Effectiveness of mass testing for control of COVID-19: a systematic review protocol

Luis Carlos Lopes-Júnior,1,1 Emiliana Bomfim,2 Denise Sayuri Calheiros da Silveira,3 Raphael Manhães Pessanha,1 Sara Isabel Pimentel Carvalho Schuab,1 Regina Aparecida Garcia Lima4

ABSTRACT

Introduction Since March 2020, when the COVID-19 outbreak has been deemed a pandemic by the WHO, the SARS-CoV-2 spreading has been the focus of attention of scientists, authorities, public health agencies and communities around the world. One of the great concerns and challenges, mainly in low-income and middle-income countries, is the identification and monitoring of COVID-19 cases. The large-scale availability of testing is a fundamental aspect of COVID-19 control, but it is currently the biggest challenge faced by many countries around the world. We aimed to synthesize and critically evaluate the scientific evidence on the influence of the testing capacity for symptomatic individuals in the control of COVID-19.

Methods and analysis A systematic review will be conducted in eight databases, such as Medical Literature Analysis and Retrieval System Online, ISI-Web of Knowledge, Cochrane Central Register of Controlled Trials, Embase, SCOPUS, Latin American and Caribbean Health Sciences Literature, PsycINFO and Chinese National Knowledge Infrastructure, from inception to 30 July 2020. No restriction regarding the language, publication date or setting will be employed. Primary outcomes will include the sensitivity as well as the specificity of the tests for COVID-19. Study selection will follow the Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist. Methodological assessment of the studies will be evaluated by the Cochrane Risk-of-Bias tool for randomised controlled trials, the MINORS for non-randomised studies and the Newcastle-Ottawa Scale for cohort or case–control studies. Findings will be structured according to the test type and target population characteristics and focused on the primary outcomes (sensitivity and specificity). Moreover, if sufficient data are available, a meta-analysis will be performed. Pooled standardised mean differences and 95% CIs will be calculated. Heterogeneity between the studies will be determined by I² statistics. Subgroup analyses will also be conducted. Publication bias will be assessed with funnel plots and Egger’s test. Heterogeneity will be explored by random effects analysis.

Ethics and dissemination Ethical approval is not required. The results will be disseminated widely via peer-reviewed publication and presentations at conferences related to this field.

PROSPERO registration number CRD42020182724.

INTRODUCTION

In December 2019, an increased number of pneumonia-like cases in Wuhan, China, led to the discovery of a new type of coronavirus—an enveloped RNA virus commonly found in humans and capable of causing respiratory, enteric, liver as well as neurological illness.1 Despite the low lethality of COVID-19, approximately 3%, its transmissibility is high,1 with respiratory contact droplet being the main means of spreading the new coronavirus.2 Since the WHO declared the COVID-19 outbreak a pandemic on 11 March 2020,3 the spread of the new coronavirus has been the focus of attention of scientists, authorities, public health agencies, government officials and communities around the world.4

Using a networked metapopulation dynamics and Bayesian inference models to
gather epidemiological factors associated with COVID-19, a recent study on SARS-CoV-2 infections in China showed that unreported infections were projected to be 55% as contagious as documented infections, per person. Besides, unreported cases were the source of infection for 79% of reported cases. A total of 213 countries, territories or areas have reported confirmed cases of SARS-CoV-2, with 8,914,787 infected and 466,718 deaths recorded as of 20 June 2020, with Brazil being the new epicentre of the pandemic, with 1,070,139 confirmed cases and 50,058 deaths so far.

One of the greatest concerns and challenges in several countries, especially low-income and middle-income countries, refers to the identification of cases. Identification platforms have undergone modifications in recent months. In addition, the coexistence of several criteria and platforms can generate serious failures in the health surveillance system, resulting in under-reporting. Indeed, the main reason for the problem with how health surveillance is being performed in several countries is the low capacity for mass testing.

Another crucial issue that the WHO has pointed out is that testing all suspected cases is essential for pandemic control. However, access to diagnostic tests remains a challenge globally, in addition to the confusion among health professionals and the population about prioritising tests and interpreting results. The limited availability of diagnostic tests and laboratory capacity for the detection of COVID-19 in many countries, for example, in Brazil, has led the Ministry of Health to limit testing only for severe cases. The Ministry of Health justified its decision by stating that, in mild cases, it does not matter if the person tests negative or positive; the treatment to be delivered is the same as if it was a suspected mild case.

It should be noted that the incubation period from infection to the appearance of the first symptoms is typically 5–7 days but up to 14 days. The final diagnosis depends on tests to detect viruses in several body fluids. Nasopharyngeal smears are more sensitive than oropharyngeal smears and are more effective at early stages of symptom development. However, the gold standard test is the detection of viral RNA by reverse-transcriptase PCR. New methods are being evaluated for faster detection of major viral sequences, and a variety of antigen detection devices have been developed; however, their performance varies widely. In South Korea, for instance, mass testing programmes, contact tracking and isolation contributed to early infection control. As the pandemic progresses, the attention is on asymptomatic patients and health professionals who are on the frontline of the COVID-19 response. Testing symptomatic patients can provide information about contact tracing, besides control and prevention of potential new infections.

Based on consolidated official data, Our World in Data raises some questions that are quite relevant in terms of differences in testing capacity. Comparing countries by their testing capacity per thousand inhabitants, there are notable differences between countries. The USA has already tested 27 784 614 individuals as of 20 June 2020, that is, 83.9 per 1000 inhabitants. However, Brazil has tested 2 409 830 individuals to date, 11.3 per 1000 inhabitants. In other words, currently, the USA has a testing capacity 7.4 times greater than that of Brazil.

With only symptomatic testing, it will be difficult to isolate patients and quarantine communicants. Thus, increasing the production of diagnostic kits and laboratory capacity are urgent issues in Brazil as well as in low-income and middle-income countries. It is hypothesised that a significant increase in large-scale testing capability would be an important advance in the control of COVID-19 in Brazil and other countries, as this is currently the biggest challenge faced by many countries around the world. Hence, this systematic review protocol, adhering to the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) reporting standards, proposes a reproducible strategy to query the scientific literature on the effectiveness of mass testing for the control of COVID-19.

**RESEARCH AIMS**

The purpose of this systematic review is to synthesise and critically evaluate the scientific evidence on the influence of the testing capacity for symptomatic individuals in the control of COVID-19.

**METHODS AND ANALYSIS**

**Search strategy**

The search strategy will be performed using resources to enhance methodological transparency and improve the reproducibility of the findings, following the PRISMA-P guidelines. In addition, using the Population/Intervention/Comparison/Outcomes (PICO) approach, we elaborated the research question of this review to ensure a systematic search of the literature: ‘What is the scientific evidence from studies about the influence of the testing capacity for symptomatic patients in COVID-19 pandemic control?’. The protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) in April 2020 (registration ID: CRD42020182724).

Article searches will be conducted in the following specialised and general databases: Medline, Embase, Cochrane Controlled Trials, PsycINFO, Web of Science, and Scopus. The grey literature will be searched in five additional sources: ProQuest Dissertations and Theses Global, Wotro, Effective Public Health Practice Projects, Public Health Literature Sources and Health Evidence. No restriction regarding the publication date, setting or language will be considered in
this systematic review. Additionally, secondary searches in other sources, such as the clinical trials website (eg, ClinicalTrials.gov), The British Library and Google Scholar, will also be performed. The reference sections of the included studies and cited studies will be manually searched for additional relevant studies. The search strategy will comprise only key terms according to a pre-established PICO strategy. Two researchers (LCL-J and EB) will independently carry out the search in all databases. Additionally, the bibliographic software EndNote (https://www.myendnoteweb.com/) as well as the Rayyan app (Qatar Computing Research Institute) will be used to store, organise and manage all the references and ensure a systematic and comprehensive search.

First, we will identify the existence of a specific subject heading index in each database (including Medical Subject Headings (MeSH) terms, Emtree terms, PsycINFO Thesaurus and DeCS-Health Science Descriptors) and their synonyms (keywords). The search terms will be combined using the Boolean operators ‘AND’ and ‘OR’. The search strategy combining MeSH terms and keywords that will be used in MEDLINE is depicted in Table 1; it will be adapted to meet each database’s specific syntax requirements.

### Table 1: Concepts and search items

<table>
<thead>
<tr>
<th>Databases</th>
<th>Search strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>CENTRAL, Cochrane Central Register of Controlled Trials; CNKI, Chinese National Knowledge Infrastructure; EMBASE, Excerpta Medica dataBASE; LILACS, Latin American and Caribbean Health Sciences Literature; MEDLINE, Medical Literature Analysis and Retrieval System Online; MeSH, Medical Subject Headings; PsycINFO, Psychology Information.</td>
<td></td>
</tr>
<tr>
<td>#2 (&quot;Coronavirus&quot; [MeSH Terms] OR “Coronavirus” [All Fields])</td>
<td></td>
</tr>
<tr>
<td>#4 #1 AND #2 AND #3</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: Inclusion and exclusion criteria

<table>
<thead>
<tr>
<th>PICO component</th>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population (P)</td>
<td>Infant, child, adolescents, young adult, adult and aged (according to MeSH terms)* of all sexes, of any ethnicity and symptomatic and/or suspect for COVID-19.</td>
<td>– Testing for other previous pandemics.</td>
</tr>
<tr>
<td>Intervention/exposure (I)</td>
<td>Testing for COVID-19.</td>
<td>–</td>
</tr>
<tr>
<td>Comparison (C)</td>
<td>Individuals symptomatic for COVID-19 who have not been tested.</td>
<td>–</td>
</tr>
<tr>
<td>Outcome (O)</td>
<td>The primary outcomes include the sensitivity as well as the specificity of the tests.</td>
<td>–</td>
</tr>
</tbody>
</table>

*In this systematic review, we will use definitions in accordance with the MeSH term indexing, such as ‘Infant’: a child between 1 and 23 months of age; ‘Child, Preschool’: a child between the ages of 2 and 5 years; ‘Child’: a person 6–12 years of age; ‘Adolescent’: a person 13–18 years of age; ‘Young Adult’: a person between 19 and 24 years of age; ‘Adult’: a person having attained full growth or maturity. Adults are 19–64 years of age; ‘Aged’: a person 65–79 years of age; ‘Aged, 80 and over’: a person 80 years of age and older. MeSH, Medical Subject Headings.
Study selection

The PICO strategy is detailed in table 2. Regarding the study design, we will include all studies with quantitative approaches (descriptive, observational and experimental studies), as well as the grey literature (editorials, opinion articles, reviews, clinical guidelines, conference proceedings, abstracts, book chapters and so on) as recommended by the Cochrane Handbook.26

Thus, studies that have investigated epidemiological and clinical aspects of testing capacity for symptomatic and suspected patients with COVID-19 will be included in this systematic review. Nevertheless, studies evaluating mass testing for severe acute respiratory syndromes other than COVID-19 will be excluded. With regard to population characteristics, people living in the community and in nursing homes, outpatients and hospitalised people will be included.

The primary outcomes of this systematic review include the sensitivity as well as the specificity of the tests for COVID-19. The sensitivity of a test corresponds to the probability of ‘true positive’. In other words, it indicates the percentage of people with the disease that correctly tested positive. Therefore, a test is highly sensitive if it identifies the actual positive cases that are clinically identified as such.27 The specificity of a test corresponds to the probability of a ‘true negative’. It indicates the true percentage of people who did not have the disease that correctly tested negative.27 These terms describe the performance characteristics of a test and can be used to gauge the effectiveness and validity of a test result.28,29

The screening and selection of studies will be carried out by two reviewers (LCL-J and EB) independently and blindly. After this selection, a third reviewer (RAGL) will be responsible for analysing and deciding on the inclusion or exclusion of each article, especially in relation to those about which there is a conflicting decision. The Rayyan application, developed by the Qatar Computing Research Institute,24 will be used as an auxiliary tool for data management.

Screening

After importing documents retrieved from the initial searches, duplicates will be removed, and two reviewers (LCL-J and EB) will independently screen the studies based on their titles and abstracts. If good agreement is achieved between reviewers (at least 80%), then each will proceed to full article screening. If there is less than 80% agreement, the articles will be reevaluated, and the disagreements will be discussed and resolved by consensus; if a disagreement persists, a third reviewer (RAGL) will make a final decision using the Rayyan app.

Data extraction

Full-text screening will be performed by the same independent investigators. To measure intercoder agreement during each screening phase, Cohen’s kappa will be performed. Once consensus is reached on the selected studies, a standardised form based on previous studies26–34 will be used for data extraction. The information to be extracted includes four domains: (1) identification of the study (article title, journal title, impact factor, authors, country of the study, language, sources of funding, publication year, host institution of the study (hospital, university, research centre, single institution and multicentre study), conflicts of interest and study sponsorship); (2) methodological characteristics (study design, study objective or research question or hypothesis, sample characteristics, eg, sample size, age, eligibility criteria, ethnicity and baseline characteristics, groups and controls, recruitment methods and study completion rates, comparator group, timeframe for follow-up, cointerventions, validated measures, costs and/or remuneration related to participation, statistical analyses and adjustments); (3) main findings and implications for clinical practice; and (4) conclusions. The same two reviewers will independently perform the data extraction. Discrepancies between the reviewers will be resolved either by discussion or, in the lack of agreement, by a third reviewer (RAGL).

Methodological appraisal

The internal validity and risk of bias for RCTs will be assessed with the appraisal tool from the Cochrane Handbook for Systematic Reviews of Interventions 5.1.0,26 which assesses the following seven domains: (I) randomisation sequence allocation; (II) allocation concealment; (III) blinding of participants and team involved; (IV) blindness of outcome evaluators; (V) incomplete outcomes; (VI) report of selective outcome; and (VII) other sources of bias. Based on the evaluation of these domains, studies are classified as at risk of low, high or uncertain bias. For assessing non-randomised controlled trials, the Methodological Index for Non-Randomized Studies (MINORS),35 will be used. This MINORS instrument contains eight items for non-comparative studies: (1) a clearly stated aim; (2) inclusion of consecutive patients; (3) prospective collection of data; (4) endpoints appropriate to the aim of the study; (5) unbiased assessment of the study endpoint; (6) follow-up period appropriate to the aim of the study; (7) loss to follow-up less than 5%; and (8) prospective calculation of the study size.35 With regard to the case–control or cohort studies, we will use the Newcastle-Ottawa Scale to evaluate the methodological quality of the studies.36 Using the Newcastle-Ottawa Scale, the case–control and cohort studies will be given star ratings in three categories: selection (maximum four stars), comparability (maximum two stars) and outcome (maximum three stars), with a maximum score of nine stars.36 The same two reviewers (LCL-J and EB) will conduct the quality assessment independently. Disagreements will be resolved by a third reviewer (RAGL).

Data synthesis

A qualitative synthesis on the RCT risk of bias will be made for the included and analysed studies. The studies will be classified according to the risk of bias as follows: ‘low’ if all the main domains were classified as ‘low risk’; ‘uncertain’
if one or two main domains were classified as ‘uncertain risk’; and ‘high’ if more than two main domains have been classified as ‘uncertain’ or ‘high risk’. When no information is available, we will assign ‘uncertain risk’.\textsuperscript{35} For assessing the non-randomised studies, each item from the MINORS will be rated from 0 to 2, which means that a score of 0 indicates that the information was not reported, 1 indicates that the information was inadequately reported and 2 indicates that the information was adequately reported.\textsuperscript{35} Regarding the case-control and cohort studies assessed by the Newcastle-Ottawa Scale, the quality of these studies will be adjudicated based on a previous study:\textsuperscript{36} good quality: Selection ≥3 stars AND Comparability ≥1 stars AND Outcome ≥2 stars; fair quality: Selection 2 stars AND Comparability ≥1 stars AND Outcome ≥2 stars; poor quality: Selection ≤1 star OR Comparability 0 stars OR ≤1 star.\textsuperscript{38}

In addition, we will complete a narrative synthesis, providing a comprehensive descriptive summary around the type of COVID-19 test, the study design and the target population characteristics that is focused on the primary outcome (the sensitivity as well as the specificity of the tests for COVID-19). In text and table formats, the methodological characteristics of the studies, subpopulation characteristics, test characteristics and sensitivity and specificity of the tests will also be presented. The assessment of the certainty of the evidence will take into consideration the precision of the synthesis findings (ie, CI if available), the number of studies and participants, the consistency of effects across studies, the risk-of-bias of the studies, how directly the included studies address the planned question (directness) and the risk of publication bias.\textsuperscript{39}

Study findings will be presented in tables or graphs in the same way as the syntheses are reported in order to facilitate the comparison of similarities and differences in designs and outcomes among studies. Key characteristics, such as study design, sample size, risk of bias, sensitivity and specificity, which may affect interpretation of the data, will also be presented. Outcomes will be analysed according to sex, population (children, adolescents, young adults, adults and aged) and the type of COVID-19 test and according to the income classification of the countries (high, upper middle, lower middle and low), based on The World Bank Classification using the Gross National Income per capita.\textsuperscript{40}

Meta-analyses will be conducted if there is sufficient homogeneity in study design and study subjects among the selected articles. Therefore, continuous and dichotomous outcomes will be pooled together for meta-analysis purposes. Quantitative data from each study will be extracted and inserted into an Excel sheet by two independent reviewers. Statistical analyses will be carried out using the Statistical Package for the Social Sciences – SPSS, V.18.0.

Standardised mean differences (SMDs) and 95\% CIs will be used to calculate the effect sizes\textsuperscript{41} 42; studies included in our meta-analysis will have reported the differences in methods of testing for COVID-19. All effect sizes will be transformed into a common metric, that is, the bias-corrected standardised difference in means (Hedges’ g), to make them comparable across studies. For continuous outcome measures, SMDs and risk ratios (RRs) for categorical outcomes from individual studies will be considered for the final assessment. The SMD was chosen as a measure of the pooled results considering the likely variability in the measuring scales for continuous outcomes.\textsuperscript{42} The effect size will be interpreted by Cohen’s proposal: 0.20 corresponds to a small effect size, 0.50 corresponds to a medium effect size and 0.80 corresponds to a large effect size.\textsuperscript{33}

A random effects model will be selected under the assumption that the studies included in the meta-analysis were carried out with heterogeneous populations. Heterogeneity will also be tested by the I\textsuperscript{2} statistic, which can quantify the heterogeneity as ranging from 0\% (no heterogeneity) to 100\% (the differences between the effect sizes can completely be explained by chance alone), and the interpretations of the percentages are as follows: 0\%–40\% indicates potentially unimportant heterogeneity, 50\%–60\% indicates moderate heterogeneity, 50\%–90\% indicates substantial heterogeneity and 75\%–100\% indicates considerable heterogeneity.\textsuperscript{42}

To explore the heterogeneity across studies, subgroup analysis will be performed using a mixed effects model according to the following variables: sex, population (children, adolescents, young adults, adults and aged), COVID-19 test type and country income classification (high, upper middle, lower middle and low).

**Patient and public involvement**

Since this is a systematic review protocol, no patients or public will be involved.

**Ethics and dissemination**

Due to the characteristics of this study design, ethical approval was not required. The findings of this systematic review will be disseminated through peer-reviewed publication as well as via different media, such as symposia and conferences related to this field. Moreover, any amendments to this protocol will be documented with reference to the saved searches and analysis methods, which will be recorded in bibliographic databases, for data collection and synthesis.

**DISCUSSION**

In this systematic review protocol, we clearly describe the studies’ designs, participants, interventions and outcomes that will be considered in line with the research question and the data sources, search strategy, data extraction, methodological quality of the studies and data synthesis approach.\textsuperscript{38} In addition, with this protocol study, we reinforce the clarity of the search strategy and minimise the risk of bias.\textsuperscript{44} These results will provide evidence to inform and customise shared decision making to the healthcare providers, stakeholders and government personnel.
Since the sensitivity and specificity of the tests for COVID-19 vary widely by test, this might be the main limitation of this systematic review, followed by the publication bias of the original studies and the methodological appraisal of the studies, which may influence the external validity.

The testing of all symptomatic patients, according to the Imperial College study and the Chinese experience, is essential to contain the epidemic. In a clinical context, although positive tests for COVID-19 are extremely useful, due caution must be taken while interpreting negative tests. Particularly, it must be taken into account the pretest probability of disease. This has important implications for healthcare professionals who interpret tests and policymakers who design diagnostic algorithms for COVID-19. The Chinese handbook of COVID-19 prevention and treatment states “if the nucleic acid test is negative at the beginning, samples should continue to be collected and tested in the following days.” False negatives carry substantial risks; for instance, patients can be transferred to wards not affected by COVID-19, leading to the spread of hospital-acquired COVID-19 infection, and caregivers can also spread the infection to vulnerable dependents. Therefore, guidelines on repeated testing are needed to reduce the risk of false negatives. Finally, physicians must ensure that patients are informed about the limitations of the tests. Patients with a single negative test, but with symptoms that are suggestive of COVID-19, should be advised to isolate themselves according to the guidelines for suspected COVID-19, since no test is 100% accurate.

Hence, this systematic review will deliver relevant evidence on the influence of the testing capacity for symptomatic individuals. Ultimately, we will provide evidence to help the health sector achieve better identification, control and timely monitoring of COVID-19 cases and to guide important strategies and health policy decision makers in several countries.

Contributors LCL-J conceptualised and designed the protocol, drafted the initial manuscript and reviewed the manuscript. All authors defined the concepts and search items, data extraction process as well as methodological appraisal of the studies. EB, DSQUs, RMP, SPSCs and LCL-J planned the data extraction and statistical analysis. RGLE, EB, DSQUs and LCL-J provided critical insights. All authors have approved and contributed to the final written manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Disclaimer The views of the authors do not necessarily reflect those of the NHS, NIHR or the Department of Health.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

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 iD
Luís Carlos Lopes-Júnior http://orcid.org/0000-0002-2424-6510

REFERENCES
44 Silagy CA, Middleton P, Hopewell S. Publishing protocols of systematic reviews: comparing what was done to what was planned. JAMA 2002;287:2831–4.