Scoping review of COVID-19-related systematic reviews and meta-analyses: can we really have confidence in their results?

Rachel Wurth 1, Michelle Hajdenberg, Francisco J Barrera, Skand Shekhar, Caroline E Copacino, Pablo J Moreno-Peña, Omar A.M. Gharib, Forbes Porter, Swapnil Hiremath, Janet E Hall, Ernesto L Schiffrin, Graeme Eisenhofer, Stefan R Bornstein, Juan P. Brito, José Gerardo González-González, Constantine A Stratakis, René Rodríguez-Gutiérrez, Fady Hannah-Shmouni

ABSTRACT

Aim The aim of this study was to systematically appraise the quality of a sample of COVID-19-related systematic reviews (SRs) and discuss internal validity threats affecting the COVID-19 body of evidence.

Design We conducted a scoping review of the literature. SRs with or without meta-analysis (MA) that evaluated clinical data, outcomes or treatments for patients with COVID-19 were included.

Main outcome measures We extracted quality characteristics guided by A Measurement Tool to Assess Systematic Reviews-2 to calculate a qualitative score. Complementary evaluation of the most prominent published limitations affecting the COVID-19 body of evidence was performed.

Results A total of 63 SRs were included. The majority were judged as a critically low methodological quality. Most of the studies were not guided by a pre-established protocol (39, 62%). More than half (39, 62%) failed to address risk of bias when interpreting their results. A comprehensive literature search strategy was reported in most SRs (54, 86%). Appropriate use of statistical methods was evident in nearly all SRs with MAs (39, 95%). Only 16 (33%) studies recognised heterogeneity in the definition of severe COVID-19 as a limitation of the study, and 15 (24%) recognised repeated patient populations as a limitation.

Conclusion The methodological and reporting quality of current COVID-19 SR is far from optimal. In addition, most of the current SRs fail to address relevant threats to their internal validity, including repeated patients and heterogeneity in the definition of severe COVID-19. Adherence to proper study design and peer-review practices must remain to mitigate current limitations.

INTRODUCTION

Since the first report of SARS-CoV-2 in December 2019, the betacoronavirus responsible for COVID-19, there has been an exponential increase in the published literature on the topic; more than 2400 articles on COVID-19 were published in a single day alone, 5 June 2020, and since December, 56,534 full-text articles related to COVID-19 have been documented on the WHO Global literature on coronavirus disease database.

The deluge of manuscripts represents the largest explosion of scientific evidence in history, and there has been increasing concern that high publication volumes, including expedited peer-review processes and increased use of preprint servers, may be compromising the scientific quality of current research. Concerns are not limited to quality; duplicate and incomplete reporting of patient data has been recognised as a significant threat to the accuracy of subsequent prevalence and effect estimates. In addition, inconsistent clinical definitions, particularly for classifying severe COVID-19, make the synthesis of information a problematic task.

The large volume and variable quality of published work on COVID-19 highlight an overwhelming need to organise and summarise findings so that the most current and accurate information can be easily accessed. Several groups, including our own, have conducted systematic reviews (SRs) with or without meta-analyses (MAs) to address this need. Using our experience and a scoping review of the literature, we will discuss the limitations of the current COVID-19-related SRs and provide suggestions for improving future research.

METHODS

Search strategy and selection criteria

A search strategy was executed in MEDLINE via PubMed from 1 December 2019 until 17 May 2020. The search strategy was limited to peer-reviewed SRs with or without MAs published in English that evaluated clinical data, outcomes or treatments for patients with COVID-19. The search terms for PubMed were “coronavirus disease 2019,” “COVID-19” and “SARS-CoV-2.” This scoping review was not registered with PROSPERO.

Study selection

A title-and-abstract and a full-text screening phase was performed by experienced reviewers in an independent and duplicated manner. Each phase was prepiloted to ensure basic understanding of the selection criteria. Substantial agreement had to be achieved to perform each phase (kappa >0.70).

Data extraction and quality assessment

For each eligible study, two reviewers independently extracted the primary outcome, country(s) of the
primary studies, journal, associated impact factor for 2019 (per Journal Citation Reports, https://jcr.clarivate.com/) and methodological quality indicators.

Methodological appraisal using AMSTAR-2
The included studies’ methodological quality was appraised independently and in duplicate by experienced reviewers using the critical domains of A Measurement Tool to Assess Systematic Reviews-2 (AMSTAR-2).

A qualitative score of critically low, low, moderate or high quality was assigned to directly reflect the number of critical flaws present across each of the domains. A quantitative score was calculated by giving 1 point for ‘yes’ and 0.5 points for ‘partial yes’ and 0 points for ‘no’ for a total of 7 points for SRs with MAs and 5 points for SRs without MAs.12 13

Supplementary methodological appraisal
Complementary evaluation of internal validity threats to the COVID-19 body of evidence was performed based on several concerns with current COVID-19 reports. This evaluation aimed to ascertain the prevalence of SRs which included primary studies that repeated patient populations, provided clinical definitions for comorbidities (e.g., hypertension was defined using specific blood pressure values) and were preprint. In addition, the articles were assessed for the presence of methods to manage the absence of a universal definition for severe COVID-19.6–8 10 Discrepancies between reviewers in the screening and data extraction phases were resolved by consensus. If consensus could not be reached, a third senior investigator was consulted (FHS, RRG or JPB).

Statistical analysis
Categorical data are summarised in frequencies and percentages, and numerical data in means and SD. Student’s t-test and Pearson’s χ² test were performed to seek an association between the methodological quality as a quantitative or qualitative score, respectively, and the inclusion of preprint primary studies, or single/multinational primary studies, or a clearly defined primary outcome. Statistical analysis was performed using SPSS V.25.0 for Mac (IBM).

RESULTS
Our search strategy yielded a total of 105 studies, of which 63 met the inclusion criteria (figure 1, online supplemental appendix 1). The majority of the SRs included primary studies from more than one country (34, 56%) and 23 (37%) included data from a single country, China. The mean±SD for impact factor was 4.36±3.37 (range, 1.42–17.37; table 1) for SRs with MAs and 4.30±3.19 (range, 0.86–13.95; table 1) for SRs without MAs.

Methodological appraisal using AMSTAR-2
The methodological quality was qualitatively judged as critically low in 27 (66%) and 16 (73%) of the included studies for SRs with MAs and SRs without MAs, respectively; only 6 (15%) and 2 (9%) were judged as high quality (table 2). The mean±SD AMSTAR-2 score for SRs with MAs was 4.49±1.47 (range, 1–7) and for SRs without MAs was 1.98±1.52 (range, 0–5) (table 1). For both SRs with and without MAs, the inclusion of multinational primary studies, preprint primary studies or a
clearly defined primary outcome did not appear to influence the qualitative score (table 1).

The complete performance of the SRs with and without MAs for the critical domains of AMSTAR-2 can be found in table 3. Across both groups of included studies (SRs with MAs and SRs without MAs), the most critical methodological flaws were lack of or inadequate pre-established study protocol (39, 62%) and discussion of risk of bias when interpreting the results (39, 62%), respectively. In addition, the majority of SRs without MAs suffered from deficient techniques for assessing the risk of bias for the included studies (15, 68%; table 3). The most prominent strength of SRs with MAs was the use of appropriate statistical methods to synthesise results (39, 95%). In addition, the use of a comprehensive literature search strategy was reported in most of the included studies (54, 86%).

Supplementary methodological appraisal

Of the 49 SRs that evaluated comorbidities, only 15 (30%) provided a clinical definition of included comorbidities. Severe COVID-19 was evaluated as an outcome in 48 SRs, although only 16 (33%) recognised heterogeneity in the definition of severe COVID-19 as a limitation of the study. Of these 16 studies, 4 stated that only one definition of severe COVID-19 was used. Almost all (15, 94%) included an MA; however, the vast majority (11, 73%) did not perform a sensitivity analysis for each definition of severe COVID-19 used in the primary studies. Only 15 (24%) of the included SRs recognised repeated patient populations as a limitation of their study, and the majority (11, 73%) of these SRs implemented a method to mitigate the risk of including repeated patients in their analysis. Finally, 15 (24%) SRs included preprint primary studies.

DISCUSSION

Main findings

In this scoping review, we aimed to appraise the methodological quality of COVID-19 SRs with and without MAs using a validated SR appraisal tool, AMSTAR-2, and complementary criteria evaluating limitations pertinent to the current pandemic literature. Overall, the quality of included SRs was judged as critically low. A small number of studies recognised limitations affecting COVID-19-related primary literature, namely, the inclusion of primary studies that repeat patient populations and heterogeneity in the definition of severe COVID-19. Ultimately, the SRs evaluated in this study suffered from several major limitations and the reported effect estimates, and conclusions should be interpreted cautiously.

Methodological appraisal using AMSTAR-2

The majority of COVID-19-related SRs evaluated in this study suffered from significant limitations, with two-thirds of the included SRs with MAs and 7 of 10 SRs without MAs judged as a critically low methodological quality. The average quantitative score was 4.49 of 7 and 1.98 of 5 for SRs with and without MAs, respectively. Although AMSTAR-2 does not recommend the use of a quantitative score to evaluate SRs, previous studies have used both qualitative and quantitative scores. We did not find any correlation between qualitative score and the inclusion of multinational primary studies, preprint primary studies or a clearly defined primary outcome.

One of the most prominent flaws of both SRs with and without MAs was lack of or inadequate pre-established study protocol. Interestingly, previous studies evaluating the methodological quality of SRs across various disciplines also reported insufficient pre-established study protocols as a predominant limitation. Therefore, this limitation appears to impact SRs holistically. Brito et al reported that SRs including randomised controlled trials (RCTs) received a higher AMSTAR score. Inclusion of RCT primary studies was extremely limited in our body of evidence, likely due to the recent emergence of COVID-19. Future appraisals of the methodological quality of COVID-19-related SRs should explore associations between primary study design type and AMSTAR score. Ultimately, unambiguous eligibility criteria for included primary studies and a structure for quantitative and qualitative synthesis are critical components to an SR. SRs conducted without a prespecified protocol may be subject to selection bias.

We also found that the majority of SRs with and without MAs failed to adequately discuss risk of bias when interpreting the results of the study. In the SR/MA recently published by the authors, the primary studies were of critical risk of bias, and therefore it is imperative to recognise this limitation to prevent any conclusions from being overstated. While retrospective studies are an initial source of information at the onset of the pandemic, failure to consider potential biases affecting retrospective studies, including, but not limited to, confounding, and collider bias creates methodological flaws. Strengths of the included studies were use of appropriate statistical methods for combining results, such as justifying the use of a random-effects or fixed-effects model and providing pre-established methods to investigate heterogeneity, as well as the use of a comprehensive literature search. Adherence to both these practices will minimise bias and help achieve more representative and reliable effect estimates.
<table>
<thead>
<tr>
<th>ID</th>
<th>Author</th>
<th>Journal</th>
<th>Journal impact factor</th>
<th>Qualitative score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Systematic reviews with meta-analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Sarma et al.</td>
<td>Journal of Medical Virology</td>
<td>2.021</td>
<td>Critically low</td>
</tr>
<tr>
<td>5</td>
<td>Di Mascio et al.</td>
<td>American Journal of Obstetrics &amp; Gynecology</td>
<td>–</td>
<td>High</td>
</tr>
<tr>
<td>7</td>
<td>Yang et al.</td>
<td>Journal of Infection</td>
<td>–</td>
<td>Critically low</td>
</tr>
<tr>
<td>15</td>
<td>Cao et al.</td>
<td>Journal of Medical Virology</td>
<td>2.021</td>
<td>Critically low</td>
</tr>
<tr>
<td>16</td>
<td>Wang et al.</td>
<td>Research</td>
<td>–</td>
<td>Critically low</td>
</tr>
<tr>
<td>17</td>
<td>Mantovani et al.</td>
<td>Liver International</td>
<td>5.175</td>
<td>Critically low</td>
</tr>
<tr>
<td>19</td>
<td>Kumar et al.</td>
<td>Diabetes &amp; Metabolic Syndrome: Clinical Research &amp; Reviews</td>
<td>–</td>
<td>Low</td>
</tr>
<tr>
<td>20</td>
<td>Parohan et al.</td>
<td>Hepatology Research</td>
<td>3.165</td>
<td>Low</td>
</tr>
<tr>
<td>21</td>
<td>Farsalinos et al.</td>
<td>Internal and Emergency Medicine</td>
<td>2.322</td>
<td>Critically low</td>
</tr>
<tr>
<td>22</td>
<td>Tong et al.</td>
<td>Otolaryngology– Head and Neck Surgery</td>
<td>2.341</td>
<td>Low</td>
</tr>
<tr>
<td>24</td>
<td>Huang et al.</td>
<td>Diabetes &amp; Metabolic Syndrome: Clinical Research &amp; Reviews</td>
<td>–</td>
<td>Critically low</td>
</tr>
<tr>
<td>27</td>
<td>Hu et al.</td>
<td>Journal of Clinical Virology</td>
<td>2.777</td>
<td>Critically low</td>
</tr>
<tr>
<td>28</td>
<td>Chang et al.</td>
<td>Journal of the Formosan Medical Association</td>
<td>3.008</td>
<td>Critically low</td>
</tr>
<tr>
<td>30</td>
<td>Henry et al.</td>
<td>Clinical Chemistry and Laboratory Medicine</td>
<td>3.595</td>
<td>Critically low</td>
</tr>
<tr>
<td>32</td>
<td>Fu et al.</td>
<td>Journal of Infection</td>
<td>4.842</td>
<td>Critically low</td>
</tr>
<tr>
<td>33</td>
<td>Cheung et al.</td>
<td>Gastroenterology</td>
<td>17.373</td>
<td>Critically low</td>
</tr>
<tr>
<td>34</td>
<td>Lippi et al.</td>
<td>Polish Archives of Internal Medicine</td>
<td>3.007</td>
<td>Critically low</td>
</tr>
<tr>
<td>35</td>
<td>Wang et al.</td>
<td>Aging</td>
<td>4.831</td>
<td>Low</td>
</tr>
<tr>
<td>36</td>
<td>Emami et al.</td>
<td>Archives of Academic Emergency Medicine</td>
<td>–</td>
<td>Critically low</td>
</tr>
<tr>
<td>37</td>
<td>Zhao et al.</td>
<td>International Journal of Infectious Diseases</td>
<td>3.202</td>
<td>High</td>
</tr>
<tr>
<td>38</td>
<td>Zhu et al.</td>
<td>Family Medicine and Community Health</td>
<td>–</td>
<td>Low</td>
</tr>
<tr>
<td>40</td>
<td>Li et al.</td>
<td>Progress in Cardiovascular Diseases</td>
<td>6.763</td>
<td>Critically low</td>
</tr>
<tr>
<td>41</td>
<td>Zhao et al.</td>
<td>Journal of Medical Virology</td>
<td>2.021</td>
<td>Moderate</td>
</tr>
<tr>
<td>42</td>
<td>Zhu et al.</td>
<td>Journal of Medical Virology</td>
<td>2.021</td>
<td>Low</td>
</tr>
<tr>
<td>43</td>
<td>Borges do Nascimento et al.</td>
<td>Journal of Clinical Medicine</td>
<td>3.303</td>
<td>High</td>
</tr>
<tr>
<td>44</td>
<td>Wang et al.</td>
<td>Diabetes Research and Clinical Practice</td>
<td>4.234</td>
<td>Critically low</td>
</tr>
<tr>
<td>45</td>
<td>Aggarwal et al.</td>
<td>Current Problems in Cardiology</td>
<td>2.966</td>
<td>High</td>
</tr>
<tr>
<td>46</td>
<td>Pranata et al.</td>
<td>Journal of the Renin-Angiotensin-Aldosterone System</td>
<td>1.417</td>
<td>Critically low</td>
</tr>
<tr>
<td>47</td>
<td>Pranata et al.</td>
<td>Journal of Stroke &amp; Cerebrovascular Diseases</td>
<td>1.787</td>
<td>Critically low</td>
</tr>
<tr>
<td>48</td>
<td>Rodriguez-Morales et al.</td>
<td>Travel Medicine and Infectious Disease</td>
<td>4.589</td>
<td>Critically low</td>
</tr>
<tr>
<td>49</td>
<td>Yang et al.</td>
<td>International Journal of Infectious Diseases</td>
<td>3.202</td>
<td>Critically low</td>
</tr>
<tr>
<td>50</td>
<td>Li et al.</td>
<td>Clinical Research in Cardiology</td>
<td>5.268</td>
<td>Critically low</td>
</tr>
<tr>
<td>63</td>
<td>Alqhtani et al.</td>
<td>PLoS One</td>
<td>2.74</td>
<td>Critically low</td>
</tr>
<tr>
<td>65</td>
<td>Zhang et al.</td>
<td>Clinical Infectious Diseases</td>
<td>8.313</td>
<td>Moderate</td>
</tr>
<tr>
<td>66</td>
<td>Singh et al.</td>
<td>Diabetes &amp; Metabolic Syndrome: Clinical Research &amp; Reviews</td>
<td>–</td>
<td>Critically low</td>
</tr>
<tr>
<td>68</td>
<td>Zhang et al.</td>
<td>Pharmacological Research</td>
<td>5.893</td>
<td>Critically low</td>
</tr>
<tr>
<td>71</td>
<td>Wang et al.</td>
<td>Clinics and Research in Hepatology and Gastroenterology</td>
<td>2.718</td>
<td>Critically low</td>
</tr>
<tr>
<td>73</td>
<td>Gao et al.</td>
<td>Journal of Infection</td>
<td>4.842</td>
<td>Critically low</td>
</tr>
<tr>
<td></td>
<td>Systematic reviews without meta-analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Yang et al.</td>
<td>Journal of Maternal-Fetal &amp; Neonatal Medicine</td>
<td>1.737</td>
<td>Low</td>
</tr>
<tr>
<td>4</td>
<td>Zaigham et al.</td>
<td>Acta Obstetricia et Gynecologica Scandinavica</td>
<td>2.77</td>
<td>Critically low</td>
</tr>
<tr>
<td>6</td>
<td>Yousefiard et al.</td>
<td>Archives of Academic Emergency Medicine</td>
<td>–</td>
<td>Critically low</td>
</tr>
<tr>
<td>8</td>
<td>Ford et al.</td>
<td>Journal of the International AIDS Society</td>
<td>5.553</td>
<td>Critically low</td>
</tr>
<tr>
<td>9</td>
<td>Ludvigsson et al.</td>
<td>Acta Paediatrica</td>
<td>2.111</td>
<td>Critically low</td>
</tr>
<tr>
<td>11</td>
<td>Balla et al.</td>
<td>Journal of Clinical Medicine Research</td>
<td>–</td>
<td>Critically low</td>
</tr>
<tr>
<td>12</td>
<td>Moujaess et al.</td>
<td>Critical Reviews in Oncology/ Hematology</td>
<td>5.833</td>
<td>Critically low</td>
</tr>
<tr>
<td>13</td>
<td>AminJafari et al.</td>
<td>International Immunopharmacology</td>
<td>3.943</td>
<td>Critically low</td>
</tr>
<tr>
<td>14</td>
<td>Singh et al.</td>
<td>Diabetes &amp; Metabolic Syndrome: Clinical Research &amp; Reviews</td>
<td>–</td>
<td>Critically low</td>
</tr>
<tr>
<td>23</td>
<td>Minotti et al.</td>
<td>Journal of Infection</td>
<td>4.842</td>
<td>Critically low</td>
</tr>
<tr>
<td>26</td>
<td>Castagnoli et al.</td>
<td>JAMA Pediatrics</td>
<td>13.946</td>
<td>High</td>
</tr>
<tr>
<td>31</td>
<td>Lovato et al.</td>
<td>Ear, Nose &amp; Throat Journal</td>
<td>0.859</td>
<td>Low</td>
</tr>
<tr>
<td>51</td>
<td>Cortegiani et al.</td>
<td>Journal of Critical Care</td>
<td>2.685</td>
<td>Critically low</td>
</tr>
<tr>
<td>52</td>
<td>Vardavas et al.</td>
<td>Tobacco Induced Diseases</td>
<td>1.434</td>
<td>Critically low</td>
</tr>
<tr>
<td>55</td>
<td>Rajendran et al.</td>
<td>Journal of Medical Virology</td>
<td>2.021</td>
<td>Critically low</td>
</tr>
</tbody>
</table>

Continued
Supplementary methodological appraisal

Of the 63 SRs evaluated in our study, less than one-quarter either identified repeat patient populations as a limitation of their study or considered repeat patient populations to be a factor when selecting studies for MA. To avoid examination of repeat publications, some MAs excluded studies that appeared to have overlap.20 38–41 In some cases, authors of the primary publications may have been contacted to explain the overlaps. In one study,40 the authors assessed information from the facilities to which patients were admitted, as well as the ‘epidemiological week’ to avoid any overlap. However, the vast majority of publications examined did not recognise repeat patient populations as a limitation in performing an MA, whereas others recognised duplicate patient populations as a limitation of the study but did not specify if or how those repeat patient populations were addressed.

In our experience, to prevent analysis of repeated patient populations, we evaluated all included studies for overlap in both hospital and time frame of enrolment, selecting the study with the largest sample size when overlap was suspected.7 We encourage authors to implement similar methods to prevent the introduction of bias and inflation of results. In the case of interventional studies, reporting populations more than once increases the chance that the CI around the pooled effect size will be lower, altering interpretations of significance values.42

One of the larger concerns in performing SRs of COVID-19 is the lack of a universal definition of severe COVID-19.9 Of the 63 SRs examined, 48 evaluated severe COVID-19 as an outcome. However, half of the SRs that recognised heterogeneity in these definitions did not address the issue in their analysis. Of the SRs that addressed heterogeneity, many outlined severe disease according to specific organisations (such as the Chinese National Health Committee, WHO guidelines, American Thoracic Society Guidelines), whereas others constructed their own definition of severe COVID-19.20 These self-constructed definitions included presentation of acute respiratory distress syndrome, use of ventilation or life support, or admission to the intensive care unit (ICU). Inconsistencies in the definition of severe COVID-19 used by primary studies creates a source of bias known as information bias, where exposure and/or outcome are incorrectly determined.43 Future SRs should be cognisant of this type of bias and consider establishing selection criteria for primary studies and conducting sensitivity analyses to determine whether effect estimates vary.44 Ultimately, until a universal definition of severe COVID-19 is established, the clinical significance of severe COVID-19 as an outcome will remain unclear.

Implications for future research

SRs related to COVID-19 suffer from significant limitations as reflected by the poor methodological quality of the majority of SRs included in our study. In table 4, we provide a brief overview of additional, highly contested limitations affecting the COVID-19 body of evidence and possible solutions to mitigate their deleterious consequences. Future such analyses should establish methods to eliminate duplicate patient populations, including evaluation of overlapping hospitals and study duration, to prevent artificial inflation of outcomes.45 In addition, heterogeneity in the definition of severe COVID-19 prevents establishing reliable associations between risk factors and this outcome. Authors should consider using surrogate quantifiable

### Table 2: Performance of systematic reviews with and without meta-analyses for the critical domains of AMSTAR-2

<table>
<thead>
<tr>
<th>Critical domain question</th>
<th>n (% of studies judged as 'no')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systematic review with meta-analysis</td>
<td></td>
</tr>
<tr>
<td>Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?</td>
<td>22 (54)</td>
</tr>
<tr>
<td>Did the review authors use a comprehensive literature search strategy?</td>
<td>4 (10)</td>
</tr>
<tr>
<td>Did the review authors provide a list of excluded studies and justify the exclusion?</td>
<td>3 (7)</td>
</tr>
<tr>
<td>Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?</td>
<td>21 (51)</td>
</tr>
<tr>
<td>If meta-analysis was performed, did the review authors use appropriate methods for statistical combination of results?</td>
<td>2 (5)</td>
</tr>
<tr>
<td>Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review?</td>
<td>24 (59)</td>
</tr>
<tr>
<td>If they performed quantitative synthesis, did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?</td>
<td>12 (29)</td>
</tr>
<tr>
<td>Systematic review without meta-analysis</td>
<td></td>
</tr>
<tr>
<td>Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?</td>
<td>17 (77)</td>
</tr>
<tr>
<td>Did the review authors use a comprehensive literature search strategy?</td>
<td>5 (23)</td>
</tr>
<tr>
<td>Did the review authors provide a list of excluded studies and justify the exclusion?</td>
<td>9 (41)</td>
</tr>
<tr>
<td>Did the review authors use a satisfactory technique for assessing the RoB in individual studies that were included in the review?</td>
<td>15 (68)</td>
</tr>
<tr>
<td>If meta-analysis was performed, did the review authors use appropriate methods for statistical combination of results?</td>
<td>NA</td>
</tr>
<tr>
<td>Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review?</td>
<td>15 (68)</td>
</tr>
<tr>
<td>If they performed quantitative synthesis, did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?</td>
<td>NA</td>
</tr>
</tbody>
</table>
definitions of severity, such as ICU admission, to mitigate this concern.86 Ultimately, addressing these limitations will help reduce bias and establish a more accurate estimation of risks associated with COVID-19 outcomes of interest.

As new studies continue to be published, a living SR model (LSR) may serve as a valuable mechanism for representing the dynamic COVID-19 literature. LSRs are continuously updated, with new searches conducted at pre-established time frames to synthesise the most up-to-date information. LSRs are justified for research questions deemed important to clinical decision-making and in the setting of rapidly evolving or emerging health issues or disease.87 Therefore, authors should consider implementing this model to establish the most current risks and clinical guidance.

Ultimately, the COVID-19 pandemic has resulted in rapid collaborations between academia, government and industry, in some cases at a multinational level, to produce an astronomical amount of data on virtually every aspect of SARS-CoV-2. Although rapid dissemination of findings essential to human health is invaluable, long-standing practices of proper study design and peer review cannot be compromised if we are to establish optimal public health policies.

**Strengths and Limitations**

Limitations of our study include only searching in MEDLINE. It is likely that additional SRs that met our inclusion criteria were missed, and therefore our conclusions may not be generalisable to COVID-19-related SRs. However, the main intention of this scoping review was to obtain a representative sample of the total SRs available in the literature and not to provide a comprehensive overview and appraisal of the totality. Second, our appraisal was limited to the critical domains of AMSTAR-2, and the additional methodological qualities assessed through the non-critical domains of AMSTAR-2 were not evaluated in this study. Our study is strengthened by inclusion of a secondary evaluation of limitations pertinent to the COVID-19 body of evidence, inclusion of repeated patient populations and heterogeneity in the definition of severe COVID-19 and by a systematic approach to the screening and data extraction of the studies included.

**CONCLUSION**

Current SRs suffer from important methodological limitations according to our systematic evaluation using the AMSTAR-2 critical domains and additional concerns pertinent to the COVID-19 current literature. The methodological flaws place these articles at high risk of bias that, if existent, could influence their results and lead to misleading conclusions. Therefore, the findings of the majority of the studies should be interpreted with caution. We encourage future SRs to take into consideration these particularities in the COVID-19 literature to obtain more reliable results and lead to a better understanding of the current pandemic.

**Table 4** Additional limitations affecting the COVID-19 body of evidence and suggestions to mitigate their deleterious consequences

<table>
<thead>
<tr>
<th>Limitation</th>
<th>Definition and example</th>
<th>Recommendations to address limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confounding</td>
<td>Extraneous factors influence the effect of interest and can lead to misrepresentation of a causal relationship.95 96</td>
<td>DAGs can be used to assess confounders in exposure–outcome relationships.86 88 94</td>
</tr>
<tr>
<td></td>
<td>Age, diabetes mellitus and obesity are likely confounders for mortality risk in patients with diabetes mellitus and hypertension.97 These variables have not always been reported in the primary literature, limiting the rigour of sensitivity analyses in MAs.</td>
<td>Comprehensive reporting of patient characteristics.7</td>
</tr>
<tr>
<td>Collider bias</td>
<td>Two variables, exposure and outcome, influence a third variable, the collider, resulting in a false association between exposure and outcome.99</td>
<td>Weighted regression analysis to account for over-representation or under-representation of certain individuals.46</td>
</tr>
<tr>
<td></td>
<td>Testing for COVID-19, as currently conducted, may over-represent patients who are symptomatic, hospitalised and have better access to healthcare. Artificial associations may be established when using non-generalisable sample populations.14</td>
<td>DAGs can be used to assess confounders in exposure–outcome relationships.86 88 94</td>
</tr>
</tbody>
</table>
| Publishing demands         | The peer-review system is designed to mitigate the possibility that critically flawed articles reach circulation.  
Example: From 1 January to 1 June 2020, JAMA has received more than 11 000 articles, compared with 4000 in the same time frame the previous year.4  
Lancet Global Health has seen a 185% increase in submissions for June 2020 compared with June 2019.9 | No compromise should be made to the rigour of the peer-review process.4 |
| raise concern regarding     |                                                                                         |                                       |
| scientific quality          |                                                                                         |                                       |
| Preprint servers            | Repositories for studies that have not yet undergone the formal peer-review process.     | AI-powered literature reviews summarise key findings in articles to aid in identifying high-quality studies.11 11 |
|                             | Examples: Since 14 July 2020, 25 publications have been retracted, 10 of which have been preprint.86 | Databases, such as the Novel Coronavirus Research Compendium, feature high-quality articles each with their own appraisal.12 |

AI, artificial intelligence; DAGs, directed acyclic graphs; MAs, meta-analyses.

Main messages

- The quality of COVID-19-related systematic reviews (SRs) included in our study was judged as critically low, and only a small number of studies recognised limitations affecting COVID-19 related primary literature.
- The most prominent methodological flaws of the included SRs were lack of or inadequate pre-established study protocol and discussion of risk of bias.
- The majority of included SRs used a comprehensive literature search strategy and appropriate statistical methods to synthesise results (when a meta-analysis was performed).
REFERENCES
11 Brainard J. Scientists are drowning in COVID-19 papers. can new tools keep them afloat? 2020.
Dijkers M. Duplicate publications and systematic reviews: problems and proposals 2018;6:12.
Joanne Brooker AS, McDonald S, Elliot J. Guidance for the production and publication of Cochrane living systematic reviews: cochrane reviews in living mode, 2019.
Grimes DA, Grimes DD. Guideline for the production and publication of Cochrane living systematic reviews: cochrane reviews in living mode. 2018;6:12.
Dijkers M. Duplicate publications and systematic reviews: problems and proposals 2018;6:12.
Joanne Brooker AS, McDonald S, Elliot J. Guidance for the production and publication of Cochrane living systematic reviews: cochrane reviews in living mode, 2019.
Appendix 1. Complete citation list of included articles.


