Accuracy of routine laboratory tests to predict mortality and deterioration to severe or critical COVID-19 in people with SARS-CoV-2

Cochrane COVID-19 Diagnostic Test Accuracy Group; Verbakel, Jan Y.; De Rop, Liselore; Stegeman, Inge; Holtman, Gea A.; Ochodo, Eleanor A.; Yang, Bada; Guleid, Fatuma; Davenport, Clare; Deeks, Jonathan J.

Published in:
Cochrane Database of Systematic Reviews

DOI:
10.1002/14651858.CD015050

Cochrane Library

Cochrane Database of Systematic Reviews

Accuracy of routine laboratory tests to predict mortality and deterioration to severe or critical COVID-19 in people with SARS-CoV-2 (Protocol)

Verbakel JY, De Rop L, Stegeman I, Holtman GA, Ochodo EA, Yang B, Guleid F, Davenport C, Deeks JJ, Dinnes J, Dittrich S, Emperador D, Hooft L, Spijker R, Van den Bruel A, Wang J, Takwoingi Y, Langendam MW, Leeflang MMG, Cochrane COVID-19 Diagnostic Test Accuracy Group

Verbakel JY, De Rop L, Stegeman I, Holtman GA., Ochodo EA, Yang B, Guleid F, Davenport C, Deeks JJ, Dinnes J, Dittrich S, Emperador D, Hooft L, Spijker R, Van den Bruel A, Wang J, Takwoingi Y, Langendam MW, Leeflang MMG, Cochrane COVID-19 Diagnostic Test Accuracy Group.
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, Issue 9. Art. No.: CD015050.
DOI: 10.1002/14651858.CD015050.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Protocol]

Accuracy of routine laboratory tests to predict mortality and deterioration to severe or critical COVID-19 in people with SARS-CoV-2

Jan Y Verbakel[1,2], Liselore De Rop[1], Inge Stegeman[3,4,5], Gea A. Holtman[6], Eleanor A Ochodo[7,8], Bada Yang[5], Fatuma Guleid[9], Clare Davenport[10,11], Jonathan J Deeks[10,11], Jacqueline Dinnes[10,11], Sabine Dittrich[12], Devy Emperador[12], Lotty Hooft[13], René Spijker[14,15], Ann Van den Bruel[1], Junfeng Wang[16], Yemisi Takwoingi[10,11], Miranda W Langendam[5], Mariska MG Leeflang[17], Cochrane COVID-19 Diagnostic Test Accuracy Group[11]

[1]Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium. [2]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom. [3]Department of Otorhinolaryngology & Head and Neck Surgery, University Medical Center Utrecht, Utrecht , Netherlands. [4]Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands. [5]Epidemiology and Data Science, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands. [6]Department of General Practice, University of Groningen, University Medical Centre Groningen, Groningen, Netherlands. [7]Centre for Evidence-based Health Care, Department of Global Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa. [8]Centre for Global Health Research, Kenya Medical Research Institute, Kisumu, Kenya. [9]KEMRI-Wellcome Trust Research Programme, Nairobi, Kenya. [10]Test Evaluation Research Group, Institute of Applied Health Research, University of Birmingham, Birmingham, UK. [11]NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham, Birmingham, UK. [12]FIND, Geneva, Switzerland. [13]Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht , Netherlands. [14]Medical Library, Amsterdam UMC, University of Amsterdam, Amsterdam Public Health, Amsterdam, Netherlands. [15]Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands. [16]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Netherlands. [17]Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands

Contact address: Jan Y Verbakel, jan.verbakel@kuleuven.be.

Editorial group: Cochrane Infectious Diseases Group.
Publication status and date: New, published in Issue 9, 2021.

Citation: Verbakel JY, De Rop L, Stegeman I, Holtman GA., Ochodo EA, Yang B, Guleid F, Davenport C, Deeks JJ, Dinnes J, Dittrich S, Emperador D, Hooft L, Spijker R, Van den Bruel A, Wang J, Takwoingi Y, Langendam MW, Leeflang MMG, Cochrane COVID-19 Diagnostic Test Accuracy Group. 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, Issue 9. Art. No.: CD015050. DOI: 10.1002/14651858.CD015050.

ABSTRACT

Objectives

This is a protocol for a Cochrane Review (diagnostic). The objectives are as follows:

To assess the accuracy of routine blood-based laboratory tests to predict mortality and deterioration to severe or critical (from mild or moderate) COVID-19 in people with SARS-CoV-2 infection.

Secondary objectives

Where data are available, we will investigate whether prognostic accuracy varies according to a specific measurement or test, reference standard, timing of outcome verification, sample type, study design, and setting, including prevalence of the target condition (either by stratified analysis or meta-regression).

BACKGROUND

On 30 December 2019, a report about a cluster of people with pneumonia of unknown origin in Wuhan, China, was publicly described in ProMED (promedmail.org/promed-posts). In January 2020, it became clear that this was caused by a new coronavirus, and was also spreading to other countries. In March 2020, the World Health Organization (WHO) declared that the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and resulting coronavirus disease-2019 (COVID-19) was a worldwide pandemic. This pandemic, in combination with the novelty of the virus, presents important diagnostic challenges.

These challenges range from understanding the value of signs and symptoms in predicting possible infection; assessing whether existing biochemical and imaging tests can identify infection and people who need critical care; to evaluating whether new diagnostic tests can provide accurate, rapid, and point of care testing, which can either identify current infection, rule out infection, identify people in need of care escalation, or test for past infection and immunity.

Prognostic accuracy studies evaluate the ability of medical tests to predict disease, or to identify people who are likely to experience an adverse medical event among those who have disease. These studies typically present results in terms of sensitivity and specificity, or the area under the curve. The current review focusses on the prognostic accuracy of individual biomarkers to predict mortality and deterioration to severe or critical COVID-19. This is fundamentally different from (1) prediction model reviews, which focus on the predictive performance of models, and critically appraise model development studies (including discrimination and calibration) and external validation studies of prediction models (Wynants 2020); and (2) prognostic factor studies, which investigate the association between a test, biomarker, or personal characteristic and a future outcome, either by itself, or over and above other known predictors. Prognostic factor studies typically present the results as a measure of association, rather than the accuracy at a given test threshold for test positivity.

This Cochrane Review will concentrate on the accuracy of routine, blood-based laboratory tests to predict death and deterioration to severe or critical COVID-19 disease, in people with SARS-CoV-2 infection. In clinical care, routine laboratory markers, such as white blood cell count, measures of anticoagulation, C-reactive protein (CRP), and procalcitonin, are used to assess a person's health status. These laboratory markers are also used in people with COVID-19, and may be useful for hospital triage, to assess whether a person with COVID-19 should receive outpatient treatment or more intensive treatment, which usually requires a hospital admission.

This review follows a generic protocol that covers the full series of Cochrane Diagnostic Test Accuracy (DTA) Reviews for the diagnosis of COVID-19 (Deeks 2020a). Therefore, we use some text in the Background and Methods sections that was originally published in that protocol, and some text overlaps with some of our other reviews (Deeks 2020b; Dinnes 2020; Islam 2021; Stegeman 2020; Struyf 2021).

Target condition being diagnosed

COVID-19 is the disease caused by infection with SARS-CoV-2. SARS-CoV-2 infection can be asymptomatic (no symptoms); or cause mild or moderate signs and symptoms (such as fever, cough, aches, lethargy, breathlessness, and fast breathing); severe signs and symptoms (which include severe respiratory distress and low oxygen saturation, indicative of severe pneumonia); or critical signs and symptoms (which require respiratory support due to acute respiratory distress syndrome (ARDS)), or lead to organ dysfunction (indicative of sepsis). People with severe or critical COVID-19 require distinctive management of their signs and symptoms; it is important to identify them.

At present, as polymerase chain reaction (PCR) testing is more prevalent compared to when the pandemic started, clinicians look to the added value of routine laboratory tests to inform their decision on whether to admit people with a suspected or confirmed SARS-CoV-2 infection to hospital, or to adapt a watchful waiting approach, if they suspect mild or moderate COVID-19. Therefore, in this review, we will focus on the distinction between mild or moderate, and severe or critical COVID-19 on the one hand, and mortality from COVID-19 on the other.

Index test(s)

Routinely available blood-based biomarkers for infection and inflammation may be considered in the investigation of people with suspected or confirmed SARS-CoV-2 infection. Evaluation of commonly available tests may be helpful to predict death or deterioration of a person with mild or moderate COVID-19 to severe or critical COVID-19.

We will collate evidence on all routine blood, plasma, and serum biomarker tests reported in the identified studies.

Clinical pathway

Standard workup for people suspected of having COVID-19 consists of assessing signs and symptoms, and doing a PCR test. However, as people with COVID-19 present with a variety of symptoms, of varying severity, and as they may deteriorate quickly, it is important to be able to predict who will deteriorate and who may not. Therefore, it is common practice to perform routine laboratory tests whenever people are assessed at the hospital (either outpatient or inpatient).

Routine laboratory tests may be used to predict deterioration from mild or moderate disease to severe or critical outcomes in people with a suspected or confirmed SARS-CoV-2 infection. In ambulatory care, the decision to refer a person with a SARS-CoV-2 infection implies the potential breach of quarantine measures. Routine laboratory tests might help to inform the decision to treat the person at home, to reduce the workload of already burdened hospitals and intensive care units. More favourable laboratory test results could support ambulatory care management of people with COVID-19, providing clinicians with information on which signs and symptoms might trigger a further diagnostic workup. For people who are hospitalized, routine laboratory tests may inform the decision to refer them to the intensive care unit (ICU), or confirm that they are stable enough to remain in the general ward.

Alternative test(s)

In emergency departments, chest X-rays, ultrasounds, and computed tomography (CT) are also widely used to assess the severity of a person's condition, especially in the case of pneumonia. Which imaging test is available may depend on the

type of hospital and available resources, e.g. a tertiary care hospital in a high-income country may have a mobile CT scanner available, while smaller hospitals may only have an X-ray and ultrasound machine. These imaging tests have the advantage that they can enable a visual assessment of the condition of the lungs.

Rationale

It is essential to understand the prognostic accuracy of tests, to inform clinicians on how to use them optimally in different settings, to help to develop effective management pathways. New evidence about routine laboratory testing is becoming available quickly.

An alternative may be to use prediction models, rather than alternative tests. However, not all laboratories measure all biomarkers and tests required to estimate a specific model. Furthermore, it would be very useful if there was a biomarker that could serve as a 'red flag': if this biomarker is positive, it means that the person needs extra care to prevent deterioration.

Therefore, we will produce a Cochrane Living Review (a systematic review that is continually updated, incorporating relevant new evidence as it becomes available) that will summarize new and existing evidence on the prognostic accuracy of routine laboratory markers.

While we want to examine the accuracy of routine laboratory markers to predict person-related outcomes in this review, the DTA framework still applies, with the addition of some adaptations to the risk of bias assessment, to account for the prognostic nature of our research objectives.

OBJECTIVES

To assess the accuracy of routine blood-based laboratory tests to predict mortality and deterioration to severe or critical (from mild or moderate) COVID-19 in people with SARS-CoV-2 infection.

Secondary objectives

Where data are available, we will investigate whether prognostic accuracy varies according to a specific measurement or test, reference standard, timing of outcome verification, sample type, study design, and setting, including prevalence of the target condition (either by stratified analysis or meta-regression).

METHODS

Criteria for considering studies for this review

Types of studies

We will keep the eligibility criteria broad to include all groups of people and all variations of a test. If the participant population is unclear, we will include the study.

We will include studies of all designs that produce estimates of prognostic accuracy, or provide data from which estimates can be computed: cross-sectional studies, case-control designs (using participants from a single original cohort), and consecutive series of participants assessing the prognostic accuracy of routine laboratory tests.

We will include only single-gate designs, in which a single group of participants who may develop the target condition or event, is recruited. We plan to include studies that based their results on individual participants, and studies that based their results on laboratory samples. We will carefully consider the limitations of different study designs, using quality assessment and analysis.

Participants

We will include studies recruiting people who present to outpatient services, or are admitted to general hospital wards with confirmed SARS-CoV-2 infection, and studies that are based on serum banks of samples from people with confirmed COVID-19.

Studies must include a minimum of 10 samples or 10 participants.

Index tests

We will collect evidence on all routine blood-based laboratory tests performed during the first encounter with the person as part of the initial routine diagnostic workup (e.g. during admission to hospital, or during first assessment), as reported in the identified studies.

These may include, but are not limited to:

- **White blood cells**
  - White blood cell count (WBC)/leukocyte count
  - Monocyte count
  - Monocytes percentage
  - Neutrophil count
  - Neutrophils percentage
  - Lymphocyte count
  - Lymphocytes percentage
  - Neutrophyl-to-lymphocyte ratio
  - Eosinophils count
  - Eosinophils percentage
- **Red blood cells**
  - Red blood cells
  - Haemoglobin
- **Biochemistry**
  - Potassium
  - Sodium
  - Chloride
  - Calcium (free)
- **Kidney function tests**
  - Serum creatinine
  - Creatinine clearance
  - Blood urea nitrogen (BUN)
  - Estimated glomerular filtration rate (eGFR)
  - Cystatin-C

- **Liver and cholestasis markers**
  - Albumin
  - Pre-albumin
  - Globulin
  - Albumin/globulin ratio (A/G)
  - Alanine aminotransferase (ALT)
  - Aspartate aminotransferase (AST)
  - Alkaline phosphatase (ALP)
  - Gamma-glutamyltransferase (GGT)
  - Total bilirubin
  - Direct bilirubin
  - Indirect bilirubin
- **Coagulation markers**
  - Platelet count
  - Activated partial thromboplastin time (APTT)
  - Thrombin time
  - Prothrombin time
  - D-dimer
  - Fibrinogen
- **Cardiac markers**
  - Brain natriuretic peptide (BNP)
  - NT-pro-BNP
  - Troponin T
  - Hypersensitive troponin T
  - Myoglobin
  - Creatine kinase
  - Creatine kinase - MB
- **Metabolic markers and markers of cell damage**
  - Lactate
  - Uric acid (UA)
  - Lactate dehydrogenase (LDH)
  - Glucose
- **Inflammation markers, immune markers, and specific subgroups of leukocytes**
  - Procalcitonin (PCT)
  - Interleukins
  - Erythrocyte sedimentation rate (ESR)
  - C-reactive protein (CRP)
  - Serum ferritin

We will interpret the term 'routine' broadly, considering that some markers will be more routine in some settings or countries than in others. A positive test is defined as an increase in values compared to the normal ranges, or as a decrease compared to normal values.

Target conditions

To be eligible, studies must identify a current SARS-CoV-2 infection.

First target condition

Death due to any cause (no specific prediction horizon prespecified)

Second target condition

Deterioration from mild or moderate to severe or critical COVID-19 cases

Reference standards

We expect the definitions for mild, moderate, severe, and critical COVID-19 to vary between publications, and to be poorly reported. Therefore, we will include any reference standard used to define severity that is provided by the authors, and document the definitions.

Unless otherwise provided by the original paper, the study will need to make a distinction between mild (to moderate) and severe (to critical) cases as defined by the WHO Clinical Management of COVID-19 interim guidance report (WHO 2020).

- Mild disease: people with symptoms who meet the case definition for COVID-19, without evidence of viral pneumonia or hypoxia
- Moderate disease (pneumonia):
  - adolescent or adult with clinical signs of pneumonia (fever, cough, dyspnoea, fast breathing), but with SpO_2 ≥ 90% on room air, and no signs of severe pneumonia (54)
  - child with clinical signs of non-severe pneumonia (cough or difficulty breathing plus fast breathing or chest indrawing, or both), and no signs of severe pneumonia. Fast breathing is defined as (breaths/minute): < 2 months old: ≥ 60; 2 to 11 months old: ≥ 50; 1 to 5 years old: ≥ 40 (55)
- Severe disease (severe pneumonia):
  - adolescent or adult with clinical signs of pneumonia (fever, cough, dyspnoea, fast breathing) plus one of the following: respiratory rate > 30 breaths/minute; severe respiratory distress; or SpO_2 < 90% on room air (54)
  - child with clinical signs of pneumonia (cough or difficulty in breathing) plus at least one of the following:
    - ☐ central cyanosis or SpO_2 < 90%; severe respiratory distress (e.g. fast breathing, grunting, very severe chest indrawing); general danger sign: inability to breastfeed or drink, lethargy or unconsciousness, or convulsions (55,56)
    - ☐ fast breathing is defined as (breaths/minute): < 2 months old: ≥ 60; 2 to 11 months old: ≥ 50; 1 to 5 years old: ≥ 40 (55)
- Critical disease: acute respiratory distress syndrome (ARDS), sepsis or septic shock. We will also categorize the following outcomes as critical disease: ICU admission, need for ventilation, and need for intubation.

While these diagnoses can be made on clinical grounds; chest imaging (radiograph, CT scan, ultrasound) may assist in making the diagnosis, and identify or exclude pulmonary complications.

We will assess, and extract if available, the prediction horizon of the biomarkers identified in included studies, aiming to allow reasonable assessment and comparison of the laboratory markers measured at baseline (e.g. admission to hospital or first assessment).

We will assess the quality of these definitions according to the criteria listed in our QUAPAS Table, and we will provide a qualitative overview of the reference standards used and reported in the included studies (Table 1).

Search methods for identification of studies

Electronic searches

We will search the Cochrane COVID-19 Study Register (covid-19.cochrane.org/).

The Cochrane COVID-19 Study Register is a specialised register built within the Cochrane Register of Studies (CRS) and is 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.

We will perform the search using the CRSweb interface (crsweb.cochrane.org), using a strategy that combines a search for diagnostic studies, a search for prognostic characteristics, and a search for severely ill patients. See Appendix 1 for search terms.

Data collection and analysis

Selection of studies

Two review authors will independently screen each title and abstract for possible inclusion. In the next step, two review authors will independently screen the full-text of each possibly relevant article. From the final list of included studies, we will perform both forward and backward citation tracking, using Microsoft Academic through EPPI-Reviewer (EPPI-Reviewer 2020).

For articles only available in languages other than English, we will use Google Translate, or review authors who can read and understand that language will perform translations. We will solve disagreements by discussion. If discussion cannot solve the dispute, we will consult a third review author.

Data extraction and management

Two review authors will independently extract data from each included study. We will assign multiple studies with the same first author to one extractor, so that they can detect preprints from already peer-reviewed, published articles. We will contact study authors, when needed, to check details and obtain missing information.

We will extract data on the country and region, the setting, the time period of the study, funding, and information needed for the 'Characteristics of included studies' tables. Studies may have defined a positive test result as a decrease or an increase compared to normal values, or as both an increase and a decrease. Where possible, we will adapt the 2 x 2 tables so that all studies included in the analyses report on the same definition of test positivity. However, if studies report both an increase and a decrease as a positive test result, we will include both. We will resolve disagreements by discussion between the two review authors, and two other review authors will check the results when these are entered into Review Manager 5 (Review Manager 2020).

Assessment of methodological quality

Two review authors will independently assess risk of bias and applicability concerns. Because we are assessing prognostic accuracy in this review, we will use the Quality Assessment of Prognostic Accuracy Studies (QUAPAS) tool, which incorporates elements of the QUADAS-2 tool, supplemented by elements of the QUIPS and PROBAST tools, and adds a fifth domain of 'analysis' to the quality appraisal (Table 1). (Whiting 2011, Wolff 2019, Hayden 2013) We will resolve disagreements by discussion between three review authors.

The focus of our review is on prognostic accuracy, and not the predictive performance of models, so we will not be critically appraising model development studies (including discrimination and calibration), and external validation studies of prediction models. Therefore, we decided that using QUIPS (prognostic factors) or PROBAST (models) as such, was unsuitable.

The other four domains of the QUAPAS tool are identical to the QUADAS-2 tool: patient selection, index test, reference standard, and flow and timing (Whiting 2011). Each domain is assessed for risk of bias, and the first three domains are also assessed for concerns of applicability. Signaling questions are included to help judge bias. Table 1 shows the definitions used to assess the methodological quality.

Statistical analysis and data synthesis

Although this review focusses on prognostic accuracy to predict patient-related outcomes, the same approach as for DTA reviews applies, given the nature of the data identified in the primary studies (2 x 2 tables for each test in each study).

Most routine laboratory tests provide test results as continuous measurements. That means that an explicit threshold is needed to provide positive and negative results, to estimate diagnostic characteristics, such as sensitivity and specificity. Some tests indicate mild (or moderate) versus severe (or critical) disease if the value is decreased relative to the normal ranges. For other tests, mild versus severe disease is indicated when the value is increased. For another group of tests, both an increase and a decrease may indicate the presence of mild or severe disease. For each test in each study, we will report the threshold used in our analyses, and whether an increase or a decrease in value is to be regarded as a positive test result.

From each study, we will include one threshold for each test. If multiple thresholds are reported, we will choose the threshold closest to a predefined value of relevance. We will present the sensitivity and specificity results in forest plots, and provide positive and negative predictive values for each study. We will report the median and interquartile range (IQR) of pre-test probability of the target condition in 2 x 2 tables from single-gate studies.

We will consider that a meta-analysis is appropriate when four or more studies report on a particular test. As studies might report different thresholds for the same test, we will use the Hierarchical Summary Receiver Operator Curve (HSROC) model

for meta-analyses, to estimate summary curves, recommended by the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Macaskill 2010). Since summary sensitivities and specificities are only clinically interpretable when the studies included in a meta-analysis use a common cut-off, we will estimate sensitivity at points on the SROC curves that correspond to the median specificity observed in the studies included in the meta-analysis. We will report the estimates for the first and third quartile specificity in the summary of findings table. We will use SAS 9.4, using PROC NLMIXED, for the meta-analyses (SAS 2015).

In order to identify the most discriminative test in the situation, we will compare the prognostic accuracy of biomarkers with a minimum estimated sensitivity of 50% (point estimate), at a minimum specificity of 50% (either median or IQR). We will perform these analyses on all studies that evaluate one of these tests (indirect comparison). We will also perform analyses that are restricted to studies that make head-to-head comparisons (i.e. assessed two of the biomarkers in the same participant), when at least four studies are included that enabled these direct comparisons. We will make test comparisons by adding a covariate for test type to the HSROC model, to assess the effect of test type on the accuracy, cut-off, or shape parameters of the model. Whenever the estimated SROC curves have the same shape, we will calculate the relative diagnostic odds ratio (RDOR) as a summary of the relative accuracy of the two biomarkers at hand. To assess the statistical significance of differences in test accuracy, we will use likelihood ratio tests for comparisons of models with and without covariate terms. If fewer than 10 primary studies are available for the head-to-head comparison, we will assume that the shape parameter of the model is equal for the biomarkers under evaluation.

Investigations of heterogeneity

If adequate data are available, we will investigate the following sources of heterogeneity: measurement technique or test type, reference standard, timing of outcome verification, sample type, study design, and setting, including prevalence of the target condition, either using stratification (where we believe it is inappropriate to combine studies), or with meta-regression models.

Summary of findings and assessment of the certainty of the evidence

We will develop a list of key findings in summary of findings tables, and determine the certainty in the summary estimates for each test and findings, using the GRADE approach (Schünemann 2020a; Schünemann 2020b).

Starting at high certainty, we will downgrade by one level when at least half of the studies are at high risk of bias for one or more domains; we will downgrade for indirectness when we have high concerns of applicability for at least one domain in at least half of the studies; we will downgrade for imprecision when fewer people with the target condition are included than would have been needed to achieve the sensitivity estimates listed, and the confidence interval is wider than 10 percentage points; and we will downgrade for inconsistency when study estimates differ more than 20 percentage points from each other.

Updating

We will undertake the searches of published literature, preprints, and new test approvals monthly, and depending on the relevance of our research question and the number of new and important studies found at each search update, we will consider updating this review with each search, if resources allow.

Sensitivity analyses

We aim to undertake sensitivity analyses considering the impact of unpublished studies. We aim to perform sensitivity analyses to investigate the impact of prospective versus retrospective data collection.

Assessment of reporting bias

We aim to publish lists of studies that we know exist but for which we have not managed to locate reports, and request information to include in review updates.

ACKNOWLEDGEMENTS

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

- the project team (Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Leeflang MMG, Spijker R, Hooft L, Van den Bruel A, Verbakel JY, McInnes MDF, Emperador D, Dittrich S, Cunningham J);
- the systematic review teams for each review:
  - Molecular, antigen, and antibody tests (Adriano A, Arevalo-Rodriguez I, Beese S, Buitrago DC, Ciapponi A, Domen J, Dretzke J, Ferrante di Ruffano L, Harris I, Mateos M, Price M, Taylor M, Taylor-Phillips S)
  - Signs and symptoms (Struyf T, Horn S)
  - Routine laboratory markers (Yang B, Langendam M, Ochodo EA, Guleid F, Holtman GA, Wang J, Stegeman I)
  - Accuracy of routine laboratory tests (De Rop L)
  - Imaging tests (Salameh JP, McGrath TA, van der Pol CB, Frank RA, Prager R, Hare SS, Dennie C, Jenniskens K, Korevaar DA, Cohen JF, van de Wijgert J, Damen JAAG);
- the wider team of systematic reviewers from the University of Birmingham, UK who assisted with title and abstract screening across the entire suite of reviews for the diagnosis of COVID-19 (Agarwal R, Baldwin S, Berhane S, Herd C, Kristunas C, Quinn L, Scholefield B).

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

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

REFERENCES

Additional references

Deeks 2020a

Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Leeflang MMG, Spijker R, et al. Diagnosis of SARS-CoV-2 infection and COVID-19: accuracy of signs and symptoms; molecular, antigen, and antibody tests; and routine laboratory markers. *Cochrane Database of Systematic Reviews* 2020, 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, Issue 6. Art. No: CD013652. [DOI: 10.1002/14651858.CD013652]

Dinnes 2020

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.pub2]

EPPI-Reviewer 2020 [Computer program]

EPPI-Centre Software. UCL Social Research Institute EPPI-Reviewer: advanced software for systematic reviews, maps and evidence synthesis.. Thomas J, Graziosi S, Brunton J, Ghouze Z, O'Driscoll P, Bond M. London: EPPI-Centre Software. UCL Social Research Institute, 2020.

Hayden 2013

Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. *Annals of Internal Medicine* 2013;**158**(4):280-6.

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]

Macaskill 2010

Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, editor(s). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration, 2010. Available from srdta.cochrane.org.

Review Manager 2020 [Computer program]

The Cochrane Collaboration Review Manager 5 (RevMan 5). Version 5.4. The Cochrane Collaboration, 2020.

SAS 2015 [Computer program]

SAS Institute SAS. Version 9.4. Cary, NC, USA: SAS Institute, 2015. Available at www.sas.com.

Schünemann 2020a

Schünemann HJ, Mustafa RA, Brozek J, Steingart KR, Leeflang M, Murad MH, et al. GRADE guidelines: 21 part 1. Study design, risk of bias and indirectness in rating the certainty across a body of evidence for test accuracy. *Journal of Clinical Epidemiology* 2020;**122**:129-41.

Schünemann 2020b

Schünemann HJ, Mustafa RA, Brozek J, Steingart KR, Leeflang M, Murad MH, et al. GRADE guidelines: 21 part 2. Test accuracy: inconsistency, imprecision, publication bias and other domains for rating the certainty of evidence for test accuracy and presenting it in evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology* 2020;**122**:142-52.

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 2021

Struyf T, Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Leeflang MMG, 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, Issue 2. Art. No: CD013665. [DOI: 10.1002/14651858.CD013665.pub2]

Whiting 2011

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529-36.

WHO 2020

World Health Organization. Clinical management of COVID-19: interim guidance, 27 May 2020. https://apps.who.int/iris/handle/10665/332196 (accessed prior to 5 August 2021).

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. *Annals of Internal Medicine* 2019;**170**(1):51-8.

Wynants 2020

Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of COVID-19: systematic review and critical appraisal. *BMJ* 2020;**369**:m1328.

ADDITIONAL TABLES

Table 1. QUAPAS domains

Domain	Participants	Index test	Target event	Flow and timing	Analysis
Description	Describe methods for recruiting participants[a,b] Describe participants[a,b] (previous testing, presentation, intended use of index test, and setting)[a]	Describe the index test (definition, method of measurement, and interpretation)[a,b]	Describe the target event (definition, method of measurement, and interpretation)[a]	Describe any participants lost to follow-up[a,b] or excluded from 2 x 2 table[b] Describe the time horizon from the index test to the target event	Describe the statistical methods[b]
Signaling questions (yes, no, unclear)	Was there consecutive or random enrolment of participants?[a] Was a case-control study design avoided?[a] Did the study avoid inappropriate exclusions?[a]	Was the method and setting for performing the index test the same for all participants?[b] Was the index test measured without knowledge of the target event?[a] If a threshold was used, was it prespecified?[a]	Was the method used to measure the target event the same for all participants?[b] Was the target event measured without knowledge of the index test results?[a] Was the method used to measure the target event valid and reliable?[a,b]	Was information on the target event available for all participants?[a,b] Was information on the index test available for all participants?[b] Was the time horizon sufficient to capture the target event?[c]	Were methods used to account for competing events?[c] Were methods used to account for censoring?[c] Were imputation methods used for missing data?[b]
Risk of bias (high, low, unclear)	Could the recruitment of participants have introduced bias?[a]	Could the conduct or interpretation of the index test have introduced bias?[a]	Could measurement of the target event have introduced bias?[a]	Could the study flow have introduced bias?[a]	Could the analysis have introduced bias?[b]
Concerns about applicability (high, low, unclear)	Are there concerns that the participants do not match the review question?[a]	Are there concerns that the index test, its conduct, or its interpretation differ from the review question?[a]	Are there concerns that the target event does not match the review question?[a]	Are there concerns that the time horizon does not match the review question?	

[a]Item of QUADAS
[b]Item of QUIPS
[c]Item of PROBAST
QUAPAS: explanation
The Quality Assessment of Prognostic Accuracy Studies (QUAPAS) tool was initiated to address the need for a risk of bias and applicability assessment tool in systematic reviews of prognostic accuracy studies. While tools such as QUADAS-2 exist to assess methodological quality in diagnostic accuracy reviews, as well as QUIPS (prognostic factors) and PROBAST (models) for prognostic reviews, no comparable tool exists for reviews of prognostic accuracy studies.
Based on this need and availability of reliable existing resources, the QUAPAS tool was created by mapping the relevant items from QUIPS and PROBAST to the existing domain-based framework and logic of QUADAS-2. This way, the format of QUADAS-2 could be used, with which many reviewers of diagnostic accuracy studies are familiar: domains, signalling questions, and a judgment on risk of bias and applicability concerns of a given primary study.
With QUAPAS, the aim was to focus on the key distinguishing factor between diagnostic and prognostic test accuracy research: the longitudinal study design and time dependent occurrence of the outcome, inherent to the research question. The domains of QUADAS-2 were modified to account for factors unique to prognostic research, while keeping intact signalling questions where potential risk of bias assessment does not differ significantly between diagnostic or prognostic questions (*Participants* and *Index Test* domains). Changes were applied to the *Reference Standard* (now called *Target Event*) and *Flow and Timing* domains to better account for bias introduced from longitudinal research question.

Inspired by QUIPS, PROBAST, and our understanding of prognostic research, a fifth domain was added, called *Analysis*, as there are complexities introduced from time-dependent analysis. In this fifth domain, two signalling questions from PROBAST were added, and one from QUIPS, as neglecting methods for handling censoring, competing events, and missing data raise concern for bias in prognostic accuracy studies. Other domains, such as *Confounding* (from QUIPS), were excluded entirely from QUAPAS, as causal questions are not relevant for test accuracy studies.

Despite the changes to the QUADAS-2 domain names and signalling questions, QUAPAS is intended to be used in the same manner. Reviewers assess risk of bias for each of the five domains, with signalling questions to aid the grading as high, low, or unclear. Concerns for applicability are graded on the same scale for four of the five domains.

APPENDICES

Appendix 1. Search strategy

1	MESH DESCRIPTOR Biomarkers EXPLODE ALL AND COVID19:INREGISTER
2	MESH DESCRIPTOR Sensitivity and Specificity EXPLODE ALL AND COVID19:INREGISTER
3	MESH DESCRIPTOR Cross-Sectional Studies EXPLODE ALL AND COVID19:INREGISTER
4	("Cross-sectional"):SG
5	((biomarker* or marker* or test or tests or diagn* or discrimin* or detect* or sensitivity or specificity or auc or predictive-value or NPV or PPV or accuracy or case-control* or cross-sectional):AB OR (biomarker* or marker* or test or tests or diagn* or discrimin* or detect* or sensitivity or specificity or auc or predictive-value or NPV or PPV or accuracy or case-control* or cross-sectional):TI) AND COVID19:INREGISTER
6	#1 OR #2 OR #3 OR #4 OR #5
7	MESH DESCRIPTOR Prognosis EXPLODE ALL AND COVID19:INREGISTER
8	MESH DESCRIPTOR Clinical Decision Rules EXPLODE ALL AND COVID19:INREGISTER
9	((prognos* OR cohort OR validat* OR predict* OR follow-up):AB OR (prognos* OR cohort OR validat* OR predict* OR follow-up):TI) AND COVID19:INREGISTER
10	#9 OR #7 OR #8
11	MESH DESCRIPTOR Mortality EXPLODE ALL AND COVID19:INREGISTER
12	MESH DESCRIPTOR Intensive Care Units EXPLODE ALL AND COVID19:INREGISTER
13	MESH DESCRIPTOR Continuous Positive Airway Pressure AND COVID19:INREGISTER
14	MESH DESCRIPTOR Critical Illness EXPLODE ALL AND COVID19:INREGISTER
15	MESH DESCRIPTOR Survival Analysis EXPLODE ALL AND COVID19:INREGISTER
16	MESH DESCRIPTOR Intubation, Intratracheal EXPLODE ALL AND COVID19:INREGISTER
17	((sever* or mortal* Or critical* Or death* Or ventilation OR intub* or intensive-care or icu or survival):AB OR (sever* or mortal* Or death* Or critical* Or ventilation OR intub* or intensive-care or icu or survival):TI) AND COVID19:INREGISTER
18	#17 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16

(Continued)

18	#6 AND #10 AND #18

CONTRIBUTIONS OF AUTHORS

Jon Deeks was unable to sign-off on the final protocol version, but co-authors agreed he fully contributed to the protocol.

Authors read and approved the final protocol version.

DECLARATIONS OF INTEREST

JYV has no known conflicts of interest.

LDR has no known conflicts of interest.

IS has no known conflicts of interest.

GAH has no known conflicts of interest.

EO has no known conflicts of interest.

BY has no known conflicts of interest.

FG has no known conflicts of interest.

CD has no known conflicts of interest.

JJD has published or been quoted in opinion pieces in scientific publications, and in the mainstream and social media related to diagnostic testing.

JD has no known conflicts of interest.

SD has no known conflicts of interest.

DE is employed by FIND, and has no known conflicts of interest. FIND is a global not-for-profit product development partnership and WHO Diagnostic Collaboration Centre. It is FIND's role to accelerate access to high-quality diagnostic tools for low-resource settings and this is achieved by supporting both R&D and access activities for a wide range of diseases, including COVID-19. FIND has several clinical research projects to evaluate multiple new diagnostic tests against published Target Product Profiles that have been defined through consensus processes. These studies are for diagnostic products developed by private sector companies who provide access to know-how, equipment/reagents, and contribute through unrestricted donations as per FIND policy and external SAC review.

LH has no known conflicts of interest.

RS has no known conflicts of interest.

AVdB has no known conflicts of interest.

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

YT has no known conflicts of interest.

MWL has no known conflicts of interest.

MMGL has no known conflicts of interest.

SOURCES OF SUPPORT

Internal sources

- Liverpool School of Tropical Medicine, UK

External sources

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

  Project number 300342-104
- National Institute for Health Research (NIHR), UK
- NIHR Birmingham Biomedical Research Centre at the University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham, UK