

BMJ Open Role and utility of COVID-19 laboratory testing in low-income and middle-income countries: protocol for rapid evidence synthesis

Ojiambo Kevin Ouma,^{1,2} Kisangala Ephraim,^{2,3} Nakalembe Loyce,^{2,4} Eve Namisango,^{2,5} Fred Nalugoda,^{2,6} Regina Ndagire,^{2,3} Rachel Nante Wangi ,^{2,3} Brenda Allen Kawala,^{2,7} Thomas Katairo ,^{2,8} Allen Eva Okullo,^{2,9} Robert Apunyo,² Daniel Semakula ,^{2,10} Ash Luwambo,^{2,11} Alison Annet Kinengyere,^{2,12} Nelson Sewankambo,^{2,13} Sheila N Balinda,¹⁴ Moses Ocan ,^{2,15} Ekwaro A Obuku^{2,16}

To cite: Ouma OK, Ephraim K, Loyce N, *et al.* Role and utility of COVID-19 laboratory testing in low-income and middle-income countries: protocol for rapid evidence synthesis. *BMJ Open* 2021;**11**:e050296. doi:10.1136/bmjopen-2021-050296

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-050296>).

Received 16 February 2021
Accepted 22 September 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Moses Ocan;
ocanmoses@gmail.com

ABSTRACT

Introduction Accurate and affordable laboratory testing is key to timely diagnosis and appropriate management of patients with COVID-19. New laboratory test protocols are released into the market under emergency use authorisation with limited evidence on diagnostic test accuracy. As such, robust evidence on the diagnostic accuracy and the costs of available tests is urgently needed to inform policy and practice especially in resource-limited settings. We aim to determine the diagnostic test accuracy, cost-effectiveness and utility of laboratory test strategies for COVID-19 in low-income and middle-income countries.

Methods and analysis This will be a multistaged, protocol-driven systematic review conducted in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for diagnostic test accuracy studies. We will search for relevant literature in at least six public health databases, including PubMed, Google Scholar, MEDLINE, Scopus, Web of Science and the WHO Global Index Medicus. In addition, we will search Cochrane Library, COVID-END and grey literature databases to identify additional relevant articles before double-screening and abstraction of data. We will conduct a structured narrative and quantitative synthesis of the results guided by the Fryback and Thornbury framework for assessing a diagnostic test. The primary outcome is COVID-19 diagnostic test accuracy. Using the GRADE approach specific to diagnostic accuracy tests, we will appraise the overall quality of evidence and report the results following the original PRISMA statement. The protocol is registered with the International Prospective Register of Systematic Reviews (PROSPERO; <https://www.crd.york.ac.uk/prospero/>).

Ethics and dissemination Ethical review was done by the School of Biomedical Sciences Research Ethics Committee and the Uganda National Council for Science and Technology. The published article will be accessible to policy and decision makers. The findings of this review will guide clinical practice and policy decisions and highlight areas for future research.

Strengths and limitations of this study

- The study will contribute to strengthening the evidence base on the effectiveness of laboratory testing strategies for COVID-19 in hospitals and community populations in low-income and middle-income countries (LMICs).
- The protocol has been written following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.
- The GRADE system will be used to ascertain the strength of the evidence base for each outcome and to report data for the primary outcome in a 'Summary of Findings table'.
- The review is limited to evidence from LMICs.
- Non-English databases will not be searched and this may introduce language bias.

PROSPERO registration number CRD42020209528.

BACKGROUND

COVID-19 is a viral pneumonia caused by a novel coronavirus, initially named 2019 novel coronavirus (2019-nCoV) and subsequently changed to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the International Committee on Taxonomy of Viruses.¹ Initial cases of COVID-19 were identified in Wuhan, Hubei Province, China in December 2019. The epidemic later spread to other countries, reaching Egypt and Africa on 14 February 2020. On 11 March 2020, the disease was declared by the WHO a global pandemic.²

Proper clinical management and control of this pandemic warrant laboratory diagnosis and testing of appropriate specimens from patients meeting the suspected case



definition for COVID-19 as a priority.³ Detection of viral nucleic acid using nucleic acid amplification tests such as reverse transcription-PCR (RT-PCR) is the gold standard for diagnosis of SARS-CoV-2 infection.¹ Real-time RT-PCR assays are characterised by rapid detection and high sensitivity and specificity and hence recommended for diagnosis of early COVID-19 infections.⁴

The RT-PCR assay is complex, time-consuming and associated with risk of eliciting false-negative and false-positive results because it is easily affected by factors such as collection time, sample type and nature of sample preservation.^{5 6} Each PCR test may cost hundreds of dollars and requires the use of sophisticated equipment and expensive reagents.⁷ According to the Ministry of Health in Uganda, each PCR diagnostic test (WHO-approved) costs approximately \$65. This high cost is a potential barrier to majority of the population. Furthermore, this method is unable to meet the principles of early detection, early isolation and early treatment and hence not favourable for prevention and control of the epidemic.⁵

Existing evidence also highlights inconsistencies in the diagnostic accuracy of these assays. More so, most of the evidence on diagnostic accuracy is largely from developed countries, where the COVID-19 curves are flattening. Low-income and middle-income settings are now the epicentre of the pandemic, yet evidence on the diagnostic accuracy of existing tests is largely lacking. This review addresses this knowledge gap on the diagnostic accuracy of available assays to further strengthen the role of testing in the COVID-19 response in these settings.

Rationale

According to a systematic review and meta-analysis of articles on diagnostic accuracy from China, Denmark, Italy, Japan, Spain, Sweden, UK, USA and Germany, the pooled sensitivity of ELISA measuring IgG or IgM was 84.3%, for lateral flow immunoassays was 66.0% and for chemiluminescent immunoassays was 97.8%.⁸ In the same study, the pooled specificity ranged from 96.6% to 99.7%.⁸ In a similar meta-analysis of studies from North and South America, Europe and China, the average sensitivity of rapid antigen tests was 56.2% and the average specificity was 99.5%.⁹ In the same study, the average sensitivity of rapid immunoassays was 95.2% and the specificity was 98.9%.⁹ Based on the findings of these review studies, the diagnostic accuracy of these assays varies and remains questionable. Also, these reviews may not be used to depict the diagnostic accuracy of assays in low-income and middle-income countries (LMICs). Therefore, there is a need to review the diagnostic test accuracy of these tests in LMICs as they are key in the fight against the pandemic.

According to Fryback and Thornbury,¹⁰ it is necessary to assure the efficacy of a diagnostic technique at six levels. This involves determining the technical quality (does the test measure what it purports to measure?), diagnostic accuracy (sensitivity and specificity of the test), diagnostic thinking efficacy (does the test help clinicians

Table 1 The PICOST model for the review question

PICOST element	Description
Population/setting	Adults (18 years and above) in LMIC settings as defined by the World Bank.
Intervention/exposure	New index laboratory test; peripheral laboratory testing strategy or mass testing (pooling).
Comparator	Reference tests for COVID-19 (gold standard) and the current standard of testing strategy (centralised and individualised).
Outcome	Types of tests available; diagnostic test accuracy (sensitivity, specificity, predictive values); costs and cost-effectiveness of the tests; relative risk of testing strategy.
Study design	Diagnostic accuracy studies of observational design (cross-sectional, case-control and cohort studies), and diagnostic strategy studies of experimental design or randomised trials on COVID-19 laboratory testing.
Timing of outcome assessment	72 hours.

LMIC, low-income and middle-income country.

come to a diagnosis?), therapeutic efficacy (does it aid in planning treatment?), whether patients benefit from the use of the test, and the societal efficacy (cost-benefit and cost-effectiveness).¹¹ This review therefore seeks to generate evidence-based recommendations that support the effectiveness of testing strategies and the utility of testing in the control and management of COVID-19 in LMICs through a rapid review.

METHODS

The evidence synthesis will be protocol-driven. The protocol is registered with the International Prospective Register of Systematic Reviews (PROSPERO; <https://www.crd.york.ac.uk/prospere/>) and will be published in a peer-reviewed journal after further development following the statement of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for diagnostic test accuracy studies (PRISMA-DTA).¹²

Review question

The review question is: what is the effectiveness of laboratory testing strategy for COVID-19 in hospitals and community populations in LMICs?

Our review will be guided by the following elements of PICOST (population/setting, intervention/exposure, comparator, outcome, study design, timing of outcome assessment) (table 1).

Outcomes

The primary outcome is the diagnostic test accuracy (sensitivity and specificity) of COVID-19 laboratory test methods in LMICs. The secondary outcomes are the types of COVID-19 tests that are available in LMICs, the effect (relative risk) of the testing strategy, and the cost and cost-effectiveness (incremental cost-effectiveness ratio, ICER) of the various COVID-19 testing algorithms.

Eligibility and selection of studies

Studies will be included if they are published in peer-reviewed journals from January 2020 to present; are studies about PCR assay tests for COVID-19 and rapid point-of-care diagnostic tests; studies conducted on adults (18 years and above) in LMIC settings; and observational studies (cross-sectional, case-control and cohort studies), systematic reviews and randomised controlled trials on COVID-19 laboratory testing.

We intend to exclude studies about index COVID-19 tests without a reference standard; clinical COVID-19 diagnosis alone without verification with any laboratory test; modelling studies on COVID-19 testing; manufacturers' brochures on COVID-19 testing; studies on children <18 years as they are an unlikely source of transmission; or COVID-19 laboratory tests not recommended by the WHO.

Data sources

Article search will be performed on the following databases: PubMed, Google Scholar, MEDLINE, Scopus, Web of Science and the WHO Global Index Medicus. Manual searches will be conducted in websites of organisations championing COVID-19 management for grey literature, including but not limited to manufacturers of COVID-19 laboratory tests; Centers for Disease Control and Prevention in Africa, China, Europe and the USA; the WHO; specialised research institutions in Africa, such as the Uganda Virus Research Institute and Kenya Medical Research Institute; and departments of health such as the Ministry of Health in Uganda, South Africa (Southern Africa), Nigeria (West Africa), and Rwanda and Kenya (Eastern Africa). A list of key experts in diagnosis and testing will be developed and contacted to obtain more information on this subject matter.

Search strategy

The search strategy was developed by our information science specialist (AAK). This search strategy was piloted in PubMed to test for precision of appropriate articles retrieved. We will identify additional relevant articles by manually searching the reference list of selected articles, consulting experts in this field, and searching targeted libraries and websites such as Cochrane and COVID-END.

Search terms

We will use the following search terms: COVID-19, 2019-nCoV, novel corona virus disease, Wuhan pneumonia, severe acute respiratory syndrome related corona virus-2, SARS-CoV-2 and corona virus disease-19. We will also use

the following Medical Subject Heading (MeSH) terms to identify the tests: testing, tests, diagnosis, diagnostics, COVID-19 'point of care tests', Wuhan corona virus tests, laboratory test, corona virus tests and corona virus testing. The search will be limited to LMICs and the search terms will be combined using Boolean operators (AND, OR, NOT) in the electronic search engines.¹³ This search string from PubMed will be adapted to the syntax of other targeted databases for this review (online supplemental file 1).

Data management, screening and selection

The EndNote software will be used for the initial management of references of the search results. These will later be exported to online open access review management software for screening, coding and analysis. The retrieved articles will be exported to EndNote and duplicates will be removed. The studies will then be screened in duplicate following a priori criteria for eligibility (online supplemental file 2). The screening will be performed independently by two review team pairs (OKO, KE, NL and EN), and any disagreements between the reviewers will be resolved by consensus, with further disagreements referred to a tie breaker (EAO or MO).

Data abstraction and coding

The data abstraction form will be developed in an Excel 2007 spreadsheet. The coding process will be performed independently by two research team members (OKO, KE, NL and EN), whose results will be reconciled. Disagreements will be resolved through discussion, and later independent senior reviewers (EAO and MO) will validate the results for quality control and assurance to ensure completeness and correctness.

The following data will be extracted from the articles in a table format: author, year of publication, author affiliation, study design, funding source and other PICOST items, as shown in [table 1](#). The outcome data items are the types of tests available, diagnostic test accuracy (sensitivity, specificity, predictive values), costs and cost-effectiveness of the tests, and relative risk of the testing strategy (online supplemental file 3).

Framework for review synthesis

Our review will be guided by the Fryback and Thornbury¹⁰ framework to establish diagnostic test efficacy, focusing on three levels. These are 'technical efficacy', 'diagnostic accuracy efficacy' and 'societal efficacy'. This six-tiered model is a continuum for diagnostic test efficacy and assesses the effectiveness of laboratory testing strategy for COVID-19 among hospitals and community populations in LMICs. The other levels are 'diagnostic thinking efficacy', 'therapeutic efficacy' and 'patient outcome efficacy' and are less applicable to this review.

Briefly, the following are the three levels of interest: (1) Technical efficacy concerns physical parameters describing the technical quality of a diagnostic test. These are derived under optimal laboratory conditions



and are prerequisites to consideration of efficacy at all subsequent levels. These include the turnaround time, type of the sample and diagnostic test algorithm, that is, single test or series of tests. (2) Diagnostic accuracy efficacy is characterised by the yield of abnormal or normal diagnoses in a case series. This will be measured as a percentage of the correct diagnoses in the case series, the positive and negative predictive values, and the sensitivity and specificity of a given COVID-19 laboratory diagnostic test. (3) Societal efficacy goes beyond the individual risk and benefit of a given COVID-19 test and denotes the cost borne by the society as whole for the diagnostic test to be acceptable for use regardless of the efficacy of the test on individual patient application at any other level. We will estimate whether a given COVID-19 laboratory test is efficacious to an extent that it is an efficient use of resources and provides medical benefits to the society given the low-income and middle-income setting. We will calculate the cost per unit output (measures from level 1 to 6) of a given COVID-19 diagnostic test and the cost-effectiveness by calculating the ICER as a difference between the costs of two given COVID-19 laboratory tests divided by the difference in their effects (measures from level 2).

To determine the relative risk/effect of the testing strategy, we will conduct regression analysis with a random effects model and estimate the relative risk ratios to identify the types of strategies which are associated with optimal strategies associated with optimal specificity and sensitivity cut-offs. Relative risk ratios and CIs will be reported.

Data synthesis

The syntheses will be in the form of summary of findings tables, simple graphs and forest plots, as applicable, using STATA V.15. The Fryback and Thornbury framework¹⁰ will guide this synthesis. First, a structured narrative synthesis of the results will be conducted. This will describe the types of data available, including the tests and the study design. Second, the quantitative synthesis will be outcome-based considering the primary outcome (diagnostic test accuracy of COVID-19 laboratory tests) and the secondary outcomes (costs, cost-effectiveness, turnaround times and the diagnostic testing strategy: centralised versus peripheral; and targeted individual testing versus pooling of samples for scale-up). We will use mixed effects model with the Duckworth-Lewis-Stern method to calculate the overall target score for accuracy. Reporting of these findings will be in line with the PRISMA-DTA statement.¹²

Risk of bias assessment

Two reviewers (EN, OKO, NL or KE) will independently evaluate the methodological quality using the Quality Assessment of Diagnostic Accuracy Studies approach (QUADAS-2 tool).¹⁴ Bias will be assessed by making judgements (high, low and unclear) on individual elements from five domains (selection bias, attrition bias, performance bias, reporting bias, detection bias and other

biases, ie, conflict of interest). Any disagreements will be resolved through discussion and involvement of a senior reviewer (MO or EAO).

Publication bias

All included articles will be assessed for publication bias based on the asymmetry of the funnel plot and/or Egger's test,¹⁵ as appropriate; these are simple rank-based data augmentation techniques which have been proven to be accurate in assessing publication bias due to missing studies.¹⁶ We will plot funnel plots and use the symmetry of the plots to detect the likelihood of publication bias among the articles included in the review. Graphically, in the absence of missing studies, the shape of the scatter plot resembles a symmetrical inverted funnel with a wide base and a narrow top. The presence of large 'holes'—most often seen close to the bottom—or asymmetry in the plot indicates publication bias, but could also be explained by other factors such as study heterogeneity.

Heterogeneity

To assess the level of statistical heterogeneity in the articles, I^2 statistics will be used.²² ¹⁷ The I^2 statistics will indicate percentage (%) heterogeneity that can be attributed to between-study variance. An I^2 of 25% indicates low heterogeneity, I^2 of 50% moderate heterogeneity and I^2 of 75% high heterogeneity. Subgroup analysis will be done on articles with low and moderate heterogeneity.

Quality assessment

To assess the quality of evidence from the reviews, we will use AMSTAR 2 (A Measurement Tool to Assess Systematic Reviews), which is a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions or both.¹⁸ The tool contains 10 domains against which the articles are assessed for quality. The overall quality of evidence will be assessed using a modified Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, where we will assign certainty of evidence ratings for the outcome variables listed above based on an approach developed by the GRADE Working Group¹⁹ and will be done in duplicate, with any disagreements resolved by consensus.

Ethics approval and consent to participate

The review protocol was reviewed and approved by the Makerere University School of Biomedical Sciences Institutional Review Board and the Uganda National Council for Science and Technology.

Patient and public involvement

There was no patient and public engagement in the design, interpretation or dissemination of the findings nor will it be required in this review since it will use already published data.

Author affiliations

¹Department of Medicine, Clinical Epidemiology Unit, Makerere University College of Health Sciences, Kampala, Uganda

²Africa Centre for Systematic Reviews and Knowledge Translation, Makerere University College of Health Sciences, Kampala, Uganda

³Department of General Medicine, Kairos Hospital, Kampala, Uganda

⁴Department of Pharmacology, College of Medicine and Health Sciences, King Ceasor University, Kampala, Uganda

⁵Cicely Saunders Institute of Palliative Care, Policy & Rehabilitation, King's College Hospital, London, UK

⁶School of Public Health, Rakai Health Sciences Program, Makerere University College of Health Sciences, Kampala, Uganda

⁷Department of Public Health, Institute of Medicine, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden

⁸Laboratory, Infectious Diseases Research Collaboration, Kampala, Uganda

⁹Department of Medicine, Makerere University, Kampala, Uganda

¹⁰Department of Medicine, Regional East African Community Health (REACH) Policy Initiative, Makerere University College of Health Sciences, Kampala, Uganda

¹¹Directorate for ICT Support, Makerere University College of Health Sciences, Kampala, Uganda

¹²Albert Cook Library, Makerere University College of Health Sciences, Kampala, Uganda

¹³Department of Medicine, Makerere University College of Health Sciences, Kampala, Uganda

¹⁴Pathogen Genomics, Phenotype and Immunity, Uganda Virus Research Institute, Entebbe, Uganda

¹⁵Department of Pharmacology and Therapeutics, Makerere University College of Health Sciences, Kampala, Uganda

¹⁶Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK

Twitter Brenda Allen Kawala @KawalaBrenda, Daniel Semakula @Dansemakula and Moses Ocan @MosesOcan

Acknowledgements We would like to acknowledge and thank MO and EAO for their guidance and unwavering support throughout the entire process of writing this protocol. This work would not be possible without accessing funding from the Africa Centre for Systematic Reviews and Knowledge Translation, College of Health Sciences, Makerere University, and the Clinical Epidemiology Unit, Makerere University, Kampala, Uganda.

Contributors Conception of the work: MO, EAO and NS. Acquisition of data: OKO, KE, NL, EN, FN, RN, RNW, BAK, TK, AEO, RA, DS, AL, AAK, NS, SNB, MO and EAO. Drafting the work: EN, NL, OKO and KE. Final approval: all authors. OKO, KE, NL, NE, FN, RN, RNW, BAK, TK, AEO, RA, DS, AL, AAK, NS, SNB, MO and EAO reviewed and approved the final manuscript.

Funding This study is funded by the Makerere University Research and Innovation Fund (MakRIF-COVID-19 fund). The funder had no role in developing the protocol.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which

permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Rachel Nante Wangi <http://orcid.org/0000-0002-6589-2535>

Thomas Katairo <http://orcid.org/0000-0002-2002-8911>

Daniel Semakula <http://orcid.org/0000-0002-0806-213X>

Moses Ocan <http://orcid.org/0000-0002-8852-820X>

REFERENCES

- World Health Organization. Diagnostic testing for SARS-CoV-2: interim guidance, 2020. Available: <https://www.who.int/publications/i/item/diagnostic-testing-for-sars-cov-2> [Accessed 11 Sept 2020].
- World Health Organization. Timeline of WHO's response to COVID-19, 2020. Available: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline>
- World Health Organization. Laboratory testing for coronavirus disease 2019 (COVID-19) in suspected human cases: interim guidance, 2 March 2020, 2020. Available: <https://apps.who.int/iris/handle/10665/331329>
- Tahamtan A, Ardebili A. Real-time RT-PCR in COVID-19 detection: issues affecting the results. *Expert Rev Mol Diagn* 2020;20:453–4.
- Yan M, Zheng Y, Sun Y, *et al*. Analysis of the diagnostic value of serum specific antibody testing for coronavirus disease 2019. *J Med Virol* 2021;93:441–7.
- Bruce EA, Huang M-L, Perchetti GA. Direct RT-qPCR detection of SARS-CoV-2 RNA from patient nasopharyngeal swabs without an RNA extraction step. *bioRxiv* 2020.
- Yeung P. Senegal to trial \$1 speedy test for covid-19. *New Sci* 2020;246:13.
- Bastos ML, Tavaziva G, Abidi SK. Diagnostic accuracy of serological tests for covid-19: systematic review and meta-analysis. *Bmj* 2020;370.
- Dinnes J, Deeks JJ, Adriano A, *et al*. Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database Syst Rev* 2020;8:CD013705.
- Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991;11:88–94.
- Sun F, Bruening W, Erinoff E. Addressing challenges in genetic test evaluation: evaluation frameworks and assessment of analytic validity, 2011. <https://www.ncbi.nlm.nih.gov/books/NBK56750/>
- Salameh J-P, Bossuyt PM, McGrath TA, *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.
- Scells H, Zuccon G, Koopman B. Automatic Boolean query refinement for systematic review literature search. *The world wide web conference*, 2019.
- Whiting PF, Rutjes AWS, Westwood ME, *et al*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
- Egger M, Davey Smith G, Schneider M, *et al*. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629–34.
- Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000;56:455–63.
- Higgins JPT. Commentary: heterogeneity in meta-analysis should be expected and appropriately quantified. *Int J Epidemiol* 2008;37:1158–60.
- Shea BJ, Reeves BC, Wells G, *et al*. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017;358:j4008.
- Schünemann HJ, Cuello C, Akl EA, *et al*. Grade guidelines: 18. How robins-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. *J Clin Epidemiol* 2019;111:105–14.