False-negative results of initial RT-PCR assays for COVID-19: A systematic review

Ingrid Arevalo-Rodriguez1,2*, Diana Buitrago-Garcia3,4, Daniel Simancas-Racines5, Paula Zambrano-Achig6, Rosa Del Campo6, Agustin Ciapponi7, Omar Sued8, Laura Martinez-Garcia2,6, Anne W. Rutjes3,4, Nicola Low3,4, Patrick M. Bossuyt9, Jose A. Perez-Molina10,11, Javier Zamora1,2,12

1 Clinical Biostatistics Unit, Hospital Universitario Ramón y Cajal-IRYCIS, Madrid, Spain, 2 CIBER of Epidemiology and Public Health, Madrid, Spain, 3 Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland, 4 Graduate School for Health Sciences, University of Bern, Bern, Switzerland, 5 Centro de Investigacion en Salud Publica y Epidemiología Clinica (CISPEC), Facultad de Ciencias de la Salud “Eugenio Espejo”, Universidad UTE, Quito, Ecuador, 6 Department of Microbiology, Ramón y Cajal University Hospital, Ramón y Cajal Health Research Institute (IRYCIS), Madrid, Spain, 7 Instituto de Efectividad Clinica y Sanitaria (IECS-CONICET), Buenos Aires, Argentina, 8 Fundación Huesped, Buenos Aires, Argentina, 9 Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam University Medical Centres, University of Amsterdam, Amsterdam, The Netherlands, 10 Infectious Diseases Department, National Referral Centre for Tropical Diseases, Hospital Universitario Ramón y Cajal, Madrid, Spain, 11 Instituto Ramón y Cajal de Investigación Sanitaria, Madrid, Spain, 12 Institute of Applied Health Research, University of Birmingham, Birmingham, United Kingdom

* ingrid.arevalo@salud.madrid.org

Abstract

Background
A false-negative case of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection is defined as a person with suspected infection and an initial negative result by reverse transcription-polymerase chain reaction (RT-PCR) test, with a positive result on a subsequent test. False-negative cases have important implications for isolation and risk of transmission of infected people and for the management of coronavirus disease 2019 (COVID-19). We aimed to review and critically appraise evidence about the rate of RT-PCR false-negatives at initial testing for COVID-19.

Methods
We searched MEDLINE, EMBASE, LILACS, as well as COVID-19 repositories, including the EPPI-Centre living systematic map of evidence about COVID-19 and the Coronavirus Open Access Project living evidence database. Two authors independently screened and selected studies according to the eligibility criteria and collected data from the included studies. The risk of bias was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool. We calculated the proportion of false-negative test results using a multilevel mixed-effect logistic regression model. The certainty of the evidence about false-negative cases was rated using the GRADE approach for tests and strategies. All information in this article is current up to July 17, 2020.
Results

We included 34 studies enrolling 12,057 COVID-19 confirmed cases. All studies were affected by several risks of bias and applicability concerns. The pooled estimate of false-negative proportion was highly affected by unexplained heterogeneity (tau-squared = 1.39; 90% prediction interval from 0.02 to 0.54). The certainty of the evidence was judged as very low due to the risk of bias, indirectness, and inconsistency issues.

Conclusions

There is substantial and largely unexplained heterogeneity in the proportion of false-negative RT-PCR results. The collected evidence has several limitations, including risk of bias issues, high heterogeneity, and concerns about its applicability. Nonetheless, our findings reinforce the need for repeated testing in patients with suspicion of SARS-CoV-2 infection given that up to 54% of COVID-19 patients may have an initial false-negative RT-PCR (very low certainty of evidence).

Systematic review registration


Introduction

Accurate laboratory tests are essential for the diagnosis and management of emerging infectious diseases. On December 31, 2019, the World Health Organization (WHO) was alerted about a cluster of patients with pneumonia in Wuhan City, Hubei province, China [1]. Chinese authorities confirmed, a week later, an outbreak of a novel coronavirus. The virus has been named as severe acute respiratory coronavirus 2 (SARS-CoV-2) [2], and the clinical disease that it causes is coronavirus disease 2019 (COVID-19), which has become a worldwide public health emergency and reached pandemic status [3, 4]. By October 16 2020, there was a total of 39,023,292 confirmed cases, and 1,099,586 confirmed cases worldwide [5].

Clinical suspicion of COVID-19 is based primarily on respiratory symptoms such as fever, cough, and shortness of breath as primary manifestations [6, 7]. The spectrum of symptoms and clinical signs associated with COVID-19 has expanded with increasing knowledge about SARS-CoV-2. Although most of the cases present mild symptoms, some cases have developed pneumonia, severe respiratory diseases, kidney failure or heart failure [8–11]. The death rate from SARS-CoV-2 infection is estimated to be of 0.68% (95% CI from 0.53 to 0.82) [12] SARS-CoV-2 mainly spreads through person-to-person contact via respiratory droplets from coughing and sneezing, and through surfaces that have been contaminated with these droplets [13]. A proportion of cases will, however, remain asymptomatic throughout the course of infection, estimated as around 20% in a range of settings [14, 15].

Because the signs of infection mentioned above are non-specific, confirmation of cases is currently based on the detection of nucleic acid amplification tests that detect viral ribonucleic acid (RNA) sequences by reverse transcription-polymerase chain reaction (RT-PCR) [16]. Different RT-PCR assays have been proposed, all of which include the N gene that codes for the viral nucleocapsid. Other alternative targets are the E gene, for the viral envelope; the S gene for the spike protein; and the Hel gene for the RNA polymerase gene (RdRp/Helicase) [16, 17]. Molecular criteria for in vitro diagnosis of COVID-19 disease are heterogeneous, and usually require the detection of two or more SARS-CoV-2 genes [18].
A person with a negative RT-PCR result at initial testing, with a subsequent positive test result, is considered as a false-negative diagnosis. Clinical practice guidelines and consensus statements recommend repeated RT-PCR testing to confirm a clinical diagnosis, especially in the presence of symptoms associated with COVID-19 [19–23]. Researchers have suggested that these failures in SARS-CoV-2 detection are related to multiple preanalytical and analytical factors, such as lack of standardisation for specimen collection, delays or poor storage conditions before arrival in the laboratory, the use of inadequately validated assays, contamination during the procedure, insufficient viral specimens and load, the incubation period of the disease, and the presence of mutations that escape detection or PCR inhibitors [18, 24].

The availability of accurate laboratory tools for COVID-19 is essential for case identification, contact tracing, and optimisation of infection control measures, as it was shown by previous epidemics caused by SARS-CoV and the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) [25–27]. Due to the significant burden on health systems around the globe caused by the COVID-19 pandemic, and the potential consequences at several levels of missing a COVID-19 case, we aimed to obtain through a systematic review of the literature, a summary estimate of the proportion of false-negatives related to the detection of SARS-CoV-2 using RT-PCR assays at the first healthcare encounter (initial testing).

Materials and methods

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy Studies (PRISMA-DTA) to report this review [28]. This manuscript reflects the third update of our literature searches with information current up to July 2020. The initial review protocol, previous reports of findings by date of search and supplementary material are available in the Open Science Framework repository (https://tinyurl.com/vvbgqya).

Criteria for considering studies for this review

We included observational studies (including diagnostic test accuracy studies) reporting the initial use of RT-PCR to detect SARS-CoV-2 RNA in patients with suspected infection by clinical or epidemiological criteria. We primarily aimed to include studies enrolling consecutive patients who received an RT-PCR test at the first healthcare encounter (initial testing), with further confirmation of SARS-CoV-2 infection of initial negative cases by an additional RT-PCR evaluation. We did not impose limits by age, gender, or study location.

We aimed to include all types of RT-PCR kits, regardless of the brand or manufacturer, the RNA extraction method used, the number of target gene assays assessed, or the cycle threshold value for positivity. We excluded studies that focused on other populations or reporting samples/specimens instead of patients (such as monitoring or discharge of COVID-19 confirmed cases, population screening and patients with comorbidities), studies only providing the absolute number of false-negatives or without clear information about numerical information, and studies reporting validation of novel assays or comparing sample collection/sample specimens (i.e. focus on agreement). Full eligibility criteria can be found in the S1 File.

Search methods for identification of studies

We carried out a comprehensive and sensitive search strategy based on search terms developed for the COVID-19 Open Access Project by researchers and librarians at the Institute of Social and Preventive Medicine, University of Bern (https://ispmbern.github.io/covid-19/living-review/collectingdata.html) in the following databases:

- MEDLINE (Ovid SP, 1946 to July 17, 2020)
Data collection and analysis

For the selection of eligible studies, two out of three reviewers (IAR, DBG or PZA) independently screened the search results based on their titles and abstract. We retrieved the full-text copy of each study assessed as potentially eligible, and two out of three reviewers (IAR, DBG or PZA) confirmed eligibility according to the selection criteria. In case of disagreements, we reached consensus by discussion. For data extraction, one reviewer extracted qualitative and quantitative data from eligible studies, and an additional reviewer checked all the extracted information for accuracy. We contacted study authors to supply missing information about critical characteristics of included studies.

Assessment of methodological quality

Two authors independently assessed the risk of bias of included studies, and disagreements were resolved through discussion. We evaluated the methodological quality using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool [30]. We decided to also apply the QUADAS-2 tool for case series studies due to the lack of tools to assess the risk of bias associated with these studies. However, for a more comprehensive assessment of the limitations of the included studies, we adapted the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Case Series [31]. This tool included items about inclusion criteria, measurement of asymptomatic status, follow-up of the course of the disease, and availability of numerator and denominator. We added questions about the representativeness of the source and target populations as well.

Statistical analysis and data synthesis

For all included studies, we extracted data to calculate the proportion of false-negative results. The numerator (false-negative cases) was the number of cases initially considered negative by RT-PCR. The denominator was the total number of SARS-CoV-2 infections, detected on an
additional test by RT-PCR, using the same or a different assay. We presented the results of estimated proportions (with 95% CIs) in a forest plot to assess the between-study variability. We aimed to calculate a summary estimate of the false-negative rate with the corresponding 95% CI using a multilevel mixed-effect logistic regression model in Stata 16. This method allowed us to estimate the between-study heterogeneity from the variance of study-specific random intercepts. We computed 90% prediction intervals to include the between-study variation. The 90% prediction interval shows the range of true false-negative proportions that can be expected in 90% of future settings, comparable to the ones included in the meta-analysis. We expressed heterogeneity in primary study results using the Tau-square statistic.

We planned to investigate the potential sources of heterogeneity using a descriptive approach and performing a random-effects meta-regression analysis, including covariates, one at each time, into the logistic model. Anticipated sources of heterogeneity included the type of specimen collected, the presence or not of clinical findings, the number of RNA targets genes under assessment, and the time of symptom evolution.

Summary of findings and certainty of the evidence
We rated the certainty of the evidence about false-negative cases following the GRADE approach for tests and strategies [32, 33]. We assessed the quality of evidence as high, moderate, low, or very low, depending on several factors, including risk of bias, imprecision, inconsistency, indirectness, and publication bias. We illustrate the consequences of the numerical findings in a population of 100 tested, according to three different prevalence estimates of the disease provided by the stakeholders involved in this review.

Results
Electronic searches yielded 2536 references from the selected databases. In addition, we obtained 186 additional references searching in other resources (Fig 1). Our initial screening of titles and abstracts identified 171 references to assess in full text. We excluded 137 studies
mostly due to: a) ineligible setting (no initial SARS-CoV-2 testing); b) incomplete or no data about false-negative cases and COVID-19 confirmed cases; c) ineligible population (i.e. pooling sample, analysis based on samples instead of patients) (S3 File). The between reviewers agreement in the selection of references in the title & abstract stage was moderate (kappa statistic = 0.56), while the agreement in the full-text selection was substantial (kappa statistic = 0.75). We included 34 studies in our synthesis [34–67] that dealt with 12057 patients (Fig 1).

The sample sizes ranged from 18 to 5,700 COVID-19 confirmed cases (median 90; interquartile range–IQR = 46.5 to 204). Twelve studies focused on the estimation of diagnostic test accuracy of alternative tests in populations with suspected COVID-19 at the beginning of the study [34, 37–41, 44, 46, 47, 51, 57, 65]. The remaining studies reported information from case series, most of which included confirmed cases of COVID-19 at the beginning of the study [35, 36, 42, 43, 45, 48–50, 52–56, 58–64, 66, 67]. One study focused its data collection only on children [53] and other only on healthcare workers [48].

We included 34 studies in our synthesis [34–67] that dealt with 12057 patients (Fig 1).

The sample sizes ranged from 18 to 5,700 COVID-19 confirmed cases (median 90; interquartile range–IQR = 46.5 to 204). Twelve studies focused on the estimation of diagnostic test accuracy of alternative tests in populations with suspected COVID-19 at the beginning of the study [34, 37–41, 44, 46, 47, 51, 57, 65]. The remaining studies reported information from case series, most of which included confirmed cases of COVID-19 at the beginning of the study [35, 36, 42, 43, 45, 48–50, 52–56, 58–64, 66, 67]. One study focused its data collection only on children [53] and other only on healthcare workers [48].

Data collection from cases ranged from January 1 [58] to April 15, 2020 [41, 48]; two studies did not provide complete information about the time of recruitment [35, 45]. Ten studies were carried out in institutions outside of China [35, 37, 41, 45, 46, 48, 49, 52, 54, 61]. The age of participants was reported in different ways in 21 studies providing information of COVID-19 confirmed cases [34–38, 44, 46, 47, 51, 53–55, 58, 59, 61–65, 67]: for studies reporting a mean, the average ranged from 2.5 [53] to 56 years [58], while for studies reporting medians, the corresponding range was 45 [44] to 63 years [54]. These 21 studies reported a total of 5331 men and 4067 women (Table 1).

In all cases, the presence of infection was confirmed after detection of SARS-CoV-2 RNA in any real-time RT-PCR assay that was repeated after a negative result. The specimens collected for the RT-PCR assessment were heterogeneous in most of the included studies; in 13 studies the authors reported the use of nasopharyngeal swabs [35–38, 45–47, 49, 52, 54, 58, 61, 66], along with oropharyngeal swabs in seven of these 13 studies [35–38, 45–47] (Table 1). The name/brand of the SARS-CoV-2 nucleic acid detection kit used was reported by 19 studies [34–37, 46–52, 55, 56, 58, 59, 61–63, 65], and 13 studies reported the target genes under assessment for positivity [35, 46, 49–52, 55, 56, 58, 59, 61, 63, 65], with the ORF1ab being the most frequently used (8 studies). Ten studies provided heterogeneous information about the time from the symptom onset to initial testing [35, 36, 39, 40, 43, 44, 49, 51, 61, 65] (Table 1).

**Quality of included studies**

After classification with the QUADAS-2 tool (Fig 2 and S4 File), the domain most affected by a high risk of bias was the flow and timing domain, as some studies had not repeated the RT-PCR testing to all patients with negative results at initial testing [37, 45, 46, 52, 54, 55]; besides, some studies did not provide information about the interval of time for the administration of a new RT-PCR assay. In the patient selection domain, the risk of bias and applicability concerns were judged as high or unclear for several studies in which participant selection was driven by prior testing with RT-PCR, Chest CT findings or serology tests. In most of the studies it was unclear if a standard testing protocol was used, or if authors restricted participant sampling to those who had received all test [34–36, 38–41, 47, 48, 51, 53, 57, 58, 60, 64, 67].

In the index test domain, several studies lacked to provide details about the criteria for positive results, such as the target genes under assessment of the SARS-CoV-2 nucleic acid detection kit used, so that their risk of bias and applicability remained unclear [34, 36–45, 47, 48, 50, 51, 53, 54, 57, 60, 64, 66, 67]. Finally, two studies were judged at unclear risk of bias and applicability in the reference standard domain, since the authors did not report the characteristics of the repeated RT-PCR and their administration in sufficient detail [39, 52]. Six studies were
Table 1. Characteristics of included studies.

<table>
<thead>
<tr>
<th>ID</th>
<th>Data collection</th>
<th>Country</th>
<th>Setting</th>
<th>Age (years)</th>
<th>% Male: % Female</th>
<th>Type of specimen</th>
<th>RT-PCR Brand</th>
<th>Target genes</th>
<th>Days from symptoms onset (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ai T 2020 [34]</td>
<td>January 6-February 6</td>
<td>China</td>
<td>Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology, Wuhan, Hubei, China</td>
<td>Mean 51 ± 15</td>
<td>46:54 b</td>
<td>Throat swab</td>
<td>• TaqMan One-Step RT-PCR Kits from Shanghai Huirui Biotechnology Co., Ltd&lt;br&gt;• Shanghai BioGerm Medical Biotechnology Co., Ltd</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Albert E 2020 [35]</td>
<td>Unclear-April 14</td>
<td>Spain</td>
<td>Hospital Clínico Universitario of Valencia</td>
<td>Median 65 years; range from 3 to 98</td>
<td>57:43 c</td>
<td>Nasopharyngeal or oropharyngeal swabs, upper RT samples</td>
<td>• LightMix Modular SARS-CoV-2 (COVID-19) E-gene/ LightMix Modular SARS-CoV-2 (COVID-19) RdRP gene from TIB MOLBIOL GmHD&lt;br&gt;• SARS-CoV-2 Real-time PCR Kit from Vircell Diagnostics&lt;br&gt;• SARS-CoV-2 (S gene)–BD Max System (Viasure Real-Time PCR Detection Kits; CerTest, Zaragoza, Spain).</td>
<td>E, RdRp, S</td>
<td>Median 5 days; range: 1–14 days</td>
</tr>
<tr>
<td>Bernheim A 2020 [36]</td>
<td>January 18-February 2</td>
<td>China</td>
<td>Hospitals from four provinces in China: Nanchang (Jiangxi Province), Zhuhai (Guangdong Province), Chengdu (Sichuan province) and Guilin (Guangxi province)</td>
<td>Mean 45 ±15.6 b</td>
<td>50:50 b</td>
<td>Bronchoalveolar lavage, endotracheal aspirate, nasopharyngeal swab, or oropharyngeal swab</td>
<td>• Sansure Biotech Inc. (Changsha, China), Shanghai Zhijiang Biotechnology Co. (Shanghai, China),&lt;br&gt;• Da An Gene Co. (Guangzhou, China).</td>
<td>Not reported</td>
<td>Range from 0 to 12</td>
</tr>
<tr>
<td>Besutti G 2020 [37]</td>
<td>March 13–23</td>
<td>Italy</td>
<td>AUSL-IRCCS di Reggio Emilia, Reggio Emilia, Italy</td>
<td>Mean 59 ± 15.8 b</td>
<td>59:41 b</td>
<td>Nasopharyngeal and oropharyngeal swabs</td>
<td>GeneFinder™ COVID-19 PLUS Real Real Amp Kit</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Chen D 2020 [38]</td>
<td>January 19-February 20</td>
<td>China</td>
<td>Five non-specialised infectious disease hospitals in Guangzhou</td>
<td>Mean 49.7 ± 15.7 a</td>
<td>43:57 a</td>
<td>Nasopharyngeal or oropharyngeal swabs</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
</tr>
<tr>
<td>Chen HJ 2020 [39]</td>
<td>January 26-February 4</td>
<td>China</td>
<td>Hainan General Hospital</td>
<td>Mean 54.3 ± 11.8 a</td>
<td>68:32 a</td>
<td>Nasopharyngeal or oropharyngeal swabs</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
</tr>
<tr>
<td>Chen ZH 2020 [40]</td>
<td>January 24-February 6</td>
<td>China</td>
<td>The Hangzhou Xixi Hospital Affiliated to Zhejiang Chinese Medical University</td>
<td>Mean 46.9 ± 11.1 a</td>
<td>55:45 a</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Mean 2; range 1 to 4.5 days</td>
<td></td>
</tr>
<tr>
<td>Cinkooglu A 2020 [41]</td>
<td>March 15-April 15</td>
<td>Turkey</td>
<td>Ege University Faculty of Medicine</td>
<td>Means from 39.9 to 51 a</td>
<td>47:53 a</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
</tr>
<tr>
<td>Dai H 2020 [42]</td>
<td>January 10-February 7</td>
<td>China</td>
<td>13 hospitals in Jiangsu</td>
<td>Mean 44.6 ± 14.8 a</td>
<td>58:42 a</td>
<td>Respiratory samples</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>ID</th>
<th>Data collection</th>
<th>Country</th>
<th>Setting</th>
<th>Age (years)</th>
<th>% Male: % Female</th>
<th>Type of specimen</th>
<th>RT-PCR Brand</th>
<th>Target genes</th>
<th>Days from symptoms onset (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duan X</td>
<td>January 10-February 8</td>
<td>China</td>
<td>The First Affiliated Hospital, College of Clinical Medicine, Medical College of Henan University of Science and Technology, Luoyang</td>
<td>Mean 52 ± 19.3 a</td>
<td>60:40</td>
<td>Nasal and oropharyngeal swab specimens</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Mean 6.64 ± 3.82 days</td>
</tr>
<tr>
<td>Fang Y</td>
<td>January 19-February 4</td>
<td>China</td>
<td>Taizhou Enze Medical Center (Group) Enze Hospital, China</td>
<td>Median 45; IQR: 39–55 a</td>
<td>57:43</td>
<td>Throat swab, sputum</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Mean 3±3</td>
</tr>
<tr>
<td>Fechner C</td>
<td>Unclear</td>
<td>Switzerland</td>
<td>Cantonal Hospital Lucerne</td>
<td>Mean 63 ± 15.7 a</td>
<td>75:25</td>
<td>Nasopharyngeal or oropharyngeal swabs</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Gietema</td>
<td>March 13–24</td>
<td>Netherlands</td>
<td>Maastricht University Medical Centre (MUMC+), the Netherlands</td>
<td>Median 66; IQR: 55–76 b</td>
<td>59:41</td>
<td>Nasopharyngeal and/or oropharyngeal swab</td>
<td>• Tib-Molbiol (Berlin, Germany)</td>
<td>• Biologeo (Netherlands)</td>
<td>Not reported</td>
</tr>
<tr>
<td>He JL</td>
<td>January 10 – February 28</td>
<td>China</td>
<td>University of Hong Kong-Shenzhen Hospital, China</td>
<td>Median 52; range: 8 to 74 a</td>
<td>50:50</td>
<td>Nasopharyngeal swab, oropharyngeal swab, endotracheal aspirate, or bronchoalveolar lavage</td>
<td>BGI Genomics (Shenzhen, China)</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Lan FY</td>
<td>March 9-April 15</td>
<td>USA</td>
<td>Massachusetts community healthcare system</td>
<td>Mean 43.6 ± 12.9 b</td>
<td>21:79</td>
<td>Nasopharyngeal swabs</td>
<td>• CDC 2019-Novel RT-PCR</td>
<td>• Roche Cobas SARS-CoV-2</td>
<td>Not reported</td>
</tr>
<tr>
<td>Lee TH</td>
<td>January-February 29</td>
<td>Singapore</td>
<td>National Centre for Infectious Diseases, Singapore</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Nasopharyngeal swabs, sputum, and stool if diarrhoea is present</td>
<td>• Laboratory developed test</td>
<td>• A’STAR Fortitude Kit (Accelerate Technologies, Singapore)</td>
<td>N +ORF1ab</td>
</tr>
<tr>
<td>Li Y 2020</td>
<td>February 2–17</td>
<td>China</td>
<td>Hankou Hospital of Wuhan, China</td>
<td>Median 57; range: 22 to 88 b</td>
<td>56:44</td>
<td>Pharyngeal swab specimens</td>
<td>Shengxiang Biotechnology Co (novel coronavirus 2019-nCoV nucleic acid detection kit (fluorescence PCR method)</td>
<td>ORF1ab d</td>
<td>Not reported</td>
</tr>
<tr>
<td>Long C</td>
<td>January 20-February 8</td>
<td>China</td>
<td>Yichang Yiling Hospital, China</td>
<td>Mean 44.8 ±18.2 a</td>
<td>56:44</td>
<td>Not reported</td>
<td>DAAN GENE d</td>
<td>ORF1ab d</td>
<td>Only duration of fever reported: 2,6 ± 1,7 days</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>ID</th>
<th>Data collection</th>
<th>Country</th>
<th>Setting</th>
<th>Age (years)</th>
<th>% Male: % Female</th>
<th>Type of specimen</th>
<th>RT-PCR Brand</th>
<th>Target genes</th>
<th>Days from symptoms onset (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long DR 2020 [52]</td>
<td>March 2–30</td>
<td>USA</td>
<td>University of Washington Virology clinical laboratory</td>
<td>Means from 56.7 to 61.6 *</td>
<td>57:43 *</td>
<td>Nasopharyngeal swabs</td>
<td>• Laboratory-developed test (LDT) two-target/two-control assay modified from the CDC</td>
<td>N1, N2, ORF1ab, E, S</td>
<td>Not reported</td>
</tr>
<tr>
<td>Ma H 2020 [53]</td>
<td>January 21-February 14</td>
<td>China</td>
<td>Wuhan Children’s Hospital</td>
<td>Mean 2.5; range: 0.9 to 7 *</td>
<td>56:44 *</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Richardson 2020 [54]</td>
<td>March 1-April 4</td>
<td>USA</td>
<td>12 hospitals in New York City, Long Island, and Westchester County, New York (Northwell Health system), USA</td>
<td>Median 63; IQR: 52–75 *</td>
<td>60:40 *</td>
<td>Nasopharyngeal swabs</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Shen N 2020 [55]</td>
<td>January 22-February 18</td>
<td>China</td>
<td>Tongji Hospital in Wuhan</td>
<td>Median 56; IQR: 42–66</td>
<td>49:51</td>
<td>Throat swabs</td>
<td>SARS-CoV-2 nucleic acid detection kit (Shanghai Huirui Biotechnology Co. Ltd)</td>
<td>N + ORF1ab</td>
<td>Not reported</td>
</tr>
<tr>
<td>Wang P 2020 [56]</td>
<td>January 25-March 16</td>
<td>China</td>
<td>First People’s Hospital of Jingmen, Hubei Province</td>
<td>Median 58; range: 21–95</td>
<td>46:54</td>
<td>Throat swabs</td>
<td>RT-PCR reagent BioGerm (Shanghai BioGerm Medical Technology Co., Ltd.)</td>
<td>N + ORF1ab</td>
<td>Not reported</td>
</tr>
<tr>
<td>Wen Z 2020 [57]</td>
<td>January 21-February 14</td>
<td>China</td>
<td>Two areas in Henan Province, China</td>
<td>Median 16; range: 12 to 98 *</td>
<td>47:53 *</td>
<td>throat-swab, sputum, or alveolar lavage fluid specimens</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Wong HYF 2020 [58]</td>
<td>January 1-March 5</td>
<td>China</td>
<td>Four tertiary and regional hospitals in Hong Kong (Queen Mary Hospital, Pamela Youde Nethersole Eastern Hospital, Queen Elizabeth Hospital, and Ruttonjee Hospital), China</td>
<td>Mean 56; range: 16 to 96 *</td>
<td>41:59 *</td>
<td>nasopharyngeal swabs and throat swabs</td>
<td>QuantiNova Probe RT-PCR Kit (QIAGEN, Hilden, Germany)</td>
<td>RdRp</td>
<td>Not reported</td>
</tr>
<tr>
<td>Wu J 2020 [59]</td>
<td>January 22-February 14</td>
<td>China</td>
<td>First People’s Hospital of Yancheng City, the Second People’s Hospital of Yancheng City, and the Fifth People’s Hospital of Wuxi, China</td>
<td>Median 46.1; IQR: 30.7 to 61.5</td>
<td>49:51</td>
<td>nose swab and/or throat swab</td>
<td>Bio-germ, Shanghai, China</td>
<td>N + ORF1ab</td>
<td>Not reported</td>
</tr>
</tbody>
</table>

(Continued)
considered as at low risk of bias in all QUADAS-II domains assessed [49, 56, 59, 61, 62, 65], while 20 were considered as at unclear risk due to at least one domain was judged with unclear risk of bias. The remaining eight studies were considered high risk of bias as at least one domain was judged at high risk [37, 38, 46, 51, 52, 54, 55, 63].

The assessment of limitations using the adapted JBI case-series tool provided a similar picture owing to the uncertainty regarding the consecutive inclusion of patients and follow-up time after the first RT-PCR result (S4 File). Additionally, due to the selection of patients, the majority of included studies were not an adequate sample of the target population (S4 File).

Table 1. (Continued)

<table>
<thead>
<tr>
<th>ID</th>
<th>Data collection</th>
<th>Country</th>
<th>Setting</th>
<th>Age (years)</th>
<th>% Male: % Female</th>
<th>Type of specimen</th>
<th>RT-PCR Brand</th>
<th>Target genes</th>
<th>Days from symptoms onset (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xie X 2020</td>
<td>January 16-February 2</td>
<td>China</td>
<td>Database of Radiology Quality Control Centre, Hunan/3 cities in Hunan Province, China</td>
<td>Not reported</td>
<td>Not reported</td>
<td>swab test; no further details provided</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Young BE 2020 [61]</td>
<td>January 23-February 3</td>
<td>Singapore</td>
<td>4 hospitals in Singapore</td>
<td>Median 47; range: 31–73 a</td>
<td>50:50 a</td>
<td>Nasopharyngeal swabs</td>
<td>QuantiTect Probe RT-PCR kit (Qiagen)</td>
<td>N, S, ORF1ab</td>
<td>Median 13; range 5–24 days</td>
</tr>
<tr>
<td>Zhang H 2020 [62]</td>
<td>January 22-February 28</td>
<td>China</td>
<td>Huagang Central Hospital and The Second Affiliated Hospital of Shandong First Medical University</td>
<td>Median 48.3; IQR: 33–56 a</td>
<td>56:44 a</td>
<td>Not reported</td>
<td>The Beijing Genomics Institute (BGI, Beijing, China)</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Zhang JJ 2020 [63]</td>
<td>December 29-February 16</td>
<td>China</td>
<td>Zhongnan Hospital of Wuhan University and No.7 hospital of Wuhan, China</td>
<td>Median 57; range: 22 to 88 a</td>
<td>53:47 a</td>
<td>Pharyngeal swab</td>
<td>Shanghai bio-germ Medical Technology Co Ltd</td>
<td>N +ORF1ab</td>
<td>Not reported</td>
</tr>
<tr>
<td>Zhao JJ [64]</td>
<td>January 11-February 9</td>
<td>China</td>
<td>Shenzhen Third People’s Hospital</td>
<td>Median 48; IQR: 35–61 a</td>
<td>49:51 a</td>
<td>Throat swabs, Nasal swabs</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Zhifeng 2020 [65]</td>
<td>January 25-February 6</td>
<td>China</td>
<td>Xiaogan Central Hospital, China</td>
<td>Range: 23 to 82 a</td>
<td>59:41 a</td>
<td>Throat swabs</td>
<td>Multiple brands d</td>
<td>N +ORF1ab</td>
<td>Mean 6,5 days d</td>
</tr>
<tr>
<td>Zhou H 2020 [66]</td>
<td>January 19-February 15</td>
<td>China</td>
<td>First Affiliated Hospital, Zhejiang University School of Medicine</td>
<td>Mean 53.3; range: 14–96 a</td>
<td>59:41 a</td>
<td>Bronchial lavage, endotracheal aspirate, or nasopharyngeal swab</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Zhou S 2020 [67]</td>
<td>January 16-February 12</td>
<td>China</td>
<td>Tongji Hospital of Tongji Medical College, Huazhong University of Science and Technology</td>
<td>Mean 52.3 ± 13.1 a</td>
<td>54:46 a</td>
<td>Pharyngeal swab</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
</tbody>
</table>

Notes:

a) Information from COVID-19 confirmed cases only;

b) Information from COVID-19 suspected (positive and negative);

c) Information from other groups reported by the authors;

d) Data provided by the corresponding author (personal communication).

CDC: Center for Disease Control and Prevention; LDT: Laboratory-developed test; ORF1ab: Open Reading Frame 1ab; RdRp gene: RNA polymerase gene; RT-PCR: Reverse transcription polymerase chain reaction; SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2.

https://doi.org/10.1371/journal.pone.0242958.1001
Findings

We analysed information from 34 studies collecting information from 12,057 patients confirmed to have SARS-CoV-2 infection and 1060 cases with RT-PCR negative findings in their initial assessment. False-negative rates ranged from 0.018 [45] to 0.58 [57], with a median of 0.11 (Fig 2).

The summary estimate of the false-negative rate was 0.13 (95% CI 0.09 to 0.19). The data were characterised by a considerable between-study heterogeneity, the 90% prediction interval ranged from 0.02 to 0.54 (tau-squared = 1.39).

Assessment of the effect of potential sources of heterogeneity was limited because stratified information for relevant subgroups was not available in most studies (Table 2). There were no

Table 2. Assessment of sources of heterogeneity.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of studies (patients)</th>
<th>Heterogeneity (Tau-squared)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days of symptoms (average/median)</td>
<td>Less than 5 days</td>
<td>3 (120)</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Five days or more</td>
<td>6 (817)</td>
<td>0.87</td>
</tr>
<tr>
<td>PCR target</td>
<td>N gene</td>
<td>8 (2911)</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>No N gene</td>
<td>5 (615)</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>ORF1ab gene</td>
<td>10 (3188)</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>No ORF1ab gene</td>
<td>3 (338)</td>
<td>0.00</td>
</tr>
<tr>
<td>Country</td>
<td>China</td>
<td>24 (4798)</td>
<td>1.31</td>
</tr>
<tr>
<td></td>
<td>Other countries</td>
<td>10 (7259)</td>
<td>0.36</td>
</tr>
<tr>
<td>Type of design</td>
<td>Accuracy</td>
<td>12 (1798)</td>
<td>1.52</td>
</tr>
<tr>
<td></td>
<td>Case series</td>
<td>22 (10259)</td>
<td>1.28</td>
</tr>
<tr>
<td>Risk of bias</td>
<td>High risk</td>
<td>8 (8947)</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>Unclear risk</td>
<td>20 (2549)</td>
<td>1.31</td>
</tr>
<tr>
<td></td>
<td>Low risk</td>
<td>6 (561)</td>
<td>0.60</td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pone.0242958.t002
differences related to the duration of symptoms at the time of first RT-PCR test based on information derived from nine studies provided means and medians of symptoms onset (Table 2). Comparison of false-negative rates of studies using different RT-PCR kits targeting (nucleocapsid N-gene and/or ORF1ab gene) makes no significant differences (Table 2). In addition, most studies (28 out of 34) reported a mix of specimen types collected for RT-PCR assessment; those reporting the use of nasopharyngeal swabs provided a range of false-negative from 0.018 to 0.33, while those reporting the additional use of oropharyngeal swabs reported a range of false-negative from 0.02 to 0.33. The analysis by country (China versus other countries) showed a potential effect on the summary estimations; the pooled estimate in non-Chinese countries was 0.06 (CI 95% = 0.04 to 0.09; 90% prediction interval 0.02 to 0.17; tau-squared = 0.36). Using meta-regression, we found a positive association of country with the false-negative rate (Table 2).

Additional post-hoc analysis by type of study did not provide a reduction of the observed heterogeneity (accuracy studies = 0.16, 95% CI 0.08 to 0.28, tau-squared = 1.52; case-series = 0.12, 95% CI 0.08 to 0.18, tau-square = 1.28). The false-negative proportion seemed higher in studies assessing test accuracy than in case series, although confidence intervals overlapped. An analysis by the global risk of bias (based on the QUADAS-II domains) showed a difference between high risk versus low-risk studies (high-risk studies = 0.08, 95% CI 0.04 to 0.14, tau-square = 0.79; low-risk studies = 0.33, 95% CI 0.20 to 0.49, Tau-square = 0.60), although the heterogeneity remains similar to those reported for the total group (Table 2).

Summary of findings under the GRADE approach

Since we could not warrant that the average estimates from the meta-analysis were valid and representative estimates of the true value of the false-negative proportion in the current practice, we used the estimated predictive interval in the analysis of the certainty of the evidence, using the GRADE approach. We illustrated the consequences of the range of false-negative rates in a population of 100 tested, according to three different prevalence estimates seen in current clinical practice by participant stakeholders and similar to those estimated by the included studies (10%, 30%, and 50%) (Fig 3). Using a prevalence of 50%, we found that 1 to 27 cases would be misdiagnosed and would not receive adequate clinical management; in addition, they could require repeated testing at some point in their hospitalisation or require another testing for competing diagnoses. The quality of the evidence was judged to be very low due to issues related to the risk of bias, indirectness, and inconsistency (Fig 3). This numerical approach should be interpreted with caution due to the multiple limitations of the evidence described above (Fig 3).

Discussion

Our systematic review identified 34 studies and 12,507 participants providing information about the proportion of false-negative (FN) cases in the detection of SARS-CoV-2 by RT-PCR assays at first use. Individual studies estimates of false-negative rate ranged from 0.018 to 0.58.

![Fig 3. Summary of findings (GRADE assessment). Notes= 1) Evidence downgraded one level due to risk of bias issues: multiple unclear risk related to patient selection and index test, several studies at high risk of bias in flow and timing Domain; 2) Evidence downgraded one level due to indirectness: unclear or high concerns about applicability of selected populations enrolled in studies; 3) Evidence downgraded one level due to inconsistency: tau-square =1.39.](https://doi.org/10.1371/journal.pone.0242958.g003)
Included studies were affected by several sources of potential bias, especially related to testing protocol to rule in/rule out the presence of SARS-CoV-2 infection, the analysis of a selected sample of COVID-19 patients, as well as the unclear report of key index test characteristics.

The meta-analysis of the FN rates showed a considerable between-study variability in estimates, which was not explained by any of the protocol defined characteristics. This variability is a limitation for the interpretation of the summary estimate of the proportion of the FN test results. Kucirka et al. also detected similar uncertainties in their bayesian modelling of false-negative rates of RT-PCR by time since exposure, based on information from seven studies and 1330 respiratory samples [69]. As an alternative, we illustrated the impact of heterogeneity by showing the number of false-negative cases expected in a cohort of 100 patients tested under three different prevalence of the disease scenarios. We based our calculations on the limits of the false-negative prediction interval. Using a prevalence of 50%, we found that in 100 persons tested, up to 27 cases would be misdiagnosed, putting them at risk of not receiving adequate clinical management or delaying isolation. We emphasised that these numerical approaches should be interpreted with caution due to very-low quality of evidence.

Our systematic review faced other challenges in its development. First, our study was initially planned as a rapid review aiming to provide a quick response to our local clinicians at the beginning of the COVID-19 pandemic. Due to the permanent involvement of clinicians managing COVID-19 patients at this point, we were able to define a review question that responds to a clinical inquiry relevant to current clinical practice [70–72]. However, due to the increasing number of publications potentially eligible to answer the review question, our approach evolved into a living systematic review with regular updates of the evidence. This manuscript reflects the third update of our literature searches with information current up to July 2020. To promote transparency in the development of this review, we have uploaded our previous results in the Open Science Framework repository for public consultation (https://tinyurl.com/vvbqyy).

A second challenge is related to the type of studies providing information about the false-negative rate associated with RT-PCR at initial testing. We expected to find studies specifically aimed to estimate the number of initial negative results of RT-PCR assays, with further confirmation of SARS-Cov-2 infection with an additional RT-PCR within the following days to the first result. On the contrary, we found that the reporting of false-negative rate was not the primary aim of any of the included studies. In some cases, these Figs were reported as descriptive statistics of the collected sample. Although we carried out a comprehensive and sensitive search strategy including major databases and repositories of preprint publications, we cannot discard that some eligible studies have not been identified yet due to the limitation of the reporting in key study sections, such as the abstract and methods.

Finally, as we have remarked in the findings section of this review, we found a considerable heterogeneity in the data insufficiently explained by the statistical analysis performed. Suggested sources of heterogeneity such as the type of specimen collected, the time to onset of symptoms (as an approach to viral load), as well as the name of the RT-PCR kit used, were partially or not reported at all by the included studies. This variability on COVID-19 testing data and the challenge to provide a pooled estimation with a useful clinical meaning have been previously remarked as the main constraint in the development of systematic reviews on this field [73]. Despite our efforts in the analysis of data, we only were able to find some reduction of this variability comparing those studies performed in China versus those carried out in other countries (i.e. USA, Singapore, and the Netherlands). We believe that information provided by Chinese studies reflects early experiences with the diagnosis of COVID-19; their findings are probably affected by several unreported issues, such as the RT-PCR kits in use (likely the first kits developed for SARS-CoV-2 detection), the lack of standardised methods for COVID-19 testing and, in general, the limited knowledge about this new infection at the beginning of 2020.
Despite the heterogeneous data available to address the review question, our study involved a rigorous assessment of potential sources of bias and applicability concerns, a formal statistical analysis and a final evaluation of the certainty of the evidence using a well-known system (GRADE). Although not all studies included in this review were accuracy studies, we decided to apply the QUADAS-II tool regardless of the type of design. However, even though QUADAS-II was not developed to evaluate case series, we preferred to standardise the quality assessment to report on a common pool of issues. We added the assessment of all studies as an appendix, using an adapted checklist tool for case-series to provide complementary information to this assessment. Due to the multiple difficulties associated with the lack of reporting of included studies, and due to the high probability of new studies being published in the short-term, we provided some recommendations for future studies candidates to be included in an update of this review:

- Inclusion of a series of consecutive patients instead of selected groups, to avoid spectrum bias.
- Description of RT-PCR scheme in use, including target genes under assessment and positivity criteria.
- Description of preanalytical steps (conservation of samples, time until being sent to the laboratory, training of personal).
- Clear reporting of the time since the onset of symptoms, especially for those patients with clinical findings at admission
- Reporting of the number of additional RT-PCR assays performed
- Details about the application of the reference standard, including the time of administration after the index test (initial RT-PCR)
- If possible, database sharing could allow re-analyses by independent researchers, including individual-patient data (IPD)-meta-analysis and increasing thus the confidence on the new evidence
- Adding serological samples to a cohort of individuals with compatible symptoms and negative PCR to warrant an independent verification of infection.

**Conclusions**

Our findings reinforce the need for repeated testing in patients with suspicion of being infected, due to either clinical or epidemiological reasons, given that up to 54% of COVID-19 patients may have an initial negative RT-PCR result (certainty of evidence: very low). The collected evidence has several limitations in terms of risk of bias and applicability. Incomplete reporting of several key factors hampered a comprehensive analysis of collected data. An update of this review is warranted when additional studies become available.

**Supporting information**

S1 File. PICO question/ eligibility criteria.
(PDF)

S2 File. Search strategies.
(PDF)

S3 File. Characteristic of excluded studies.
(PDF)
Author Contributions

**Conceptualization:** Ingrid Arevalo-Rodriguez, Diana Buitrago-Garcia, Daniel Simancas-Racines, Rosa Del Campo, Agustin Ciapponi, Anne W. Rutjes, Nicola Low, Jose A. Perez-Molina, Javier Zamora.

**Data curation:** Ingrid Arevalo-Rodriguez, Daniel Simancas-Racines, Rosa Del Campo, Agustin Ciapponi, Laura Martinez-García, Jose A. Perez-Molina, Javier Zamora.

**Formal analysis:** Ingrid Arevalo-Rodriguez, Agustin Ciapponi, Omar Sued, Laura Martinez-García, Javier Zamora.


**Project administration:** Ingrid Arevalo-Rodriguez, Agustin Ciapponi, Javier Zamora.

**Supervision:** Anne W. Rutjes, Nicola Low, Patrick M. Bossuyt, Jose A. Perez-Molina, Javier Zamora.


**Visualization:** Ingrid Arevalo-Rodriguez.

**Writing – original draft:** Ingrid Arevalo-Rodriguez, Diana Buitrago-Garcia, Daniel Simancas-Racines, Paula Zambrano-Achig, Rosa Del Campo, Agustin Ciapponi, Omar Sued, Laura Martinez-García, Anne W. Rutjes, Nicola Low, Jose A. Perez-Molina, Javier Zamora.


References


