [DOI: 10.1097/RCT.0000000000001054]

Moroni 2021
Moroni C, Cozzi D, Albanesi M, Cavigli E, Bindi A, Luvarà S, et al. Chest X-ray in the emergency department during COVID-19 pandemic descending phase in Italy: correlation with patients' outcome. *Radiologia Medica* 2021 May;126(5):661-8. [DOI: 10.1007/s11547-020-01327-3]

Murphy 2020
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. [DOI: 10.1148/radiol.202020187]

Narinx 2020
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. [DOI: 10.1007/s10140-020-01849-3]

Nivet 2021
Nivet H, Crombé A, Schuster P, Ayoub T, Pourriol L, Favard N, et al. The accuracy of teleradiologists in diagnosing COVID-19 based on a French multicentric emergency cohort. *European Radiology* 2021 May 1;31(5):2833-44. [DOI: 10.1007/s00330-020-07345-z]

O'Neill 2020
O'Neill SB, Byrne D, Müller NL, Jalal S, Parker W, Nicolaou S, et al. Radiological Society of North America (RSNA) Expert consensus statement related to chest CT findings in COVID-19 versus CO-RADS: comparison of reporting system performance among chest radiologists and end-user preference. *Canadian Association of Radiologists' Journal* 2020;0846537120968919:Epub 2020 Nov 3. PMID: 33138634. [DOI: 10.1177/0846537120968919]

Ohana 2021
Ohana M, Muller J, Severac F, Bilbault P, Behr M, Oberlin M, et al. Temporal variations in the diagnostic performance of chest CT for Covid-19 depending on disease prevalence: Experience from North-Eastern France. *European Journal of Radiology* 2021 Jan;134:109425. [DOI: 10.1016/j.ejrad.2020.109425]

Ooi 2021
Ooi MW, Liong SY, Baguley N, Sharman A, Tuck J. Role of complementary Ct chest in patients presenting with acute abdominal symptoms during covid-19 pandemic: a UK experience. *Clinical Imaging* 2021 Jan;69:289-92. [DOI: 10.1016/j.clinimag.2020.09.00]

Pagano 2021
Pagano A, Finkelstein M, Overbey J, Steinberger S, Ellison T, Manna S, et al. Portable chest radiography as an exclusionary test for adverse clinical outcomes during the COVID-19 pandemic. *Chest* 2021 Jul;160(1):238-48. [DOI: 10.1016/j.chest.2021.01.053]

Palmisano 2021
Palmisano A, Scotti GM, Ippolito D, Morelli MJ, Vignale D, Gandola D, et al. Chest CT in the emergency department for suspected COVID-19 pneumonia. *Radiologia Medica* 2021 Mar 1;126(3):498-502. [DOI: 10.1007/s11547-020-01302-y]

Pare 2020
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 Jun 19;21(4):771-8. [DOI: 10.5811/westjem.2020.5.47743]

Patel 2020
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;203(4). [DOI: 10.1101/2020.05.26.20114082]

Patrucco 2021
Patrucco F, Carriero A, Falaschi Z, Paschè A, Gavelli F, Airoldi C, et al. COVID-19 diagnosis in case of two negative nasopharyngeal swabs: association between chest CT and bronchoalveolar lavage results. *Radiology* 2021 Jan 5;298(3):E152-5. [DOI: 10.1148/radiol.2020203776]

Peng 2020a
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: 10.1101/2020.04.07.20057315]

Pivetta 2021
Pivetta E, Goffi A, Tizzani M, Locatelli SM, Porrino G, Losano I, et al. Lung ultrasonography for the diagnosis of SARS-CoV-2 pneumonia in the emergency department. *Annals of Emergency Medicine* 2021 Apr 1;77(4):385-94. [DOI: 10.1016/j.annemergmed.2020.10.008]

Puylaert 2020
Puylaert CA Scheijmans JC, Borgstein AB, Andeweg CS, Bartels-Rutten A, Beets GL, et al. Yield of Screening for COVID-19 in asymptomatic patients prior to elective or emergency surgery using chest CT and RT-PCR (SCOUT): multicenter study. *Annals of Surgery* 2020 Dec;272(6):919-24. [DOI: 10.1097/SLA.0000000000004218]

Ravikanth 2021
Ravikanth R. Diagnostic accuracy and false-positive rate of chest CT as compared to RT-PCR in coronavirus disease 2019 (COVID-19) pneumonia: a prospective cohort of 612 cases from India and review of literature. *Indian Journal of Radiology & Imaging* 2021 Jan;31(Suppl 1):S161-9. [DOI: 10.4103/ijri.IJRI_377_20]

Reginelli 2021
Reginelli A, Grassi R, Feragalli B, Belfiore MP, Montanelli A, Patelli G, et al. Coronavirus disease 2019 (COVID-19) in Italy: double reading of Ccest CT examination. *Biology (Basel)* 2021 Jan 25;10(2):89. [DOI: 10.3390/biology10020089]

Rona 2021
Rona G, Voyvoda N, Arifoğlu M, Karaaslan A, Çetin C. The efficacy of chest computed tomography in pediatric patients with suspected COVID-19. *Journal of Computer Assisted Tomography* 2021 Mar 1;45(2):337-41. [DOI: 10.1097/RCT.0000000000001127]

Roy Choudhury 2020
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 Jul;6(3):00357-2020. [DOI: 10.1183/23120541.00357-2020]

Saeed 2020
Saeed GA, Helali AA, Almazrouei S, Shah A, Ahmed LA. Chest CT features of COVID-19 in the region of Abu Dhabi, UAE: a single institute study. *Chinese Journal of Academic Radiology* 2021;4:248–56. [DOI: 10.1007/s42058-021-00075-1]

Salehi-Pourmehr 2020
Salehi-Pourmehr H, Pourfathi H, Tarzamni MK, Ghojazadeh M, Naghili B, Zarrintan A, et al. Diagnostic value of chest CT in Iranian patients with suspected COVID-19. *Caspian Journal of Internal Medicine* 2020;11(Suppl 1):527-30. [DOI: 10.22088/cjim.11.0.527]

Schalekamp 2020
Schalekamp S, Bleeker-Rovers CP, Beenen LF, Quarles van Ufford HM, Gietema HA, Stoger JL, et al. Chest CT in the emergency department for diagnosis of COVID-19 pneumonia: Dutch experience. *Radiology* 2020 Nov 17;298(2):E98-106. [DOI: 10.1148/radiol.2020203465]

Schmid 2020
Schmid B, Feuerstein D, Lang CN, Fink K, Steger R, Rieder M, et al. Lung ultrasound in the emergency department - a valuable tool in the management of patients presenting with respiratory symptoms during the SARS-CoV-2 pandemic. *BMC Emergency Medicine* 2020 Dec 7;20(1):96. [DOI: 10.1186/s12873-020-00389-w]

Schulze-hagen 2020
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]

Shah 2021
Shah JV, Shah C, Shah S, Gandhi N, Dikshit NA, Patel P, et al. HRCT chest in COVID-19 patients: An initial experience from a private imaging center in western India. *Indian Journal of Radiology & Imaging* 2021 Jan;31(Suppl 1):S182-6. [DOI: 10.4103/ijri.IJRI_405_20]

Skalidis 2020
Skalidis I, Nguyen VK, Bothorel H, Poli L, Da Costa RR, Younossian AB, et al. Unenhanced computed tomography (CT) utility for triage at the emergency department during COVID-19 pandemic. *American Journal of Emergency Medicine* 2021 Aug;46:260-5. [DOI: 10.1016/j.ajem.2020.07.058]

Song 2020a
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 Jun;7(6):ofaa171. [DOI: 10.1093/ofid/ofaa171]

Sorlini 2021
Sorlini C, Femia M, Nattino G, Bellone P, Gesu E, Francione P, et al. The role of lung ultrasound as a frontline diagnostic tool in the era of COVID-19 outbreak. *Internal and Emergency Medicine* 2021 Apr;16(3):749-56. [DOI: 10.1007/s11739-020-02524-]

Speidel 2021
Speidel V, Conen A, Gisler V, Fux CA, Haubitz S. Lung assessment with point-of-care ultrasound in respiratory coronavirus disease (COVID-19): a prospective cohort study. *Ultrasound in Medicine & Biology* 2021 Apr 1;47(4):896-901. [DOI: 10.1016/j.ultrasmedbio.2020.12.021]

Steuwe 2020
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. [10.1088/1361-6498/aba16a]

Stevens 2020
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 Feb;27(1):90-4. [DOI: 10.1016/j.radi.2020.06.010]

Sukhija 2021
Sukhija A, Mahajan M, Joshi PC, Dsouza J, Seth ND, Patil KH. Radiographic findings in COVID-19: Comparison between AI and radiologist. *Indian Journal of Radiology and Imaging* 2021 Jan;31(Suppl 1):S87-93. [10.4103/ijri.IJRI_777_20]

Sverzellati Nicola 2021
Sverzellati N, Ryerson CJ, Milanese G, Renzoni EA, Volpi A, Spagnolo P, et al. Chest x-ray or CT for COVID-19 pneumonia? Comparative study in a simulated triage setting. *European Respiratory Journal* 2021 Sep 9;58(3):2004188. [DOI: 10.1183/13993003.04188-2020]

Teichgraber 2021
Teichgräber U, Malouhi A, Ingwersen M, Neumann R, Reljic M, Deinhardt-Emmer S, et al. Ruling out COVID-19 by chest CT at emergency admission when prevalence is low: the prospective, observational SCOUT study. *Respiratory Research* 2021 Jan 12;22(1):13. [10.1186/s12931-020-01611-w]

Tsakok 2020
Tsakok M, Shaw R, Murchison A, Ather S, Xie C, Watson R, et al. Diagnostic accuracy of initial chest radiograph compared to SARS-CoV-2 PCR in patients with suspected COVID-19. *BJR Open* 2020 Aug 5;2(1):20200034. [DOI: 10.1259/bjro.20200034]

Wang 2020a
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 Mar 28;45(3):229-35. [DOI: 10.11817/j.issn.1672-7347.2020.200152]

Wehbe 2021
Wehbe RM, Sheng J, Dutta S, Chai S, Dravid A, Barutcu S, et al. Deep COVID-XR: an artificial intelligence algorithm to detect COVID-19 on chest radiographs trained and tested on a large U.S. clinical data set. *Radiology* 2021 Apr;299(1):E167-76. [DOI: 10.1148/radiol.2020203511]

Xiaocheng 2020
Xiaocheng X, Haiyan J, Xiaoping C, Yi Z, Huiyuan S. Preliminary analysis of visits to fever clinics during the epidemic of new coronavirus pneumonia. *Academic Journal of Second Military Medical University* 2020;41(8):828-31. [DOI: 10.16781/j.0258-879x.2020.08.0828]

Xiong 2020
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 Apr 28;100(16):1223-9. [DOI: 10.3760/cma.j.cn112137-20200228-00499]

Yassa 2020
Yassa M, Yirmibes C, Cavusoglu G, Eksi H, Dogu C, Usta C, et al. Outcomes of universal SARS-CoV-2 testing program in pregnant women admitted to hospital and the adjuvant role of lung ultrasound in screening: a prospective cohort study. *Journal of Maternal-Fetal & Neonatal Medicine* 2020 Nov;33(22):3820-6. [DOI: 10.1080/14767058.2020.1798398]

Yates 2021
Yates A, Dempsey PJ, Vencken S, MacMahon PJ, Hutchinson BD. Structured reporting in portable chest radiographs: An essential tool in the diagnosis of COVID-19. *European Journal of Radiology* 2021 Jan;134:109414. [DOI: 10.1016/j.ejrad.2020.109414]

Chapter 4. Prognostic accuracy of imaging findings for predicting morbidity and mortality in patients with COVID-19: systematic review protocol

List of Authors

1. Nayaar Islam, BHSc. School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa. Department of Radiology, University of Ottawa. Clinical Epidemiology Program, Ottawa Hospital Research Institute.
2. Mohammed Kashif Alghita, BSc. Department of Radiology, University of Ottawa. Clinical Epidemiology Program, Ottawa Hospital Research Institute.
3. Sanam Ebrahimzadeh, PhD. Department of Radiology, University of Ottawa. Clinical Epidemiology Program, Ottawa Hospital Research Institute.
4. Haben Dawit, BSc. Department of Radiology, University of Ottawa. Clinical Epidemiology Program, Ottawa Hospital Research Institute.
5. Ross Prager, MD. Department of Medicine, University of Ottawa.
6. Gonzalo G Alvarez, MD, MPH. Clinical Epidemiology Program, Ottawa Hospital Research Institute. Department of Medicine, The Ottawa Hospital.
7. Jérémie F Cohen, MD, PhD. Obstetrical, Perinatal and Pediatric Epidemiology Research Team (EPOPé), Centre of Research in Epidemiology and StatisticS (CRESS), Inserm UMR1153, Université de Paris.
8. Daniël A Korevaar, MD, PhD. Department of Respiratory Medicine, Amsterdam UMC, University of Amsterdam.
9. Jonathan J Deeks, PhD. Test Evaluation Research Group, Institute of Applied Health Research, University of Birmingham. NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham.
10. Jan Y Verbakel, MD, PhD. Department of Public Health and Primary Care, KU Leuven. Nuffield Department of Primary Care Health Sciences, University of Oxford.

27. Johanna AAG Damen, PhD. Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam University Medical Centers, University of Amsterdam
11. Eleanor A Ochodo, MD, PhD. Centre for Evidence-based Health Care, Department of Global Health, Faculty of Medicine and Health Sciences, Stellenbosch University. Centre for Global Health Research, Kenya Medical Research Institute.
12. Anirudh Venugopalan Nair, MD. Department of Clinical Imaging, Al Wakra Hospital, Hamad Medical Corporation.
13. Jacqueline Dinnes, PhD. Test Evaluation Research Group, Institute of Applied Health Research, University of Birmingham. NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham.
14. Carole Dennie, MD. Department of Radiology, University of Ottawa. Department of Medical Imaging, The Ottawa Hospital.
15. Ann Van den Bruel, MD, PhD. Academic of Primary Care, KU Leuven.
16. Janneke van de Wijgert, MD, MPH, PhD. Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam University Medical Centers, University of Amsterdam. Institute of Infection, Veterinary, and Ecological Sciences, University of Liverpool
17. Lindsey Sikora, MI. Sciences Library, University of Ottawa.
18. René Spijker, PhD. Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam University Medical Centers, University of Amsterdam. Medical Library, Amsterdam UMC, University of Amsterdam, Amsterdam Public Health.
19. Samanjit S Hare, MD. Department of Radiology, Royal Free London NHS Trust.

20. Matthew DF McInnes, MD, PhD. School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa. Department of Radiology, University of Ottawa. Clinical Epidemiology Program, Ottawa Hospital Research Institute.

Preface

Objective

To develop a protocol to evaluate the accuracy of imaging findings (on chest computed tomography (CT), chest X-ray, and ultrasound of the lungs) to predict morbidity (including complications, escalation of care, and hospital admission length) and mortality across all time horizons in patients with COVID-19.

Funding

1. Frederick Banting and Charles Best Canada Graduate Scholarship - Master's (Canadian Institutes of Health Research)
2. The Cochrane Infectious Diseases Group (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).

Role of funders

None of the listed funding bodies had any role in designing or conducting the study; collecting, managing, analyzing, or interpretating the data; or preparing, reviewing, or approving the manuscript.

Appendices

One appendix (4.1) is included at the end of this document.

Ethics approval

This study did not require ethics approval.

Contributions of co-authors

- Concept and design: Islam, McInnes.
- Search strategy: Sikora, Spijker.
- Drafting of the manuscript: Islam, McInnes.
- Critical revision of the manuscript for important intellectual content: Alghita, Dawit, Prager, Alvarez, Cohen, Korevaar, Deeks, Verbakel, Damen, Ochodo, Nair, Dinnes, Dennie, Van den Bruel, Leeflang, van de Wijgert, Hare.
- Administrative, technical, or material support: McInnes.

Citation example (currently accepted for publication with minor revisions)

Islam N, Alghita MK, Ebrahimzadeh S, Dawit H, Prager R, Alvarez GG, Cohen JF, Korevaar DA, Deeks JJ, Verbakel JY, Damen JAAG, Ochodo EA, Venugopalan Nair A, Dinnes J, Dennie C, Van den Bruel A, van de Wijgert J, Sikora L, Spijker R, Hare SS, McInnes MDF. Prognostic accuracy of imaging findings for predicting morbidity and mortality in patients with COVID-19 [Protocol]. Cochrane Database of Systematic Reviews [Year], Issue [Issue]. Art. No.: CD015318. DOI: 10.1002/14651858.CD015318

Abstract

This is a protocol for a Cochrane Review. The objectives are as follows: to evaluate the accuracy of imaging findings (on chest computed tomography (CT), chest X-ray, and ultrasound of the lungs) to predict morbidity (including complications, escalation of care, and hospital admission length) and mortality across all time horizons in patients with COVID-19. Secondary objectives: investigations of whether prognostic accuracy varies according to covariates of interest (including imaging technique, timing of imaging test, test used to confirm COVID-19, duration of symptoms, timing of outcome confirmation, study design, study setting, participant age, and presence or absence of pulmonary embolism) will be conducted, when data are available.

4.1 Introduction

The most commonly used reference standard test for the diagnosis of COVID-19 is the reverse transcriptase polymerase chain reaction (RT-PCR) test (WHO 2020). Other tests, including imaging tests, may also be used to assist with diagnosis and patient management. The Cochrane systematic review on 'Thoracic imaging tests for the diagnosis of COVID-19' found that chest computed tomography (CT) and ultrasound of the lungs, but not chest X-ray, could be useful in ruling out a diagnosis of COVID-19 (i.e. high sensitivity) (Ebrahimzadeh 2022; Islam 2021; Islam 2020; Salameh 2020). The review also found that chest CT, chest X-ray, and ultrasound of the lungs offer limited capability in differentiating COVID-19 from other respiratory illnesses (i.e. low specificity) (Ebrahimzadeh 2022; Islam 2021; Islam 2020; Salameh 2020). However, as the pandemic continues, there remain uncertainties regarding the prognostic value of patients with COVID-19, including whether imaging tests can predict morbidity and mortality.

Typical workup for individuals suspected of having COVID-19 includes RT-PCR testing and the assessment of signs and symptoms. As patients can present with a wide variety of symptoms, can be classified as having varying disease severities, and may deteriorate quickly, prognostic information is important to anticipate which patients may require closer monitoring and escalation of care. In emergency departments, chest CT, chest X-ray, and ultrasound have been used to indicate patients' disease severity (Redmond 2020).

Potential roles of imaging tests in patient management pathways are as follows.

- Imaging tests could be key in determining whether an individual with COVID-19 needs to be admitted or if the disease can be managed in the community or at home.

- Imaging tests could help anticipate whether an admitted patient is at risk of requiring a higher level of care (e.g. intensive care unit (Burian 2020) or tertiary care) and closer monitoring and follow-up.
- Imaging tests could predict poor prognosis, and thus be useful in determining when a patient should be transferred from a community clinic or hospital to a tertiary care centre before the patient becomes too ill to transfer.
- Imaging tests could also be important in identifying patients who are at risk of requiring specific treatments (e.g. oxygen, intubation, steroids, immunomodulators, antivirals) in the near future, depending on the stage of the infection as seen on imaging.

Research on the prognostic value of imaging in patients with COVID-19 is evolving; however a robust systematic review and meta-analysis on the topic is needed. This Cochrane Review will synthesize evidence on the prognostic accuracy of different imaging findings to predict short-term morbidity (including complications, escalation of care, and hospital admission length) and mortality in patients with COVID-19. The results from this review will inform clinical pathways and policies for management of patients with COVID-19. As the pandemic continues and eventually ends, imaging findings and outcomes of COVID-19 will change; thus, the objectives and findings of this review will be applicable within the context of the COVID-19 pandemic.

4.2 Objectives

To evaluate the accuracy of imaging findings (on chest CT, chest X-ray, and ultrasound of the lungs) to predict morbidity, including:

- complications (e.g. development of hypoxia, severe acute kidney injury (AKI), delirium);
- escalation of care (e.g. hospital admission, intensive care unit (ICU) admission, length of ICU admission, non-invasive ventilation, intubation, mechanical ventilation, ECMO (Extra corporeal Membrane Oxygenation), need for renal replacement therapy, need for transfusion); and
- length of any hospital admission.

and mortality, including

- disease-specific mortality; and
- all-cause mortality

across all time horizons in patients with COVID-19.

The secondary objectives is to investigate, when data are available, whether prognostic accuracy varies according to covariates of interest (including imaging technique, timing of imaging test, test used to confirm COVID-19, duration of symptoms, timing of outcome confirmation, study design, study setting, participant age, and presence or absence of pulmonary embolism (PE) on CT pulmonary angiography (CTPA)).

4.3 Methods

This review follows the format of a generic protocol used across the full series of Cochrane COVID-19 Diagnostic Test Accuracy Reviews (Deeks 2020a). Thus, some text included in the 'Introduction' and 'Methods' sections was originally published in other protocols and reviews from this series (Deeks 2020a; Deeks 2020b; Dinnes 2021; Islam 2020; Islam 2021; Salameh 2020; Stegeman 2020; Struyf 2021; Verbakel 2021).

Criteria for considering studies for this review

Types of studies

The eligibility criteria will be kept broad to be able to include all settings, and all variations of a test. Studies of all designs that include participants with COVID-19, evaluate the ability of imaging test findings or 'scores' to predict morbidity and/or mortality, and produce estimates of prognostic accuracy in terms of sensitivity and specificity (or provide data from which any of these values can be computed) will be included. An example of a study that would be included is Salahshour 2021.

Participants

Studies that recruited participants who have proven COVID-19, as confirmed by any of the following tests, and who are symptomatic, will be included.

- 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;
- positive on WHO criteria for COVID-19 (WHO 2020);
- positive on CDC criteria for COVID-19 (CDC 2021);
- positive serology for SARS-CoV-2 antibodies in addition to consistent symptomatology;
- positive on study-specific list of criteria for COVID-19 which includes:
- other criteria (symptoms, other tests, infected contacts).

Studies that include mixed populations (both symptomatic and asymptomatic participants) will be included only if data for symptomatic participants are reported separately. There will be no age or gender restrictions. Studies with fewer than 10 participants who underwent evaluation with an imaging test(s) will be excluded.

Index tests

This review will focus on imaging test findings and 'scores' (e.g. numerical scores indicating probability or severity of COVID-19 assigned using an established radiological scoring system) detected on chest CT, chest X-ray, and ultrasound of the lungs, all of which are commonly used imaging tests used for assessment of patients with COVID-19. Further details on these index tests are provided in Appendix 4.1.

The imaging test could be performed at any point(s) in the management pathway, and the intended role of the test at the time it was applied could be a diagnostic or prognostic test used concurrently with, or as an add-on test to diagnostic tests or other prognostic tests. Individual and combined imaging modalities, as well as single and repeat testing will be accepted. Imaging findings alone, or in combination with clinical evaluations and/or biomarkers will be accepted. Specific imaging findings include, but are not limited to, ground glass opacities, consolidation, pleural effusion, and pneumomediastinum. Specific radiological scoring systems include, but are not limited to, the COVID-19 Reporting and Data System (CO-RADS) (Prokop 2020), the British Society of Thoracic Imaging (BSTI) COVID-19 Reporting Templates (BSTI 2020), and the CT Severity Score (Yang 2020). While some of these scoring systems have been primarily developed to indicate the probability of COVID-19 on imaging for diagnosis, they are also largely based on the severity of lung involvement in patients with known COVID-19. It was thus

determined, in consultation with clinicians, that using these scoring systems as a proxy for 'imaging severity' (severity of lung involvement as seen on imaging) would be appropriate and that their linkage with prognostic outcomes would be reasonable.

Only index tests interpreted by humans will be included. Studies involving interpretation by an algorithm (machine learning/artificial intelligence (AI)) will be included only if data pertaining to diagnostic accuracy of human interpretation are also provided.

Target conditions

This review will focus on people with symptomatic COVID-19, which is the disease caused by acute infection with SARS-CoV-2 (Datta 2020), who had thoracic imaging as part of their evaluation or care. Individuals with SARS-CoV-2 infection can be asymptomatic; however, these people are not considered to have COVID-19 and are therefore not within the scope of this review. The target condition in this review is the accuracy of imaging tests for the prediction of outcomes concerning morbidity and mortality in COVID-19 patients.

The first target condition is morbidity, including complications (e.g. hypoxia, severe AKI, delirium), escalation of care (e.g. ICU admission, non-invasive ventilation, intubation, mechanical ventilation, ECMO, need for renal replacement therapy, need for transfusion), and length of hospital admission. Each type of morbidity outcome will be analyzed separately.

The second target condition is mortality following a diagnosis of COVID-19, including disease-specific mortality, and all-cause mortality. Each type of mortality outcome will be analyzed separately.

For each morbidity outcome and mortality outcome, all time horizons for which the outcome is predicted (e.g. 30 days, 60 days, etc.) reported in primary studies will be considered

and results will be grouped and synthesized accordingly. If there is a range of time horizons that are not common, studies will be divided into two groups of time horizons: short-term and long-term. The cut-off value used to separate studies into these two groups will be determined based on the range of time horizons across included studies.

Reference Standard

In this prognostic accuracy review, the reference standard section is used to define the target conditions (i.e. outcomes of interest).

With respect to the first target condition of morbidity, all reference standards for outcomes concerning complications (e.g. hypoxia, severe AKI, delirium) will accepted; reference standards may include combinations of clinical, laboratory, and/or imaging findings. Morbidity outcomes concerning escalation of care (e.g. ICU admission, non-invasive ventilation, intubation, mechanical ventilation) and length of hospital admission are objective. Similarly, the second target condition of mortality (including all-cause mortality and disease-specific mortality) consists of objective outcomes.

Search methods for identification of studies

An electronic search will be performed in the Cochrane COVID-19 Study Register (covid-19.cochrane.org), which is a specialized register built within the Cochrane Register of Studies (CRS) and maintained by Cochrane Information Specialists. The register contains study reports from several sources, including:

- monthly searches of the Cochrane Central Register of Controlled Trials (CENTRAL)
- daily searches of PubMed

- weekly searches of Embase.com
- weekly searches of medRxiv
- weekly searches of the WHO International Clinical Trials Registry Platform (ICTRP)
- daily searches of ClinicalTrials.gov

Complete data sources and search methods for the register are available at: community.cochrane.org/about-covid-19-study-register. In addition, the WHO COVID-19 global literature on coronavirus disease database will be searched, with the search restricted to sources not available within the Cochrane COVID-19 Register.

A search strategy that combines a search for patients with confirmed COVID-19, a search for imaging tests, and a filter for prediction studies, will be implemented.

Selection of studies

Screening of studies will be piloted by two reviewers, and a third experienced review author will resolve disagreements to ensure that inclusion and exclusion criteria are applied in a consistent manner.

At the first stage of screening, titles and abstracts will be screened for potentially relevant studies; at the second stage of screening, the full texts of potentially relevant studies will be screened for review eligibility. The review authors will screen studies independently, in duplicate. A third, experienced review author will resolve disagreements about initial title and abstract screening. Disagreements about eligibility assessments at full text screening will be resolved through discussion between three review authors.

For articles only available in languages other than English, Google Translate will be used, or review authors who can read and understand that language will perform translations.

Data extraction and management

Data extraction will first be piloted by two reviewers, and a third experienced review author will resolve disagreements to ensure that data is collected in a consistent manner. The review authors will perform data extraction independently, in duplicate. Disagreements will be resolved through discussion between three review authors.

The following items will be extracted.

1. Study design (e.g. cohort, case-control) study setting (including country), and study dates.
2. Age of study participants, participant eligibility and recruitment method (e.g. consecutive, random), number of participants, participants' duration of symptoms and/or COVID-19 disease course, number of participants with medical outcome(s) of interest, number of imaging studies (and if more than one study was done per participant), and other relevant participant parameters (e.g. demographics, co-morbidities, details of medical treatments).
3. Details on the medical outcome(s) of interest, as applicable (e.g. morbidities, mortality), including definition and method for measurement of outcomes, length of follow-up, and timing of outcome(s) relative to disease course. If a study reports multiple different measures and/or outcomes of interest, all relevant measures and outcomes will be extracted.

4. Specific imaging findings, features, and/or ‘scores’ as identified by CT, chest X-ray and/or ultrasound of the lungs, and timing of imaging test(s) relative to disease course.
5. Details regarding interpretation of the imaging test(s) (i.e., CT, chest X-ray and/or ultrasound of the lungs) including technical parameters, level of training of readers, number of readers, and the inter-observer variability.
6. Details of the test(s) used to confirm COVID-19. If RT-PCR was performed, timing of test, number of tests and method of acquisition (or similar details regarding other tests).
7. Details on sample size calculation, including if and how this was conducted, and the calculated numbers of participants, events, and person-time of follow-up.
8. 2x2 contingency tables of the number of true positives, false positives, false negatives and true negatives or summary statistics from which prognostic accuracy (sensitivity and specificity) can be computed. If a study reports accuracy data for both an AI algorithm and one or more radiologists, only the 2x2 contingency table corresponding to the radiologist accuracy data will be extracted.
9. Number of participants with any missing value, with respect to imaging test findings and/or medical outcome(s) of interest (i.e., morbidity, mortality).
10. Information that will be required for the risk of bias assessment outlined in the following section, 'Assessment of methodological quality'.

Authors of included studies will be contacted to obtain missing data and/or additional information, if necessary.

Assessment of methodological quality

The review authors will assess the risk of bias and applicability concerns independently, in duplicate, using the Quality Assessment of Prognostic Accuracy Studies (QUAPAS) tool (Table 1) for prognostic accuracy studies (Lee 2020). Risk of bias assessment will first be piloted by two reviewers, and a third experienced review author will resolve disagreements to ensure that the assessment is completed in a consistent manner.

The QUAPAS tool includes components of the QUADAS-2 tool (which consists of four domains: patient selection, index test, reference standard and flow and timing (Whiting 2011)) supplemented by components of the Quality in Prognostic Studies (QUIPS) (Hayden 2013) and Prediction model study Risk of Bias Assessment (PROBAST) (Wolff 2019) tools, with a fifth domain of "Analysis" added to the quality appraisal. Disagreements will be resolved through discussion between three review authors.

Since this review will not assess predictive performance of prognostic factors or models, the review authors determined that using QUIPS (for prognostic factors) or PROBAST (for prognostic models) would not be appropriate.

Assessment of the certainty of evidence

First, the authors will determine their key findings. Of the target conditions that this review will evaluate, the mortality target conditions (including disease-specific mortality, and all-cause mortality) will have primary importance, but morbidity target conditions will also have high importance. The review authors will determine the certainty in the summary estimates corresponding to the key findings for each imaging test and target condition, using the GRADE approach (Schünemann 2020). Following the GRADE guidance (Schünemann 2020), each meta-analysis will start with having high certainty, and will be downgraded by one level when at least

half of the studies have a high risk of bias on one or more domains. Each meta-analysis will be downgraded by one level for indirectness when at least half of the studies in the meta-analyses have high concerns regarding applicability on at least one domain. Each meta-analysis will be downgraded by one level for imprecision when fewer people with the target condition are included than would have been needed to achieve the sensitivity estimates listed, with a width of the confidence interval of at most 10 percentage points. Each meta-analysis will be downgraded by one level for inconsistency when study estimates differ more than 20 percentage points from each other.

Statistical analysis and data synthesis

Prognostic accuracy will be presented (as estimates of sensitivity and specificity with 95% confidence intervals) using forest plots and summary tables, separately for each combination of the following: imaging test or imaging finding(s), individual medical outcome of interest, and timing of outcome confirmation relative to disease course and/or imaging test. Data will be analyzed on a participant level, not a lesion or lung-segment level, since this is what determines care.

When four or more studies report a particular imaging test or imaging finding(s), medical outcome of interest, and timing of outcome confirmation relative to disease course and/or imaging test, meta-analysis will be performed using a bivariate model for meta-analysis which takes into account for the within- and between-study variance, and the correlation between sensitivity and specificity across studies (Chu 2006; Reitsma 2005). For studies in which data using different radiological scoring systems and at different thresholds (i.e., ‘scores’) for the same imaging test are reported, data will be synthesized for each scoring system and threshold

separately. Meta-analyses will be undertaken using the metandi function in STATA (Harbord 2009; StataCorp 2019).

Investigations of heterogeneity

Heterogeneity will be investigated by visual inspection of forest plots and summary receiver operating characteristics (SROC) plots. Specific sources of heterogeneity will be investigated if adequate data are available on covariates of interest (e.g. imaging technique, timing of imaging test, test used to confirm COVID-19, duration of symptoms, timing of outcome confirmation, study design, study setting, participant age, presence or absence of PE on CTPA), either using stratification (when it is deemed inappropriate to combine studies) or through meta-regression models with the meqrlogit function in STATA (StataCorp 2019).

The impacts of population vaccination status and different COVID-19 variants will also be considered. Results will be stratified by vaccination status and variant, if known. If studies do not provide information on these variables, results will instead be stratified by study period as an indirect method for evaluating the impact of changes in vaccination status and variants, since both have both progressed over the course of the pandemic.

Sensitivity Analyses

Sensitivity analyses will be done by restricting meta-analyses to studies that have a low risk of bias for all QUAPAS domains, if available. Additional sensitivity analyses will be done by restricting meta-analyses to studies that control for underlying comorbidities that may contribute to the outcome measures of interest, if available.

References

BSTI 2020
COVID-19 BSTI Reporting Templates and Codes. British Society of Thoracic Imaging 2020 May 22.

Burian 2020
Burian E, Jungmann F, Kaissis GA, Lohöfer FK, Spinner CD, Lahmer T, et al. Intensive Care Risk Estimation in COVID-19 Pneumonia Based on Clinical and Imaging Parameters: Experiences from the Munich Cohort. Journal of Clinical Medicine 2020 May;9(5):1514. [DOI: 10.3390/jcm9051514]

CDC 2021
Coronavirus Disease 2019 (COVID-19) 2021 Case Definition. Centers for Disease Control and Prevention 2021 Aug 24.

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]

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 Dec 8;324(22):2251. [DOI: 10.1001/jama.2020.22717]

Deeks 2020a
Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Leeflang MM, Spijker R, et al. Diagnosis of SARS-CoV-2 infection and COVID-19: accuracy of signs and symptoms; molecular, antigen, and antibody test; and routine laboratory markers [Protocol]. Cochrane Database of Systematic Reviews 2020 Apr 24, Issue 4. Art. No.: CD013596. DOI: 10.1002/14651858.CD013596.

Deeks 2020b
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 Jun 25, Issue 6. Art. No.: CD013652. DOI: 10.1002/14651858.CD013652.

Dinnes 2021
Dinnes J, Deeks JJ, Berhane S, Taylor M, Adriano A, Davenport C, et al. Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. Cochrane Database of Systematic Reviews 2021 Mar 24, Issue 3. Art. No.: CD013705. DOI: 10.1002/14651858.CD013705.pub2.

Ebrahimzadeh 2022
Ebrahimzadeh S, Islam N, Dawit H, Salameh J-P, Kazi S, Fabiano N, et al. Thoracic imaging tests for the diagnosis of COVID-19. Cochrane Database of Systematic Reviews 2022, Issue 5. Art. No.: CD013639. DOI: 10.1002/14651858.CD013639.pub5.

Foroutan 2020
Foroutan F, Guyatt G, Zuk V, Vandvik PO, Alba AC, Mustafa R, et al. GRADE Guidelines 28: Use of GRADE for the assessment of evidence about prognostic factors: rating certainty in identification of groups of patients with different absolute risks. Journal of Clinical Epidemiology 2020 May 1;121:62–70. [DOI: 10.1016/j.jclinepi.2019.12.023]

GRADE 2013
GRADE Handbook. GRADEpro Guideline Development Tool.

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

Hayden 2013
Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med 2013 Feb 19;158(4):280–6. [DOI: 10.7326/0003-4819-158-4-201302190-00009]

Huguet 2013
Huguet A, Hayden JA, Stinson J, McGrath PJ, Chambers CT, Tougas ME, et al. Judging the quality of evidence in reviews of prognostic factor research: adapting the GRADE framework. Systematic Reviews 2013 Sep 5;2(1):71. [DOI: 10.1186/2046-4053-2-71]

Islam 2020
Islam N, Salameh J-P, Leeflang MMG, Hooft L, McGrath TA, van der Pol CB, et al. Thoracic imaging tests for the diagnosis of COVID-19. Cochrane Database of Systematic Reviews 2020 Nov 26, Issue 11. Art. No.: CD013639. DOI: 10.1002/14651858.CD013639.pub3.

Islam 2021
Islam N, Ebrahimzadeh S, Salameh J-P, Kazi S, Fabiano N, Treanor L, et al. Thoracic imaging tests for the diagnosis of COVID-19. Cochrane Database of Systematic Reviews 2021 Mar 16, Issue 3. Art. No.: CD013639. DOI: 10.1002/14651858.CD013639.pub4.

Lee 2020
Lee J, Vali Y, Zafarmand M, Bossuyt P. Quality Assessment of Prognostic Accuracy Studies (QUAPAS): an extension of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool for systematic reviews of prognostic test accuracy studies. In: Abstracts of the 26th Cochrane Colloquium, Santiago, Chile, Cochrane Database of Systematic Reviews. Vol. 1. [DOI: 10.1002/14651858.CD201901]

Mahase 2020
Mahase E. Covid-19: What do we know about “long covid”? BMJ 2020 Jul 14;370:m2815. [DOI: 10.1136/bmj.m2815]

Parmar 1998
Parmar MKB, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Statistics in Medicine 1998;17(24):2815–34. [DOI: 10.1002/(SICI)1097-0258(19981230)17:24<2815:AID-SIM110>3.0.CO;2-8]

Prokop 2020
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 Apr 27;296(2):E97–104. [DOI: 10.1148/radiol.2020201473]

Redmond 2020
Redmond CE, Nicolaou S, Berger FH, Sheikh AM, Patlas MN. Emergency Radiology During the COVID-19 Pandemic: The Canadian Association of Radiologists Recommendations for Practice. Can Assoc Radiol J 2020 Nov 1;71(4):425–30. [DOI: 10.1177/0846537120930344]

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]

Riley 2019
Riley RD, Moons KGM, Snell KIE, Ensor J, Hooft L, Altman DG, et al. A guide to systematic review and meta-analysis of prognostic factor studies. BMJ 2019 Jan 30;364:k4597. [DOI: 10.1136/bmj.k4597]

Salameh 2020
Salameh J-P, Leeflang MMG, Hooft L, Islam N, McGrath TA, van der Pol CB, et al. Thoracic imaging tests for the diagnosis of COVID-19. Cochrane Database of Systematic Reviews 2020 Sep 30, Issue 9. Art. No.: CD013639. DOI: 10.1002/14651858.CD013639.pub2.

Salahshour 2021
Salahshour F, Mehrabinejad MM, Nassiri Toosi M et al. Clinical and chest CT features as a predictive tool for COVID-19 clinical progress: introducing a novel semi-quantitative scoring system. Eur Radiol 15 Jan 2021;31:5178–5188. [DOI: 10.1007/s00330-020-07623-w]

Schünemann 2020
Schünemann HJ, Santesso N, Vist GE, Cuello C, Lotfi T, Flottorp S, Davoli M, Mustafa R, Meerpohl JJ, Alonso-Coello P, Akl EA. Using GRADE in situations of emergencies and urgencies: certainty in evidence and recommendations matters during the COVID-19 pandemic, now more than ever and no matter what. Journal of Clinical Epidemiology 2020 Nov;127:202-207. [DOI: 10.1016/j.jclinepi.2020.05.030]

StataCorp 2019
Stata Statistical Software [Computer program]. 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 Nov 19, Issue 11. Art. No.: CD013787. DOI: 10.1002/14651858.CD013787.

Struyf 2021
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. Cochrane Database of Systematic Reviews 2021 Feb 23, Issue 2. Art. No.: CD013665. DOI: 10.1002/14651858.CD013665.pub2.

Tierney 2007
Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials 2007 Jun 7;8(1):16. [DOI: 10.1186/1745-6215-8-16]

Verbakel 2021
Verbakel JY, Rop LD, Stegeman I, Holtman GA, Ochodo EA, Yang B, et al. Accuracy of routine laboratory tests to predict mortality and deterioration to severe or critical COVID-19 in people with SARS-CoV-2 [Protocol]. Cochrane Database of Systematic Reviews 2021 Sep 17, Issue 9. Art. No.: CD015050. DOI: 10.1002/14651858.CD015050.

Whiting 2011
Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529–36. [DOI: 10.7326/0003-4819-155-8-201110180-00009]

WHO 2020
World Health Organization (WHO). WHO COVID-19 Case definition. www.who.int/publications-detail-redirect/WHO-2019-nCoV-Surveillance_Case_Definition-2022.1 (accessed 12 March 2022).

WHO 2021
World Health Organization (WHO). Timeline: WHO's COVID-19 response. www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline (accessed 10 March 2022).

Wolff 2019
Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med 2019 Jan 1;170(1):51–8. [DOI: 10.7326/M18-1376]

Yang 2020
Yang R, Li X, Liu H, Zhen Y, Zhang X, Xiong Q, et al. Chest CT Severity Score: An Imaging Tool for Assessing Severe COVID-19. Radiology: Cardiothoracic Imaging 2020 Apr 1;2(2):e200047. [DOI: 10.1148/ryct.2020200047]

Chapter 5. Conclusion

The Cochrane 'living' systematic review on diagnostic accuracy of imaging for COVID-19 has identified, based on the most up-to-date evidence (Chapter 3), that for diagnosing COVID-19 in suspected individuals: both chest computed tomography (CT) and ultrasound of the lungs are sensitive and moderately specific, while chest X-ray is moderately sensitive and moderately specific. This means that chest CT and ultrasound of the lungs may be useful for ruling out COVID-19 (given the high sensitivities) but may not perform well in distinguishing COVID-19 from other respiratory conditions (given the moderate specificities), in settings where these imaging modalities are available. When comparing imaging modalities, chest CT and ultrasound were determined to have similar sensitivities, both of which are higher than that of chest X-ray, while the specificities of all modalities were similar. Across the first three versions of the 'living' systematic review, the specificity of chest CT in individuals with suspected COVID-19 increased notably with each review version. Between the third version (Chapter 2) and the most recent version (Chapter 3), there were minor changes in accuracy observed for chest CT, chest X-ray, and ultrasound of the lungs.

In individuals who had an initial negative reverse transcriptase polymerase chain reaction (RT-PCR) result and a positive RT-PCR result on follow-up, the rate of positive chest CT imaging was 76%. This suggests that chest CT may be moderately effective for detecting COVID-19 in people with a false negative RT-PCR result. Among asymptomatic individuals with confirmed COVID-19, chest CT had a low sensitivity for detecting the disease, which suggests that chest CT imaging would not be useful as a screening test.

With respect to potential sources of heterogeneity, overall, reference standard conduct and the definition for a positive index test did not appear to impact diagnostic accuracy. It was also determined that when using formal scoring systems to assess chest CT images for COVID-

19, as the threshold for a positive index test is increased, sensitivity decreases, and specificity increases.

With each version of this 'living' systematic review, the review methodologies have been refined. The third version (Chapter 2) excluded studies with case-control designs due to the bias involved in such studies, as well as studies that reported an overview of index test findings in participants without explicitly classifying index tests as positive or negative. The fourth version (Chapter 3) then also excluded studies that excluded participants with normal index tests during participant recruitment due to the associated selection bias, studies that did not report the definition used by radiologists to classify a positive index test, studies that did not use RT-PCR as part of the reference standard, and studies that included imaging as part of the reference standard due to incorporation bias.

This increasingly restrictive eligibility criteria, contributed to the quality of included studies being improved in the third version of the review (Chapter 2) and again in the fourth version (Chapter 3). The risk of bias assessments showed that the fourth version (Chapter 3), compared to the third version (Chapter 2) had a higher number and higher proportion of studies with a low risk of bias. This increasingly restrictive eligibility criteria and improvement in the quality of included studies likely also influenced the results of the review. Between the second version and the third version (Chapter 2), specificity of chest CT in individuals with suspected COVID-19 increased notably; between the third version (Chapter 2) and the fourth version (Chapter 3), specificity of chest CT remained similar but with increased precision.

The findings of these reviews should be interpreted carefully, due to several limitations. First, about half of all studies in the most recent review version (Chapter 3) had a high or unclear risk of bias. Thus, the results of these primary studies could have been inaccurate due to

potentially biased study methods, and this could have led to biased accuracy estimates generated in the meta-analyses. However, the number and proportion of studies with a high or unclear risk of bias decreased between the third version (Chapter 2) and the most recent version (Chapter 3). Second, while RT-PCR is considered the best available diagnostic test, it is not always sensitive, and thus has limitations when used as a reference standard. RT-PCR sensitivity can depend on multiple factors, including 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). It is possible that chest CT may be more sensitive than the reference standard in some situations, and some patients identified as having a false-positive diagnosis on CT may have been missed by the RT-PCR test. However, in both the third version (Chapter 2) and the fourth version (Chapter 3) of the review, the meta-regression analyses for chest CT studies that performed RT-PCR testing at least twice for all 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 initial negative results, did not identify different accuracy estimates between the groups. Third, it was not possible to evaluate the impact of sex, gender, race, and ethnicity on accuracy estimates in this review series, due insufficient available data in primary studies. However, based on country of study origin, the findings in these reviews are likely most applicable to individuals who are Caucasian. Fourth, while investigations of heterogeneity for diagnostic accuracy of chest X-ray and ultrasound were conducted for the first time in the most recent review version (Chapter 3), there was a low number of studies included. Due to this small sample size, the accuracy estimates had low precision, and the analyses may have been underpowered to detect a difference between groups. Finally, the comparisons of diagnostic

accuracy between different imaging modalities were mainly based on non-comparative design studies, each of which evaluated one type of imaging test. Thus, it was not possible to directly compare the performances of different modalities in the same population.

Leading a 'living' systematic review series during the COVID-19 pandemic, alongside the real-time generation of primary research on the topic, involved many unique processes. The evidence base has grown and evolved rapidly since the start of the COVID-19 pandemic, and as such, this 'living' systematic review has been constantly updated, until the completion of the fourth and most recent review version (Chapter 3). The review methodologies were also refined between review versions. With each version of the review, a higher number of studies were included, and inclusion was limited to better quality studies, leading to periodically updated findings supported by increasing precision and primary study quality. Furthermore, as new studies became available for inclusion, additional sources of heterogeneity could be investigated, and additional primary and secondary objectives could be formally assessed.

All decisions to revise review methodologies, and address additional review objectives, were guided and approved through discussion with the Cochrane COVID-19 Diagnostic Test Accuracy Group. Systematic review experts, clinicians, and representatives of the World Health Organization in this group were consulted to determine whether these changes would be feasible to implement, and relevant in the practical application of the review findings, within the context of the evidence base and stage of the pandemic at that time.

With regards to future research, there is no immediate plan to conduct another version of this 'living' systematic review. However, given the recent changes in the proportion of the global population that has been vaccinated, as well as the changes in circulating variants of COVID-19, there may be a need to update this review in the future. The fourth version of this review

included studies published until February 2021, before new variants began circulating; thus, the most recent review was based only on the original (alpha) strain of COVID-19. In addition to this, as the COVID-19 pandemic has evolved, the prognostic value of imaging for COVID-19 has become of interest. Thus, a Cochrane systematic review protocol has been developed to evaluate the prognostic accuracy of imaging findings for predicting morbidity and mortality outcomes in individuals with COVID-19 (Chapter 4). The process of conducting this systematic review is planned to begin once the protocol has received final approval for publication.

Overall, the series of Cochrane reviews comprising this work has and will continue to provide up-to-date evidence on the diagnostic and prognostic value of imaging for COVID-19. This evidence could help inform decisions regarding diagnosis, screening, isolation, and patient management during the COVID-19 pandemic.

Appendix

Appendix 1.1 Glossary

- **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
- **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
- **Index test:** the test that is being assessed (the index test will often be a new test)
- **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 specialized 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:** specialized 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 2.1 Details on target condition and index test

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, stomachache, 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. This review focused on people suspected of having COVID-19 who had thoracic imaging as part of their evaluation or care.

Index tests

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, this review 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. This review considered any and all variations of chest radiography protocols that evaluated the lungs. 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) were not included.

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).

All possible technical parameters (e.g. type of probe, transducer frequency, use of contrast) were considered. 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 tests

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.
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.
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 Commission 2020).

Alternative tests

Other Cochrane diagnostic test accuracy 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.

Appendix 2.2 Protocol deviations

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 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 2.3). 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 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; and 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.

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.

Appendix 2.3 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 2020; 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 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 more 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 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 possible causes of this

not match the review question?	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 2.4 Search details

Search strategy

1. Living search from the University of Bern

The COVID-19 living search results of the Institute of Social and Preventive Medicine (ISPM) at the University of Bern were used. This search includes PubMed, Embase and preprints indexed in bioRxiv and medRxiv databases.

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

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

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

Searches undertaken by Cochrane to develop the Cochrane COVID-19 Study Register were used. 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. Search strategies were designed for maximum sensitivity, to retrieve all human studies on COVID-19. Language limits were not applied.

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 screened 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 were included and deduplicated these against the Cochrane COVID-19 Study Register.
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-

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 the review 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 2.5 Additional figures and tables

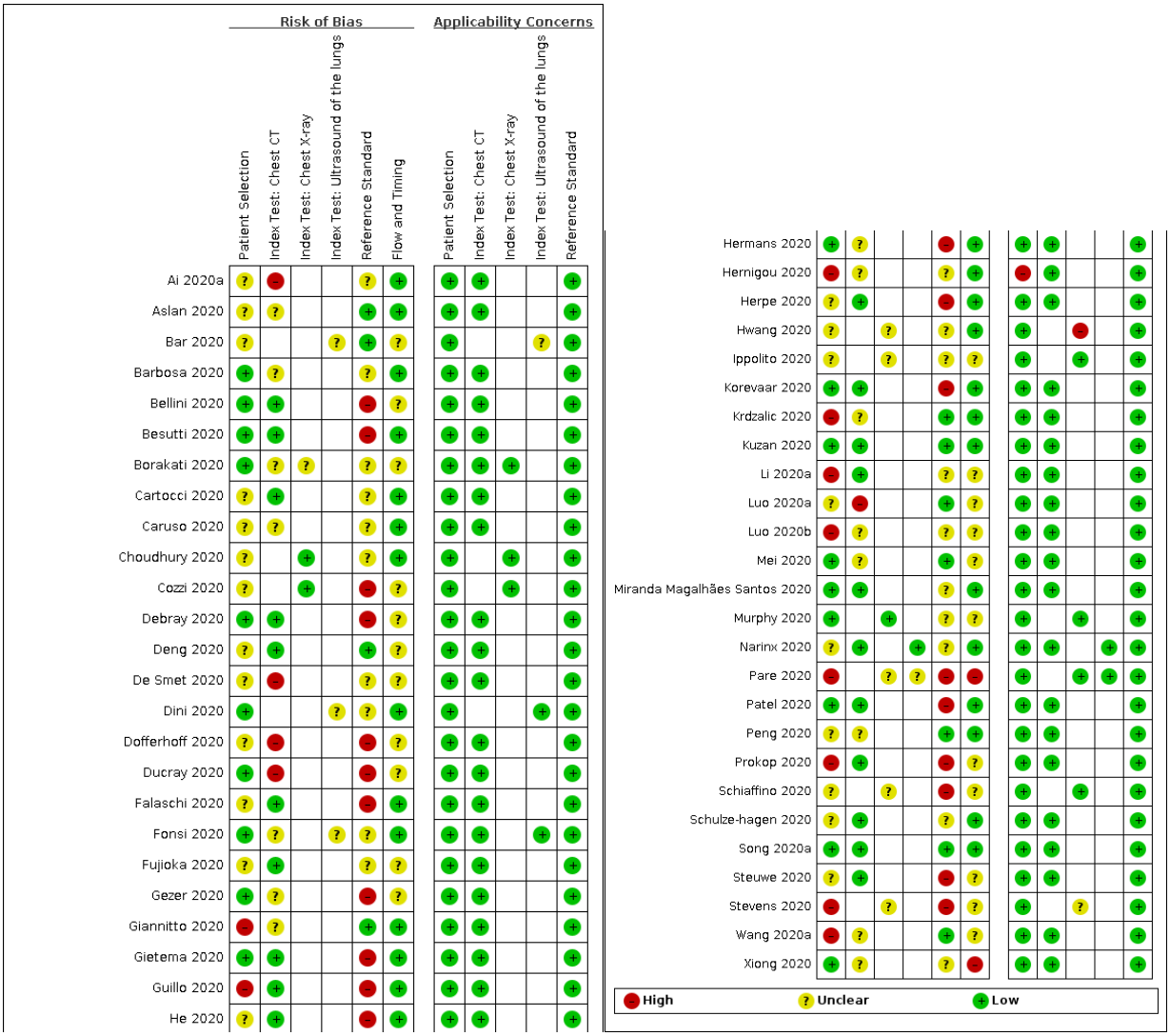


Figure 2.A1. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study.

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]

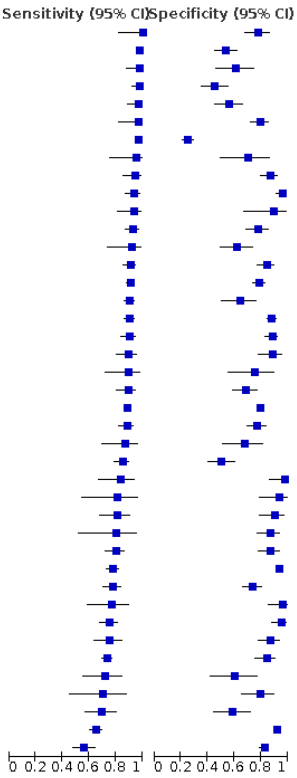


Figure 2.A2. Forest plot of chest CT in suspected cases.

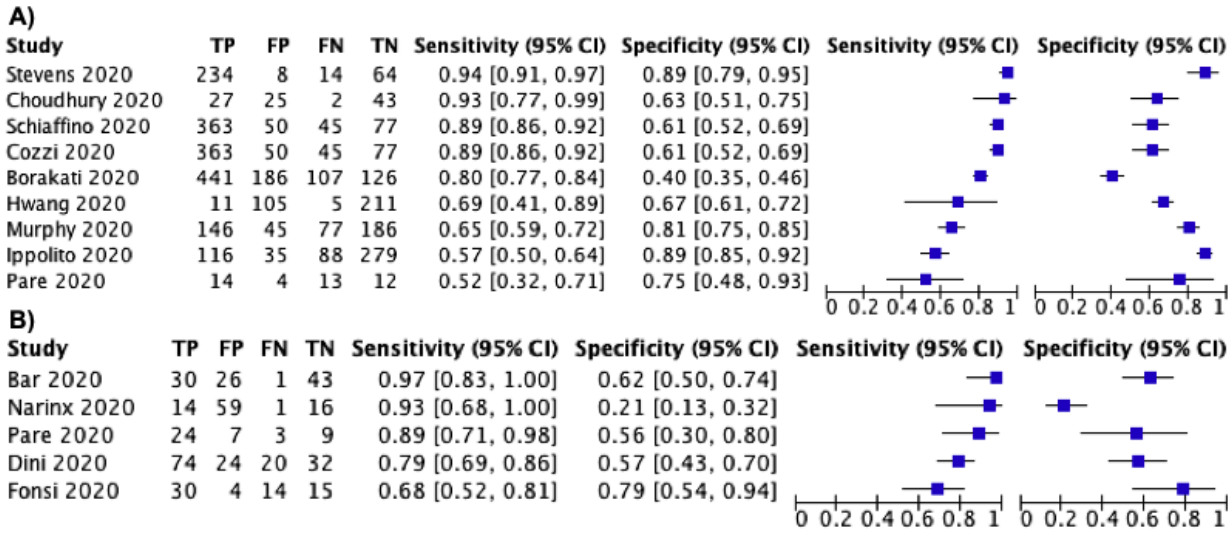


Figure 2.A3. Forest plot of A) chest X-ray and B) ultrasound in suspected cases.

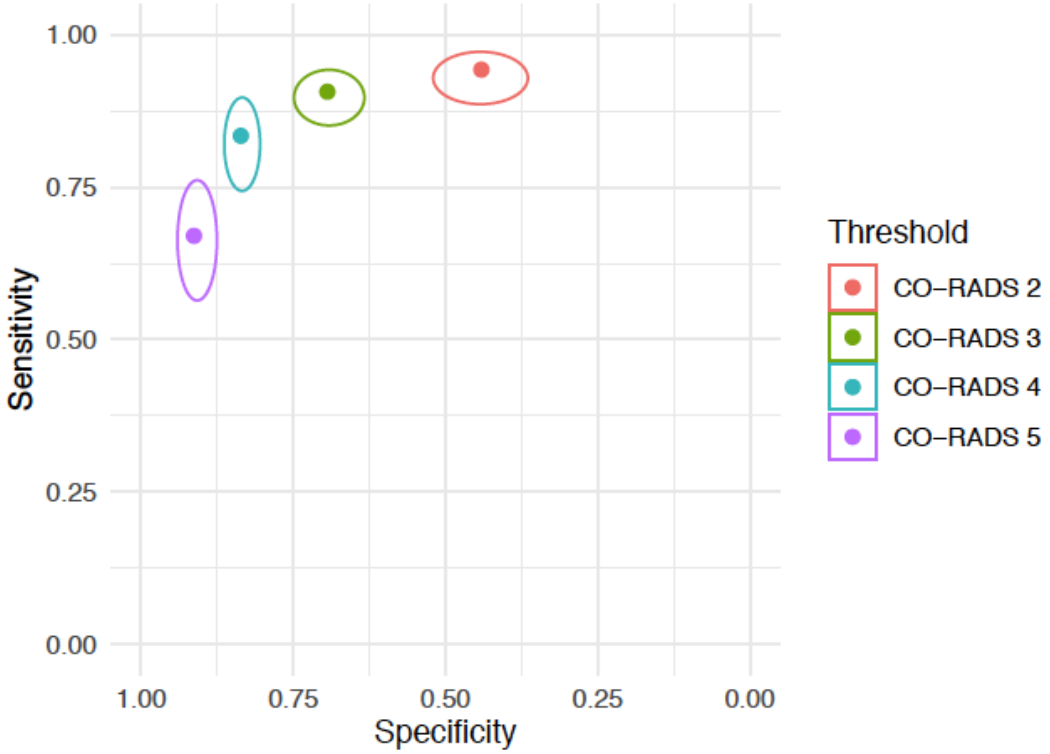


Figure 2.A4. Pooled sensitivity and specificity estimates and 95% confidence intervals at varying CO-RADS thresholds: CO-RADS 2 (n=7), CO-RADS 3 (n=9), CO-RADS 4 (n=7), and CO-RADS 5 (n=7).

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]

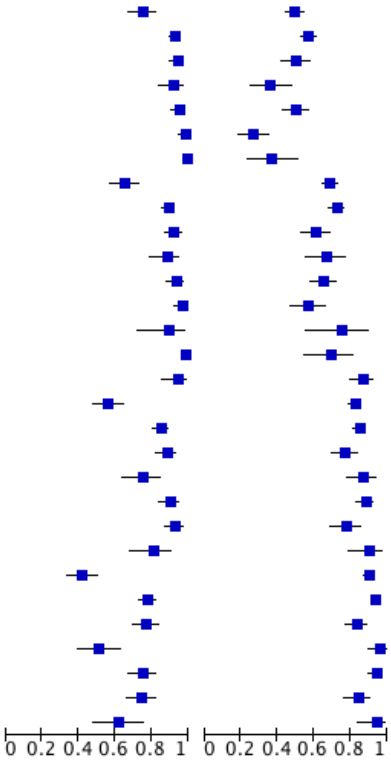


Figure A5. Forest plot of chest CT studies in suspected cases that used the CO-RADS scoring system. Threshold (i.e. 2, 3 4, or 5) indicated in parentheses ahead of Study ID. Grouped by threshold.

Appendix 3.1 Details on target condition and index test

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 and can have a wide variety of symptoms, including fever, sore throat, diarrhoea, dyspnoea, headache, chest pain, stomach-ache, nausea, loss of taste, loss of smell, myalgia (muscle pain), 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.

Index tests

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, this review 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. This review considered any and all variations of chest radiography protocols that evaluated the lungs. 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) were not included.

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).

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

Clinical pathway

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 tests

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 additional tests such as RT-PCR.
2. 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 diagnostic accuracy. 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.
4. Thoracic imaging used to detect COVID-19 in asymptomatic patients.

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 organizations 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 categorize patients as low risk, moderate to high risk, and proven COVID-19 cases (China National Health Commission 2020).

Alternative tests

Other Cochrane diagnostic test accuracy 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; Dinnes 2021).
5. Electronic and animal noses (Leeflang 2021).

Appendix 3.2 Protocol deviations

Inclusion criteria

The exclusion of: (1) case-control studies; (2) 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; (3) studies that excluded participants with normal index test results; (4) studies that did not clearly report the definition used by radiologists to classify a positive index test; (5) studies that did not include RT-PCR as a part of the reference standard; and (6) studies that included imaging as a part of the reference standard 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 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 3.3). 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 the review authors.

For the objective on rate of positive imaging in repeat RT-PCR positive results, an additional bias assessment was conducted to determine if there was participant selection bias. If all participants who had a negative RT-PCR result underwent a repeat RT-PCR test, this was considered ‘low risk’. If some but not all participants who had a negative RT-PCR result underwent a repeat RT-PCR test due to an alternate indication of infection (i.e. symptoms), this was considered ‘high risk’. If some but not all participants who had a negative RT-PCR result underwent a repeat RT-PCR test, but it is unclear what this selection was based on, this was considered ‘unclear risk’.

Primary objective

An additional primary objective, on ‘the accuracy of imaging in asymptomatic participants’ was evaluated in this review, which is a medication from the original protocol and previous versions of the review.

Secondary objectives

We did not address the planned secondary objective, ‘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’, due to insufficient available data (McInnes 2020).

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 there were few studies that 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.

Appendix 3.3 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 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 had to be confirmed by 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), alone or in combination with other criteria as follows. 1. positive on WHO criteria for COVID-19 2. positive on China CDC criteria for COVID-19 3. positive serology for SARS-CoV-2 antibodies in addition to consistent symptomatology 4. positive on study-specific list of criteria for COVID-19 (e.g., symptoms, imaging findings, other tests, infected contacts) A negative diagnosis for COVID-19 had to be confirmed by the following. 1. suspected COVID-19 with negative RT-PCR test results, whether tested once or more than once 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, as long as RT-PCR is included as a part of the reference standard, and imaging is not included as a part of the reference standard. 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 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 more or less equally suspected of having COVID-19 and were included and if the numbers in the flow chart show not too many excluded participant (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.
For studies included for rate of positive imaging in repeat RT-PCR+ results objective: Could the selection of patients have introduced bias?	YES: if only some (and not all) included participants underwent repeat RT-PCR testing, and it is clear that a non-consecutive or non-random selection procedure was employed; e.g. based on symptom status, or based on index test findings NO: if participants who underwent repeat RT-PCR testing were selected in a random or consecutive manner from the total included participants UNCLEAR: if the selection method was unclearly reported.
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 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 not match the review question?	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 possible 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

Abbreviations: **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 3.4 Additional figures and tables

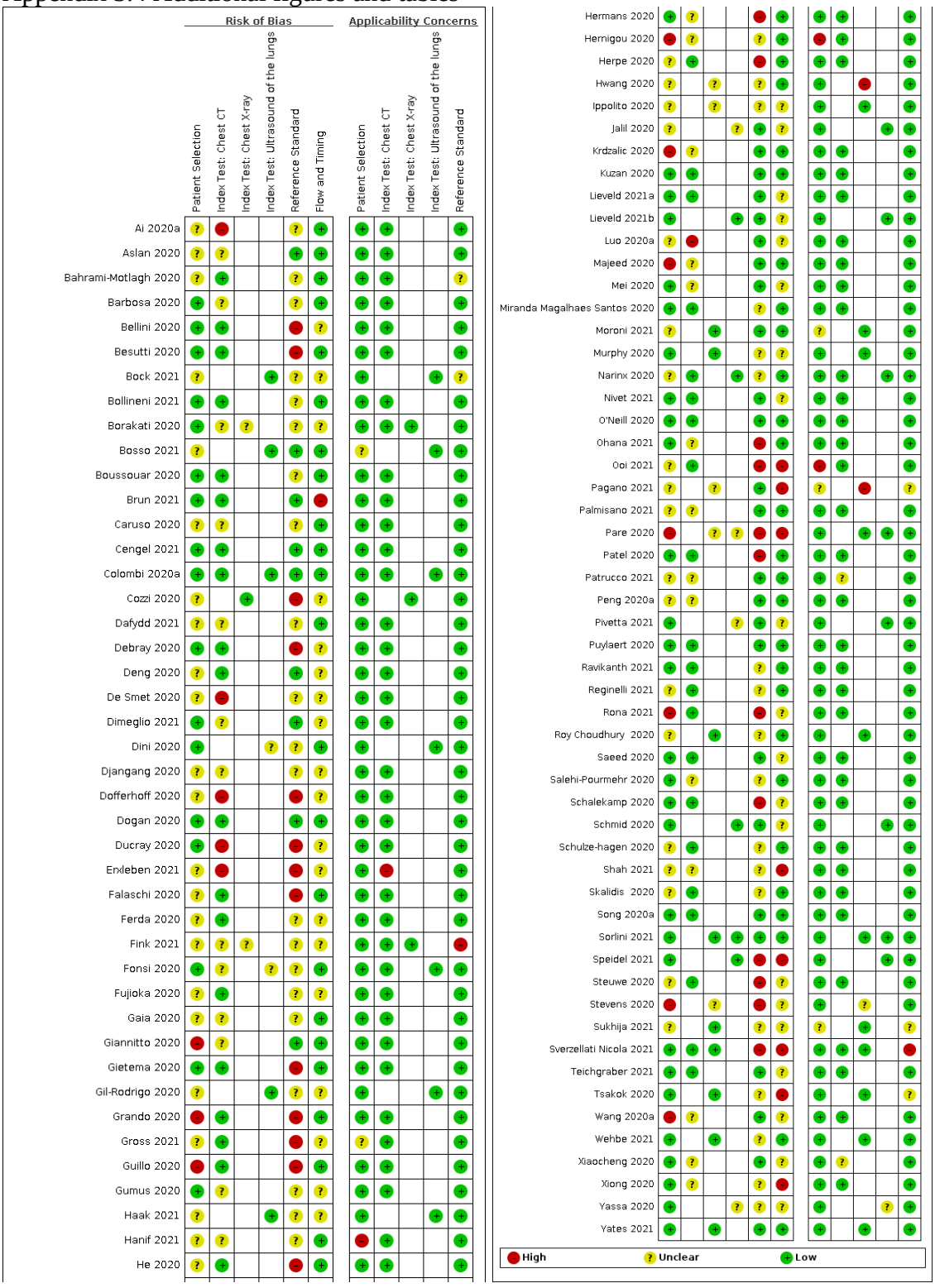


Figure 3.A1. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study.

Study	TP	FP	FN	TN	Formal scoring system 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]
Bollineni 2021	144	69	0	27	-	1.00 [0.97, 1.00]	0.28 [0.19, 0.38]
Palmisano 2021	95	11	1	36	-	0.99 [0.94, 1.00]	0.77 [0.62, 0.88]
Shah 2021	146	18	2	2	-	0.99 [0.95, 1.00]	0.10 [0.01, 0.32]
Deng 2020	423	71	10	83	-	0.98 [0.96, 0.99]	0.54 [0.46, 0.62]
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]
Bahrami-Motlagh 2020	86	47	3	27	-	0.97 [0.90, 0.99]	0.36 [0.26, 0.48]
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]
Sverzellati Nicola 2021	181	39	9	19	-	0.95 [0.91, 0.98]	0.33 [0.21, 0.46]
Xiong 2020	19	8	1	19	-	0.95 [0.75, 1.00]	0.70 [0.50, 0.86]
Reginelli 2021	309	22	19	28	-	0.94 [0.91, 0.96]	0.56 [0.41, 0.70]
Schulze-hagen 2020	65	16	4	106	CO-RADS 3	0.94 [0.86, 0.98]	0.87 [0.80, 0.92]
Ravikanth 2021	453	31	28	100	-	0.94 [0.92, 0.96]	0.76 [0.68, 0.83]
Ferda 2020	30	15	2	263	-	0.94 [0.79, 0.99]	0.95 [0.91, 0.97]
Fonsi 2020	41	2	3	17	-	0.93 [0.81, 0.99]	0.89 [0.67, 0.99]
Nivet 2021	225	43	19	226	-	0.92 [0.88, 0.95]	0.84 [0.79, 0.88]
Barbosa 2020	23	25	2	41	RSNA 4	0.92 [0.74, 0.99]	0.62 [0.49, 0.74]
Colombi 2020a	313	49	28	96	-	0.92 [0.88, 0.94]	0.66 [0.58, 0.74]
Dimeglio 2021	104	30	10	167	-	0.91 [0.84, 0.96]	0.85 [0.79, 0.89]
Gaia 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	CO-RADS 4	0.90 [0.84, 0.95]	0.88 [0.83, 0.92]
Gross 2021	18	7	2	69	CO-RADS 4	0.90 [0.68, 0.99]	0.91 [0.82, 0.96]
Lieveld 2021a	210	65	25	441	CO-RADS 4	0.89 [0.85, 0.93]	0.87 [0.84, 0.90]
Krdzalic 2020	25	7	3	21	CO-RADS 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]
Grando 2020	76	4	10	69	RSNA 4	0.88 [0.80, 0.94]	0.95 [0.87, 0.98]
Dofferhoff 2020	136	36	18	122	CO-RADS 4	0.88 [0.82, 0.93]	0.77 [0.70, 0.84]
Boussouar 2020	480	124	65	394	-	0.88 [0.85, 0.91]	0.76 [0.72, 0.80]
Luo 2020a	26	14	4	29	-	0.87 [0.69, 0.96]	0.67 [0.51, 0.81]
Schalekamp 2020	460	101	76	433	CO-RADS 4	0.86 [0.83, 0.89]	0.81 [0.78, 0.84]
Brun 2021	148	23	26	110	-	0.85 [0.79, 0.90]	0.83 [0.75, 0.89]
Borakati 2020	162	55	29	56	-	0.85 [0.79, 0.90]	0.50 [0.41, 0.60]
Teichgraber 2021	11	8	2	144	-	0.85 [0.55, 0.98]	0.95 [0.90, 0.98]
Skalidis 2020	55	18	10	27	-	0.85 [0.74, 0.92]	0.60 [0.44, 0.74]
Ohana 2021	919	148	172	955	-	0.84 [0.82, 0.86]	0.87 [0.84, 0.89]
Djangang 2020	79	24	15	0	-	0.84 [0.75, 0.91]	0.00 [0.00, 0.14]
Miranda Magalhaes Santos 2020	30	1	6	38	RSNA 4	0.83 [0.67, 0.94]	0.97 [0.87, 1.00]
Cengel 2021	330	90	66	48	-	0.83 [0.79, 0.87]	0.35 [0.27, 0.43]
Erxleben 2021	28	52	6	183	-	0.82 [0.65, 0.93]	0.78 [0.72, 0.83]
O'Neill 2020	149	18	33	45	-	0.82 [0.75, 0.87]	0.71 [0.59, 0.82]
Hernigou 2020	13	2	3	29	-	0.81 [0.54, 0.96]	0.94 [0.79, 0.99]
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	CO-RADS 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]
Fujioka 2020	57	10	19	68	CO-RADS 4	0.75 [0.64, 0.84]	0.87 [0.78, 0.94]
Debray 2020	119	4	40	78	-	0.75 [0.67, 0.81]	0.95 [0.88, 0.99]
Besutti 2020	438	16	158	84	-	0.73 [0.70, 0.77]	0.84 [0.75, 0.91]
Peng 2020a	28	13	11	20	-	0.72 [0.55, 0.85]	0.61 [0.42, 0.77]
Xiaocheng 2020	7	13	3	68	-	0.70 [0.35, 0.93]	0.84 [0.74, 0.91]
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]
Saeed 2020	44	6	20	14	-	0.69 [0.56, 0.80]	0.70 [0.46, 0.88]
Mei 2020	274	39	145	447	-	0.65 [0.61, 0.70]	0.92 [0.89, 0.94]
Salehi-Pourmehr 2020	129	84	72	283	-	0.64 [0.57, 0.71]	0.77 [0.72, 0.81]
Fink 2021	45	1	27	146	-	0.63 [0.50, 0.74]	0.99 [0.96, 1.00]
Patrucco 2021	11	4	7	24	RSNA 4	0.61 [0.36, 0.83]	0.86 [0.67, 0.96]
Majeed 2020	40	37	29	101	-	0.58 [0.45, 0.70]	0.73 [0.65, 0.80]
Bellini 2020	80	76	62	354	CO-RADS 4	0.56 [0.48, 0.65]	0.82 [0.78, 0.86]
Rona 2021	23	11	25	48	-	0.48 [0.33, 0.63]	0.81 [0.69, 0.90]
Hanif 2021	35	3	43	13	-	0.45 [0.34, 0.57]	0.81 [0.54, 0.96]
Dogan 2020	150	91	286	264	RSNA 4	0.34 [0.30, 0.39]	0.74 [0.69, 0.79]

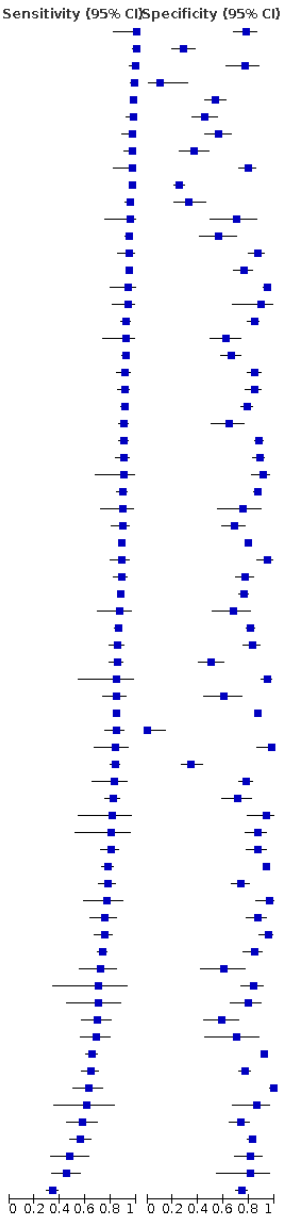
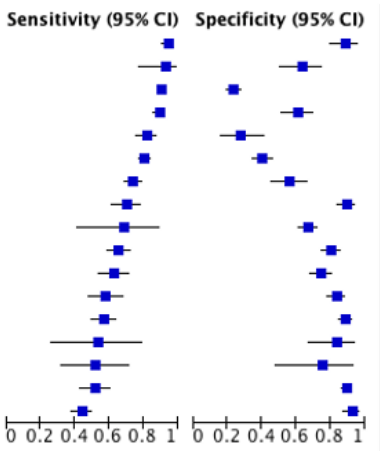


Figure 3.A2. Forest plot of chest CT in suspected cases.

A

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Stevens 2020	234	8	14	64	0.94 [0.91, 0.97]	0.89 [0.79, 0.95]
Roy Choudhury 2020	27	25	2	43	0.93 [0.77, 0.99]	0.63 [0.51, 0.75]
Pagano 2021	2088	288	223	89	0.90 [0.89, 0.92]	0.24 [0.19, 0.28]
Cozzi 2020	363	50	45	77	0.89 [0.86, 0.92]	0.61 [0.52, 0.69]
Sverzellati Nicola 2021	155	42	35	16	0.82 [0.75, 0.87]	0.28 [0.17, 0.41]
Borakati 2020	441	186	107	126	0.80 [0.77, 0.84]	0.40 [0.35, 0.46]
Sorlini 2021	207	41	73	52	0.74 [0.68, 0.79]	0.56 [0.45, 0.66]
Wehbe 2021	94	18	40	148	0.70 [0.62, 0.78]	0.89 [0.83, 0.93]
Hwang 2020	11	105	5	211	0.69 [0.41, 0.89]	0.67 [0.61, 0.72]
Murphy 2020	146	45	77	186	0.65 [0.59, 0.72]	0.81 [0.75, 0.85]
Tsakok 2020	83	54	49	158	0.63 [0.54, 0.71]	0.75 [0.68, 0.80]
Moroni 2021	58	38	42	189	0.58 [0.48, 0.68]	0.83 [0.78, 0.88]
Ippolito 2020	116	35	88	279	0.57 [0.50, 0.64]	0.89 [0.85, 0.92]
Fink 2021	8	6	7	30	0.53 [0.27, 0.79]	0.83 [0.67, 0.94]
Pare 2020	14	4	13	12	0.52 [0.32, 0.71]	0.75 [0.48, 0.93]
Yates 2021	74	47	69	392	0.52 [0.43, 0.60]	0.89 [0.86, 0.92]
Sukhija 2021	130	12	165	150	0.44 [0.38, 0.50]	0.93 [0.87, 0.96]



B

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Pivetta 2021	101	6	6	115	0.94 [0.88, 0.98]	0.95 [0.90, 0.98]
Colombi 2020a	319	103	22	42	0.94 [0.90, 0.96]	0.29 [0.22, 0.37]
Narinx 2020	14	59	1	16	0.93 [0.68, 1.00]	0.21 [0.13, 0.32]
Gil-Rodrigo 2020	25	5	2	26	0.93 [0.76, 0.99]	0.84 [0.66, 0.95]
Sorlini 2021	264	34	23	63	0.92 [0.88, 0.95]	0.65 [0.55, 0.74]
Lieveld 2021b	79	29	7	71	0.92 [0.84, 0.97]	0.71 [0.61, 0.80]
Jalil 2020	33	5	3	28	0.92 [0.78, 0.98]	0.85 [0.68, 0.95]
Bock 2021	11	25	1	46	0.92 [0.62, 1.00]	0.65 [0.53, 0.76]
Speidel 2021	10	9	1	29	0.91 [0.59, 1.00]	0.76 [0.60, 0.89]
Haak 2021	24	30	3	43	0.89 [0.71, 0.98]	0.59 [0.47, 0.70]
Pare 2020	24	7	3	9	0.89 [0.71, 0.98]	0.56 [0.30, 0.80]
Dini 2020	74	24	20	32	0.79 [0.69, 0.86]	0.57 [0.43, 0.70]
Schmid 2020	30	22	9	74	0.77 [0.61, 0.89]	0.77 [0.67, 0.85]
Yassa 2020	17	16	6	257	0.74 [0.52, 0.90]	0.94 [0.91, 0.97]
Bosso 2021	19	3	7	24	0.73 [0.52, 0.88]	0.89 [0.71, 0.98]

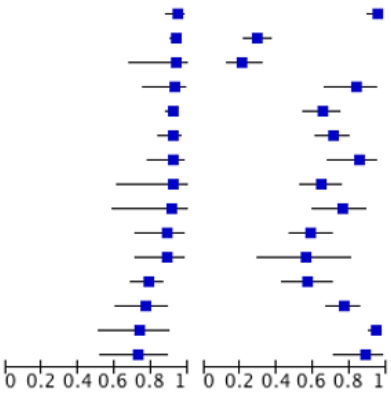
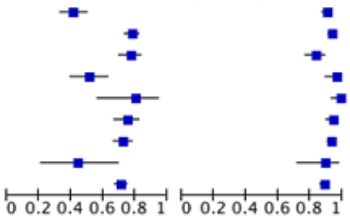


Figure 3.A3. Forest plot of A) chest X-ray and B) ultrasound of the lungs in suspected cases.

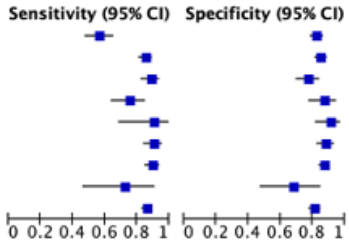
A

Study	TP	FP	FN	TN	Publication Status	Sensitivity (95% CI)	Specificity (95% CI)
Bellini 2020	59	41	83	389	Published	0.42 [0.33, 0.50]	0.90 [0.87, 0.93]
De Smet 2020	279	33	79	468	Published	0.78 [0.73, 0.82]	0.93 [0.91, 0.95]
Dofferhoff 2020	119	26	35	132	Published	0.77 [0.70, 0.84]	0.84 [0.77, 0.89]
Fujioka 2020	39	3	37	75	Published	0.51 [0.40, 0.63]	0.96 [0.89, 0.99]
Gross 2021	16	1	4	75	Published	0.80 [0.56, 0.94]	0.99 [0.93, 1.00]
Hermans 2020	100	11	33	175	Published	0.75 [0.67, 0.82]	0.94 [0.90, 0.97]
Lieveld 2021a	170	35	65	471	Published	0.72 [0.66, 0.78]	0.93 [0.91, 0.95]
Patrucco 2021	8	3	10	25	Published	0.44 [0.22, 0.69]	0.89 [0.72, 0.98]
Schalekamp 2020	381	60	155	474	Published	0.71 [0.67, 0.75]	0.89 [0.86, 0.91]



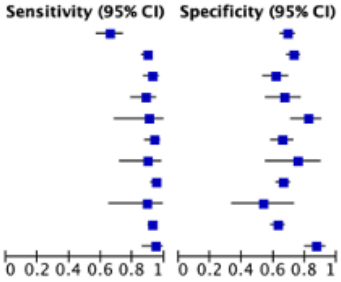
B

Study	TP	FP	FN	TN	Publication Status	Sensitivity (95% CI)	Specificity (95% CI)
Bellini 2020	80	76	62	354	Published	0.56 [0.48, 0.65]	0.82 [0.78, 0.86]
De Smet 2020	304	76	54	425	Published	0.85 [0.81, 0.88]	0.85 [0.81, 0.88]
Dofferhoff 2020	136	36	18	122	Published	0.88 [0.82, 0.93]	0.77 [0.70, 0.84]
Fujioka 2020	57	10	19	68	Published	0.75 [0.64, 0.84]	0.87 [0.78, 0.94]
Gross 2021	18	7	2	69	Published	0.90 [0.68, 0.99]	0.91 [0.82, 0.96]
Hermans 2020	120	22	13	164	Published	0.90 [0.84, 0.95]	0.88 [0.83, 0.92]
Lieveld 2021a	210	65	25	441	Published	0.89 [0.85, 0.93]	0.87 [0.84, 0.90]
Patrucco 2021	13	9	5	19	Published	0.72 [0.47, 0.90]	0.68 [0.48, 0.84]
Schalekamp 2020	460	101	76	433	Published	0.86 [0.83, 0.89]	0.81 [0.78, 0.84]



C

Study	TP	FP	FN	TN	Publication Status	Sensitivity (95% CI)	Specificity (95% CI)
Bellini 2020	93	134	49	296	Published	0.65 [0.57, 0.73]	0.69 [0.64, 0.73]
De Smet 2020	319	138	39	363	Published	0.89 [0.85, 0.92]	0.72 [0.68, 0.76]
Dofferhoff 2020	142	61	12	97	Published	0.92 [0.87, 0.96]	0.61 [0.53, 0.69]
Fujioka 2020	67	26	9	52	Published	0.88 [0.79, 0.94]	0.67 [0.55, 0.77]
Gross 2021	18	14	2	62	Published	0.90 [0.68, 0.99]	0.82 [0.71, 0.90]
Hermans 2020	124	64	9	122	Published	0.93 [0.88, 0.97]	0.66 [0.58, 0.72]
Krdzalic 2020	25	7	3	21	Published	0.89 [0.72, 0.98]	0.75 [0.55, 0.89]
Lieveld 2021a	223	172	12	334	Published	0.95 [0.91, 0.97]	0.66 [0.62, 0.70]
Patrucco 2021	16	13	2	15	Published	0.89 [0.65, 0.99]	0.54 [0.34, 0.72]
Schalekamp 2020	495	200	41	334	Published	0.92 [0.90, 0.94]	0.63 [0.58, 0.67]
Schulze-hagen 2020	65	16	4	106	Published	0.94 [0.86, 0.98]	0.87 [0.80, 0.92]



D

Study	TP	FP	FN	TN	Publication Status	Sensitivity (95% CI)	Specificity (95% CI)
Bellini 2020	107	216	35	214	Published	0.75 [0.67, 0.82]	0.50 [0.45, 0.55]
De Smet 2020	331	215	27	286	Published	0.92 [0.89, 0.95]	0.57 [0.53, 0.61]
Dofferhoff 2020	145	79	9	79	Published	0.94 [0.89, 0.97]	0.50 [0.42, 0.58]
Fujioka 2020	70	50	6	28	Published	0.92 [0.84, 0.97]	0.36 [0.25, 0.48]
Gross 2021	19	33	1	43	Published	0.95 [0.75, 1.00]	0.57 [0.45, 0.68]
Hermans 2020	127	93	6	93	Published	0.95 [0.90, 0.98]	0.50 [0.43, 0.57]
Lieveld 2021a	227	271	8	235	Published	0.97 [0.93, 0.99]	0.46 [0.42, 0.51]
Patrucco 2021	18	25	0	3	Published	1.00 [0.81, 1.00]	0.11 [0.02, 0.28]
Schalekamp 2020	511	324	24	210	Published	0.96 [0.93, 0.97]	0.39 [0.35, 0.44]

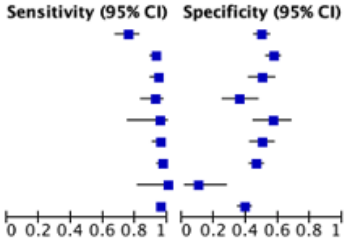
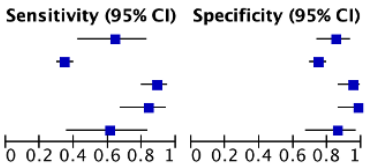


Figure 3.A4. Forest plot of chest CT studies in suspected cases that used the CO-RADS scoring system at varying thresholds: A) CO-RADS 5, B) CO-RADS 4, C) CO-RADS 3, and D) CO-RADS 2.

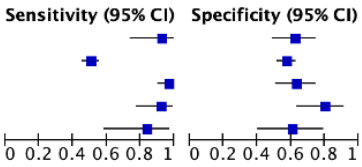
A

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Barbosa 2020	16	10	9	56	0.64 [0.43, 0.82]	0.85 [0.74, 0.92]
Dogan 2020	150	91	286	264	0.34 [0.30, 0.39]	0.74 [0.69, 0.79]
Grando 2020	76	4	10	69	0.88 [0.80, 0.94]	0.95 [0.87, 0.98]
Miranda Magalhaes Santos 2020	30	1	6	38	0.83 [0.67, 0.94]	0.97 [0.87, 1.00]
Patrucco 2021	11	4	7	24	0.61 [0.36, 0.83]	0.86 [0.67, 0.96]



B

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Barbosa 2020	23	25	2	41	0.92 [0.74, 0.99]	0.62 [0.49, 0.74]
Dogan 2020	220	152	216	203	0.50 [0.46, 0.55]	0.57 [0.52, 0.62]
Grando 2020	83	27	3	46	0.97 [0.90, 0.99]	0.63 [0.51, 0.74]
Miranda Magalhaes Santos 2020	33	8	3	31	0.92 [0.78, 0.98]	0.79 [0.64, 0.91]
Patrucco 2021	15	11	3	17	0.83 [0.59, 0.96]	0.61 [0.41, 0.78]



C

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Dogan 2020	241	285	195	70	0.55 [0.50, 0.60]	0.20 [0.16, 0.24]
Grando 2020	83	44	3	29	0.97 [0.90, 0.99]	0.40 [0.28, 0.52]
Miranda Magalhaes Santos 2020	33	22	3	17	0.92 [0.78, 0.98]	0.44 [0.28, 0.60]
Patrucco 2021	18	25	0	3	1.00 [0.81, 1.00]	0.11 [0.02, 0.28]

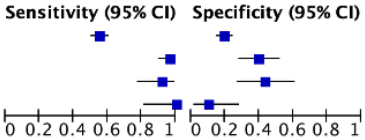


Figure 3.A5. Forest plot of chest CT studies in suspected cases that used the RSNA scoring system at varying thresholds: A) RSNA 4, B) RSNA 3, and C) RSNA 2.

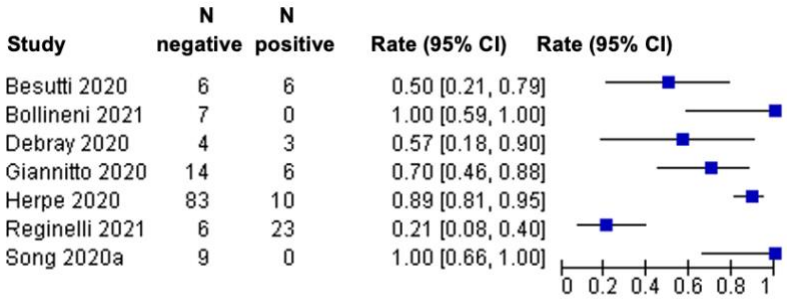


Figure 3.A6. Forest plot of positive chest CT imaging in participants with repeat RT-PCR positive results where initial RT-PCR was negative. N positive=number of participants with an initial negative RT-PCR test and a positive result on repeat RT-PCR testing, who had chest CT imaging positive for COVID-19. N negative=number of participants with an initial negative RT-PCR test result and a positive result on repeat RT-PCR testing, who had chest CT imaging negative for COVID-19. Rate=N positive / (N positive + N negative).

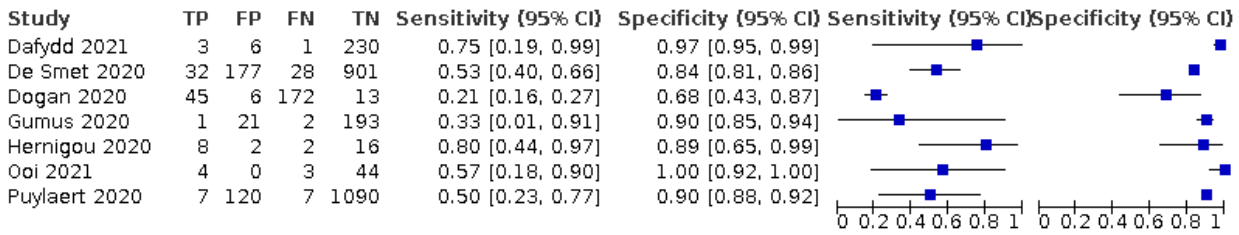


Figure 3.A7. Forest plot of positive chest CT imaging in asymptomatic participants.

Appendix 4.1 Details on index tests

Chest CT

Chest CT refers to the acquisition of images of the chest using computed tomography. This review will consider all variations of CT imaging protocols, except for those that specifically evaluated the coronary arteries or the heart, and did not include the entire lungs in the field of view. These variations include, but are 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, supine) view or posterior-anterior (PA, prone) view. This review will consider all variations of chest radiography protocols that evaluated the lungs, except for those 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 (POCUS), done at the bedside by a physician, as well as ‘consultative’ ultrasound, which is done by a technologist and subsequently interpreted by a physician (typically a radiologist). This review will consider all possible technical parameters (e.g. type of probe, transducer frequency, use of contrast), but will not consider ultrasound done with the intended purpose of evaluating only the heart or vessels of the chest.

References

China National Health Commission 2020
China National Health Commission. 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).

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]

Dinnes 2021
Dinnes J, Deeks JJ, Berhane S, Taylor M, Adriano A, Davenport C, et al. Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database of Systematic Reviews* 2021, Issue 3. Art. No: CD013705. [DOI: 10.1002/14651858.CD013705]

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]

Leeflang 2021
Leeflang MMG, Bell K, Deeks JJ, Dinnes J, Doust J, Korevaar DA, et al. Electronic and animal noses for detecting SARS-CoV-2 infection. *Cochrane Database of Systematic Reviews* 2021 Jun 28, Issue 6. Art. No: CD015013. [DOI: 10.1002/14651858.CD015013]

Li 2020
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]

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]

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]

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 SoftwareEPPI-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.