

Quality of early evidence on the pathogenesis, diagnosis, prognosis and treatment of COVID-19

Sarah Yang ¹, Allen Li ², Ali Eshaghpour ³,
Sofia Ivanisevic,¹ Adrian Salopek,¹ John Eikelboom,⁴
M Crowther⁴

10.1136/bmjebm-2020-111499

► Additional material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjebm-2020-111499>).

¹McMaster University Faculty of Health Sciences, Hamilton, Ontario, Canada

²University of Ottawa Faculty of Medicine, Ottawa, Ontario, Canada

³McMaster University Michael G DeGroot School of Medicine, Hamilton, Ontario, Canada

⁴Department of Medicine, McMaster University, Hamilton, Ontario, Canada

Correspondence to:

Ali Eshaghpour, McMaster University Michael G DeGroot School of Medicine, Hamilton, ON L8S 4S5, Canada; ali.eshaghpour@medportal.ca



Check for updates

© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Yang S, Li A, Eshaghpour A, *et al*. *BMJ Evidence-Based Medicine* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjebm-2020-111499

Abstract

Since the initial description of the SARS-CoV-2 outbreak and its declaration as a worldwide pandemic, the number of publications on the novel virus has increased rapidly. We studied the trends and quality of evidence in early SARS-CoV-2 publications. A comprehensive search of MEDLINE and EMBASE was performed for papers published between 1 January 2020 and 21 April 2020. Two reviewers independently screened titles and abstracts and subsequently full texts for eligibility in this systematic review. The search yielded 2504 citations published between January and February 2020 or an unspecified date, 109 of which remained for extraction after screening. Data extracted included study design, year of publication, country of basis, journal of publication, impact factor of publishing journal, study sample size, number of citations and topic of investigation. Study design-specific critical appraisal tools were used to evaluate the scientific rigour of all included papers: the Joanna Briggs Institute checklist was used for case series, Scale for the Assessment of Narrative Review Articles scale for narrative reviews, Newcastle-Ottawa scale for cohort studies and AMSTAR 2 for systematic reviews. The overall quality of the literature was low-moderate. Of 541 papers that reported clinical characteristics, 295 were commentaries/expert opinions and 36 were case reports. There were no randomised clinical trials, 45 case series studies, 58 narrative reviews, 1 cohort study and 5 systematic reviews. We encourage clinicians to be attentive to these findings when utilising early SARS-CoV-2 evidence in their practices.

Introduction

SARS-CoV-2 was first identified in December 2019 in Wuhan, China and has since been declared a worldwide pandemic.^{1 2} Since the initial description of the outbreak, the number of publications, including commentaries, expert opinions, mathematical modelling studies, observational studies, as well as the results of several randomised clinical trials (RCTs)³⁻⁵ in patients with SARS-CoV-2 has increased rapidly. Online resource centres have been set up to consolidate SARS-CoV-2-related literature with fast turn-around times for article submission and publication.⁶ The body of

literature surrounding SARS-CoV-2 has thereby rapidly increased.⁷

High-quality evidence is desirable to guide clinical practice.⁸ In the hierarchy of quality of original science involving patients, RCTs generally provide the most reliable evidence, followed by cohort studies, case-control studies, case series and case reports.⁹ Systematic reviews, narrative reviews and commentaries/expert opinions are also sources of summative scientific information. Systematic reviews consolidate evidence through a structured method of collecting data, appraising evidence and synthesising results.¹⁰ Robust systematic reviews that include a meta-analysis of high-quality RCTs can provide more reliable evidence than the constituent RCTs themselves.¹¹ A meta-analysis can also assess for publication bias and provide an estimate of effect size towards the research question. Narrative reviews are less reliable because they do not employ a systematic approach to summarising the evidence,¹² and are therefore prone to bias.¹³ Commentaries/expert opinions are similar to case reports and case series as they are frequently anecdotal.¹¹

In this paper, we critically examine the quality of evidence from papers on SARS-CoV-2 infection published in January and February 2020.

Methods

This systematic review was reported following standards outlined in the Preferred Reporting Items of Systematic Reviews and Meta-Analyses (PRISMA) statement.¹⁴ The research protocol was formulated prospectively and is available online here.

Search strategy

A comprehensive search of MEDLINE and EMBASE was conducted on 21 April 2020, with the aid of a medical librarian at McMaster University. We limited our search to papers published from 1 January 2020 to 21 April 2020 and searched for articles using the following keywords: 'COVID-19', 'SARS-CoV-2', '2019-nCoV', '2019 novel coronavirus', 'Wuhan coronavirus', 'new coronavirus' and 'coronavirus'. Results were uploaded to Endnote (Clarivate Analytics) and duplicates were removed.

Study selection

The articles from Endnote were uploaded onto Microsoft Excel to filter papers based on

publication date. Studies published between January and February 2020 or an unspecified publication date were included. Titles and abstracts were then uploaded onto Covidence (Veritas Health Innovation Ltd) for independent screening. Screening was performed in duplicate across two stages. First, two review authors (SI and AS) independently assessed titles and abstracts. Next, included articles underwent full-text analysis by the same two authors to assess eligibility for data extraction. Any discrepancies in eligibility during a given stage were resolved through joint discussions with a third reviewer (SY). Inclusion criterion was *any studies pertaining to clinical outcome or characteristic of SARS-CoV-2*. Exclusion criteria for abstract and title screening included: *studies whose full text could not be sourced, studies not pertaining to SARS-CoV-2, in vitro and in vivo lab studies, studies not published in January or February 2020, studies not published in the English language and virus transmission models*. In full-text screening, we added the exclusion criterion of *commentaries or case reports* to record the number of these types of publications.

Data extraction

Data were extracted from each study independently by two authors (SI and AS), and discrepancies were resolved through discussion until agreement was reached between the reviewers. Information extracted included journal of publication, impact factor of journal, country of journal basis, date of publication, number of times the paper has been cited, study design and topic of study. Journal impact factors were determined using the 2018 InCites Journal Citation Report, and the number of citations of each study was obtained using Google Scholar (Google LLC, Mountain View, Cali.).

Assessment of methodological quality

Methodological quality was independently assessed by two authors (SI and AS), and discrepancies in assessment were resolved through joint discussion until consensus was reached. Tools used for evaluation include AMSTAR-2 for systematic reviews,¹⁵ Joanna Briggs Institute (JBI) Checklist for case series,¹⁶ Scale for the Assessment of Narrative Review Articles (SANRA) for narrative reviews¹⁷ and the Newcastle–Ottawa scale for non-randomised studies.¹⁸ Selections of unclear (U) or N/A were assigned 0 points, and total scores for each study on their corresponding scale were tabulated. Given that AMSTAR-2 does not generate a numerical score, these studies were reported as low, medium or high quality per AMSTAR-2.

Methods of data analysis and synthesis

Descriptive data were recorded using Microsoft Excel (Microsoft, Redmond, Wash.) and summarised in tables and graphs. Quality assessment data were summarised in subgroups with values reported as median (IQR). The number of patients per study type as well as across all included studies were summarised using median (IQR). Statistical analyses for Spearman's rank correlation between quality of the article and journal impact factor were calculated using GraphPad Prism 7 (GraphPad Software, San Diego, Cali.). Journals whose impact factors could not be found were not included in these analyses but were included in the general descriptive results.

Results

Search results

After removing 625 duplicates, our initial search yielded a total of 2504 citations. 1963 studies were excluded during title and abstract screening, leaving 541 citations that pertained to a

Table 1 Type of studies included in analysis and number of each study

Type of study	Number of studies
Case series	45
Narrative reviews	58
Cohort study	1
Systematic reviews	5
Case reports	36

clinical outcome or characteristic of SARS-CoV-2 for full-text screening. Of the 541 studies, 432 studies were excluded. Out of these 432 excluded studies, 295 were commentaries, 36 were case reports and 101 were excluded based on reasons listed in the PRISMA flow chart. 109 citations were left to be extracted. The PRISMA flow chart of study selection can be found online in the online supplemental file.

General characteristics of studies

The countries where studies were performed spanned 15 countries. The number of studies included in each country can be found in online supplemental table 7. The USA was the country of origin of most of the included studies (35/109, 32.1%) followed by the UK (29/109, 26.6%) and China (16/109, 14.7%). 59/109 (41.8%) of the included studies were narrative reviews, 45/109 (31.9%) were case series, 1/109 (0.7%) was a cohort study and 5/109 (3.5%) were systematic reviews (table 1). We did not find any published randomised clinical trials. The most common study topic was an overview of SARS-CoV-2. The median number of patients in case series and cohort studies was 46 (IQR=66). The median number of patients included in all study types can be found in table 2.

Methodological quality of studies

The quality scales that we used to grade papers assessed certain methodological processes including quality of literature search strategies, presentation of data and use of consecutive participant inclusion. While this approach does not provide a comprehensive critical appraisal of a paper, it enables standard comparisons between papers. The methodological quality of each type of study is summarised in table 3. The specific scoring and description of each study can be found in online supplemental tables 1–4.

Case series

The JBI checklist, which consists of a 10-point scale, was used to score case series. No studies received a perfect score and one study received a 0/10 score.¹⁹ The median score was 7 (2). *Use of valid methods for identification of all participants' condition*

Table 2 Median number of patients enrolled in each study type

Type of study	Median number of patients (IQR)*
All studies (excluding systematic review and narrative review)	46 (66)
Case series	41 (68)
Narrative reviews	36 178 (36 136.5)
Cohort studies	57 (n/a)†
Systematic reviews	25 760 (24 706)

*Studies that did not state patient numbers were not included in calculations.

†Only one cohort study was found.

n/a, not applicable.

Table 3 Types of study designs included, and respective quality assessment scales used

Study design	Number of articles	Quality scale	Median (IQR)/total score
Case series	45	JBI Checklist for Case Series	7 (2)/10
Narrative review	58	SANRA	9 (2)/12
Cohort studies	1	Newcastle-Ottawa Scale for non-randomised studies	5 (0)/9
Systematic review	5	AMSTAR-2	—*

*A numerical score AMSTAR-2 score for systematic reviews could not be generated as the tool scored studies using low, moderate or high metric values.

JBI, Joanna Briggs Institute; SANRA, Scale for the Assessment of Narrative Review Articles.

(40/45, 88.9%) and *clear reporting of the demographics of participants* were the highest scoring criteria (43/45, 95.6%), while the least reported criteria were *consecutive inclusion of participants* and *clear reporting of the presenting site(s)/clinic(s) demographic information* (0/45, 0%).

Narrative reviews

A total of 58 narrative reviews were identified with the median score being 9 (2) on the 12-point SANRA scale. Two studies scored perfectly.^{20,21} The most reported criteria (where most studies scored 2/2 for that specific criterion) were *key statements supported by references* (56/58, 96.6%), *appropriate scientific reasoning* (56/58, 96.6%) and *appropriate presentation of data* (53/58, 91.4%). The least reported criterion (where most studies scored 0/2) was *description of the literature search* (2/58, 3.4%).

Cohort study

The cohort study received a score of five on the nine-point Newcastle-Ottawa Scale for non-randomised studies. The study scored highest in the *selection* category (3/4), and lowest in the *outcome* category (1/3).

Systematic reviews

The AMSTAR 2 checklist, which is scaled from *critically low to high overall*, was used to score the systematic reviews. Of the five total systematic reviews we included, three received a score of 'low'. These studies had one critical flaw and could not be justified to have provided an accurate and comprehensive summary of its included studies. The remaining two systematic reviews received a 'moderate score' as they had more than one non-critical weakness but provided an accurate and comprehensive summary of the available studies.

A two-tailed Spearman's rank correlation test revealed no significant correlation between study quality and the impact factor of the publishing journal for case series studies ($r=0.11$; CI: -0.21 to 0.41 ; $p=0.49$) or narrative reviews ($r=0.12$; 95% CI: -0.17 to 0.40 ; $p=0.40$).

Discussion

In this systematic review, we evaluated clinical studies of COVID-19 published between January and February 2020. Of 541 papers that reported clinical characteristics, 295 were commentaries/expert opinions and 36 were case reports. There were no RCTs, 45 case series studies, 58 narrative reviews, 1 cohort study and 5 systematic reviews. For studies that included more than one patient, the majority consisted of narrative reviews ($n=58$) and

case series ($n=45$). China was the originating country for over half of the literature (56%) with most of the studies being focused on the diagnosis of COVID-19 (51%). There was no significant correlation between study quality and the impact factor of its publishing journal.

Commentaries and expert opinions can offer practical knowledge for practitioners,^{22,23} especially when SARS-CoV-2 was first described and little was known about the diagnosis and management. Similarly, case reports can offer insights into the clinical presentation and possible approaches to management in the early stages of the pandemic, although they are severely limited in generalisability due to their $n=1$ sample size. Both case reports and case series lack a control group and are susceptible to selection and reporting biases, but can provide information that informs and develops hypotheses that can be tested using more rigorous study methodology.²⁴ The initial case series articles documenting the recovery of patients taking lopinavir-ritonavir therapy which was later investigated through an RCT serves as an example.²⁵ Well-designed cohort studies overcome many of the limitations of case series but can still introduce bias. A recent methodological evaluation of the cohort studies and RCTs examining chloroquine and hydroxychloroquine therapy for SARS-CoV-2 found concerning methodological issues.²⁶ While it is important to disseminate information rapidly, especially during a pandemic of a novel virus, it is essential for both authors and journals to ensure scientific rigour.⁷ The consequences of not adhering to sound scientific principles can be quite drastic, as seen with the recent retraction of a major chloroquine and hydroxychloroquine study in the *Lancet* that, in retrospect, inappropriately halted many RCTs worldwide. Furthermore, when primary studies are not undertaken with rigour, review articles and meta-analyses inadvertently become at risk of bias as they consolidate and summarise primary literature.

In the setting of a pandemic, each study design can hold value if methodologically sound, but it is important for clinicians to recognise their strengths and limitations. Though RCTs provide the most reliable evidence concerning efficacy and safety of interventions, they can be costly to conduct, require ethics and regulatory approvals and require adequate follow-up.²⁷ Furthermore, other study designs might be needed to bridge knowledge gaps where RCTs cannot, such as exploring disease pathogenesis and characteristics. Systematic reviews can quickly summarise the conclusions across these studies to help disseminate knowledge quickly and efficiently to clinicians.

A strength of our study is that it provides a practical view of the body of literature used to inform early clinical decisions and highlights a need for earlier high-quality controlled study designs such as RCTs and controlled-cohort studies. Furthermore, the use of SANRA and the JBI checklist allowed for analysis of narrative review and case series quality, which are seldom measured. Our study also highlights specific weaknesses in methodology used, allowing authors to incorporate this information in future publications. For example, most narrative reviews did not describe their search strategy, and most case series did not adequately report the presenting site/clinic demographic information.

To the best of our knowledge, there have been no studies published examining the quality of literature in the early phases of other diseases or pandemics. While other studies extracted similar data to our paper, they had a greater focus on literature characteristics without necessarily evaluating literature quality. Further, they included a broader range of studies including non-clinical papers.^{28,29} Our study also has

limitations. While we identified overall poor study designs and moderate quality, we cannot objectively determine whether the quality of evidence was inferior to the usual quality of evidence over a comparable timeframe. Our search was limited to articles published in the English language in January and February. In these very early stages, lower thresholds of quality may be acceptable as the body of literature develops. The metric of citation rates via google scholar may be inflated, but nevertheless provides a platform to compare the studies.³⁰

In the context of the low-moderate quality COVID-19 evidence base, it is important for clinicians to continue practicing evidence-based medicine. Clinicians should be competent in appraising the literature and applying relevant findings to the appropriate patient population. It is important to not inform practice solely on expert opinions and commentaries. Furthermore, while high impact journals are expected to publish higher quality studies, our study shows this was not the case in early COVID-19 research. This further highlights the need for clinicians to use appropriate appraisal skills for independent studies, and not apply findings based on the publishing journal.

Conclusion

Our study demonstrates that early COVID-19 literature was of, at best, moderate quality, and very limited in terms of study design, with only one cohort study and no RCTs. The quality of the published literature did not correlate with the impact factor of the publishing journal. While these studies are helpful in informing clinical decisions early in the pandemic, clinicians should be cautious in adopting evidence from the early literature into their practices. Further research is warranted to determine how the quality of evidence changes over time, and to see how the emerging evidence informs RCT design.

Contributors AL, AE and MC conceived and structured the project. SY, SI and AS collected data and analysed the data. SY, AL, AE, JE and MC interpreted the data. All authors contributed to the drafting and critical revision of the manuscript, and approved the final version.

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.

Competing interests MC reports grants and other from Bayer, personal fees from Pfizer, other from Alnylam, personal fees from CSL Behring, personal fees from Servier Canada, personal fees from Diagnostica Stago and personal fees from Asahi Kasei. JE reports consulting fees and/or honoraria from Astra-Zeneca, Bayer Boehringer-Ingelheim, Bristol-Myer-Squibb, Daiichi-Sankyo, Eli-Lilly, Glaxo-Smith-Kline, Pfizer, Janssen, Sanofi-Aventis, Servier and grants and/or in-kind support from Astra-Zeneca, Bayer, Boehringer-Ingelheim, Bristol-Myer-Squibb, Glaxo-Smith-Kline, Pfizer, Janssen, Sanofi-Aventis.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

This article is made freely available for use in accordance with BMJ's website terms and conditions for the duration of the covid-19 pandemic or until otherwise determined by BMJ. You may use, download and print the article for any lawful, non-

commercial purpose (including text and data mining) provided that all copyright notices and trade marks are retained.

ORCID iDs

Sarah Yang <http://orcid.org/0000-0003-3811-6378>

Allen Li <http://orcid.org/0000-0001-6615-7295>

Ali Eshaghpour <http://orcid.org/0000-0002-7827-5827>

References

- 1 Helmy YA, Fawzy M, Elasad A, *et al*. The COVID-19 pandemic: a comprehensive review of taxonomy, genetics, epidemiology, diagnosis, treatment, and control. *J Clin Med* 2020;9:1225.
- 2 WHO. WHO announces COVID-19 outbreak a pandemic, 2020. Available: <http://www.euro.who.int/en/health-topics/healthemergencies/coronavirus-covid-19/news/news/2020/3/whoannounces-covid-19-outbreak-a-pandemic> [Accessed 19 Apr 2020].
- 3 Gates B. Responding to Covid-19 - A Once-in-a-Century Pandemic? *N Engl J Med* 2020;382:1677-9.
- 4 Tuite AR, Fisman DN. Reporting, epidemic growth, and reproduction numbers for the 2019 novel coronavirus (2019-nCoV) epidemic. *Ann Intern Med* 2020;172:567.
- 5 Mullard A. Flooded by the torrent: the COVID-19 drug pipeline. *Lancet* 2020;395:1245-6.
- 6 Horbach S. Pandemic publishing: medical journals drastically speed up their publication process for Covid-19. *bioRxiv* 2020.
- 7 Bauchner H, Fontanarosa PB, Golub RM. Editorial evaluation and peer review during a pandemic: how journals maintain standards. *JAMA* 2020;324:453-4.
- 8 Sackett DL, Rosenberg WMC, Gray JAM, *et al*. Evidence based medicine: what it is and what it isn't. 1996. *Clin Orthop Relat Res* 2007;455:3-5.
- 9 Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 1989;95:2S-4.
- 10 Higgins JPT, Green S, Cochrane Collaboration. *Cochrane Handbook for systematic reviews of interventions*. Chichester, West Sussex ; Hoboken NJ: Wiley-Blackwell, 2008.
- 11 Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *J Clin Nurs* 2003;12:77-84.
- 12 Greenhalgh T, Thorne S, Malterud K. Time to challenge the spurious hierarchy of systematic over narrative reviews? *Eur J Clin Invest* 2018;48:e12931.
- 13 Pae C-U. Why systematic review rather than narrative review? *Psychiatry Investig* 2015;12:417-9.
- 14 Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 2009;62:1006-12.
- 15 Shea BJ, Grimshaw JM, Wells GA, *et al*. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.
- 16 Munn Z, Barker TH, Moola S, *et al*. Methodological quality of case series studies: an introduction to the JBI critical appraisal tool. *JBI Database System Rev Implement Rep* 2019.
- 17 Baethge C, Goldbeck-Wood S, Mertens S. SANRA—a scale for the quality assessment of narrative review articles. *Res Integr Peer Rev* 2019;4:5.
- 18 Wells G, Shea B, O'Connell D. *The Newcastle-Ottawa scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses*, 2013.
- 19 Hodcroft EB. Preliminary case report on the SARS-CoV-2 cluster in the UK, France, and Spain. *Swiss Med Wkly* 2020;150. doi:10.4414/smw.2020.20212. [Epub ahead of print: 24 Feb 2020].
- 20 Liao BWaYK X. Novel coronavirus infection during the 2019-2020 epidemic: preparing intensive care units—the experience in Sichuan Province, China..
- 21 Jin Y-H, Cai L, Cheng Z-S, *et al*. A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia (standard version). *Mil Med Res* 2020;7.
- 22 Hohmann E, Brand JC, Rossi MJ, *et al*. Expert opinion is necessary: Delphi panel methodology facilitates a scientific approach to consensus. *Arthroscop J Arthroscopic Related Surg* 2018;34:349-51.

- 23 Glasby J, Beresford P. Commentary and Issues : Who knows best? Evidence-based practice and the service user contribution. *Crit Soc Policy* 2006;26:268–84.
- 24 Chan K, Bhandari M. Three-minute critical appraisal of a case series article. *Indian J Orthop* 2011;45:103–4.
- 25 Cao B, Wang Y, Wen D, *et al.* A trial of Lopinavir–Ritonavir in adults hospitalized with severe Covid-19. *N Engl J Med Medicine* 2020;382:1787–99.
- 26 Alexander PE, Debono VB, Mammen MJ, *et al.* COVID-19 coronavirus research has overall low methodological quality thus far: case in point for chloroquine/hydroxychloroquine. *J Clin Epidemiol* 2020;123:120–6.
- 27 Kabisch M, Ruckes C, Seibert-Grafe M, *et al.* Randomized controlled trials: part 17 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 2011;108:663–8.
- 28 Haghani M, Bliemer MCJ. Covid-19 pandemic and the unprecedented mobilisation of scholarly efforts prompted by a health crisis: Scientometric comparisons across SARS, MERS and 2019-nCov literature. *bioRxiv* 2020.
- 29 Haghani M, Bliemer MCJ, Goerlandt F, *et al.* The scientific literature on coronaviruses, COVID-19 and its associated safety-related research dimensions: a scientometric analysis and scoping review. *Saf Sci* 2020;129:104806.
- 30 Harzing AWK, van der Wal R. Google Scholar as a new source for citation analysis. *Ethics Sci Environ Polit* 2008;8:61–73.