



Characteristics, quality and volume of the first 5 months of the COVID-19 evidence synthesis infodemic: a meta-research study

Rebecca Abbott ¹, Alison Bethel,¹ Morwenna Rogers,¹ Rebecca Whear,¹ Noreen Orr,¹ Liz Shaw,² Ken Stein,¹ Jo Thompson Coon ¹

10.1136/bmjebm-2021-111710

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjebm-2021-111710>).

¹NIHR ARC South West Peninsula, University of Exeter Medical School, University of Exeter, Exeter, UK

²University of Exeter Medical School, University of Exeter, Exeter, UK

Correspondence to:

Dr Rebecca Abbott, NIHR ARC South West Peninsula, University of Exeter Medical School, University of Exeter, Exeter EX2 1LS, UK; r.a. abbott@exeter.ac.uk



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Abbott R, Bethel A, Rogers M, *et al.* *BMJ Evidence-Based Medicine* 2022;**27**:169–177.

Abstract

Objective The academic and scientific community has reacted at pace to gather evidence to help and inform about COVID-19. Concerns have been raised about the quality of this evidence. The aim of this review was to map the nature, scope and quality of evidence syntheses on COVID-19 and to explore the relationship between review quality and the extent of researcher, policy and media interest.

Design and setting A meta-research: systematic review of reviews.

Information sources PubMed, Epistemonikos COVID-19 evidence, the Cochrane Library of Systematic Reviews, the Cochrane COVID-19 Study Register, EMBASE, CINAHL, Web of Science Core Collection and the WHO COVID-19 database, searched between 10 June 2020 and 15 June 2020. **Eligibility criteria** Any peer-reviewed article reported as a systematic review, rapid review, overview, meta-analysis or qualitative evidence synthesis in the title or abstract addressing a research question relating to COVID-19. Articles described as meta-analyses but not undertaken as part of a systematic or rapid review were excluded. **Study selection and data extraction** Abstract and full text screening were undertaken by two independent reviewers. Descriptive information on review type, purpose, population, size, citation and attention metrics were extracted along with whether the review met the definition of a systematic review according to six key methodological criteria. For those meeting all criteria, additional data on methods and publication metrics were extracted.

Risk of bias For articles meeting all six criteria required to meet the definition of a systematic review, AMSTAR-2 ((A MeaSurement Tool to Assess systematic Reviews, version 2.0) was used to assess the quality of the reported methods.

Results 2334 articles were screened, resulting in 280 reviews being included: 232 systematic reviews, 46 rapid reviews and 2 overviews. Less than half reported undertaking critical appraisal and a third had no reproducible search strategy. There was considerable overlap in topics, with discordant findings. Eighty-eight of the 280 reviews met all six systematic review criteria. Of these, just 3 were rated as of moderate or high quality on AMSTAR-2, with the majority having

Summary box

What is already known about this subject?

⇒ Poorly conducted systematic reviews can lead to inaccurate representations of the evidence, inaccurate estimates of treatment effectiveness, misleading conclusions and reduced applicability.

What are the new findings?

⇒ Most COVID-19 evidence syntheses described as systematic or rapid review were of low quality and missed out cornerstones of best practice. Less than a half reported critically appraising their included studies, and a third had no reproducible search strategy. Review conduct and publication were rapid. Interest, as measured by altmetrics and citations, was high, and not correlated with quality.

How might it impact clinical practice in the foreseeable future?

⇒ By being reported as ‘systematic reviews’, many readers may regard evidence syntheses as high-quality evidence, irrespective of the actual methods undertaken. The challenge especially in times such as this pandemic is to provide indications of trustworthiness in evidence that is available in ‘real time’. Researchers, peer reviewers and journal editors need to ensure that robust methods have been used for research denoted as systematic reviews.

critical flaws: only a third reported registering a protocol, and less than one in five searched named COVID-19 databases. Review conduct and publication were rapid, with 52 of the 88 systematic reviews reported as being conducted within 3 weeks, and a half published within 3 weeks of submission. Researcher and media interest, as measured by altmetrics and citations, was high, and was not correlated with quality.

Discussion This meta-research of early published COVID-19 evidence syntheses found low-quality reviews being published at pace, often with short publication turnarounds. Despite being of low quality and many lacking robust methods, the reviews received substantial attention across both academic and public platforms, and the attention was not related to the quality of review methods.

Interpretation Flaws in systematic review methods limit the validity of a review and the generalisability of its findings. Yet, by being reported as ‘systematic reviews’, many readers may well regard them as high-quality evidence, irrespective of the actual methods undertaken. The challenge especially in times such as this pandemic is to provide indications of trustworthiness in evidence that is available in ‘real time’.

PROSPERO registration number CRD42020188822.

Introduction

Since the emergence of the COVID-19 in December 2019 in Wuhan, China, there has been a proliferation of research related to its epidemiology, diagnosis, treatment, prevention and impact. As of 10 January 2021, there were over 77 000 records on PubMed alone that included COVID-19 somewhere in the title or abstract. On 5 May 2020, when we first drafted the protocol for this review, there were more than 60 published systematic reviews on COVID-19 when searching by title in PubMed alone, and as of 10 January 2021, this stands at 1820. The COVID-19 Evidence Reviews resource (<http://covid19reviews.org/index.cfm>) suggests that at this time, there are over 4000 systematic reviews, rapid reviews and evidence summaries on COVID-19 that have either been published, or in the process of being carried out.

Making sense of research by bringing together studies in systematic reviews, with or without meta-analysis, is a well-established method in medicine and health research.¹ Cochrane and other evidence-based health programmes have promoted the use of systematic review methods globally.²⁻⁴ However, poorly conducted reviews can lead to inaccurate representations of the evidence, inaccurate estimates of treatment effectiveness, misleading conclusions and reduced applicability,⁵ limiting their usefulness and ultimately contributing to research waste. There are concerns that in the panic to get answers to help manage the COVID-19 pandemic, systematic reviews are being rushed, with many of the cornerstones of robust methods being omitted.⁶ With the sense of urgency, there is also the possibility of duplication of systematic reviews answering the same research question, contributing further to research waste. Registration of review protocols, on a database such as the International Register of Prospective Systematic Reviews (PROSPERO), is recommended as a best practice to help prevent such duplication.⁷ Whether a priori protocols are being written, let alone registered, in the rush to produce evidence is not known. Rapid reviews, streamlined versions of systematic reviews which aim to be more expedient for policy-makers, have been a common approach in the COVID-19 context.^{8,9} While their contribution to informing decision-making and health policy has been documented, there remains uncertainty about how to limit the full systematic review process, and what effect this might have on the findings and consequent decision quality.¹⁰⁻¹³

Our objectives, therefore, were to map the nature and scope of systematic review evidence on COVID-19 between December 2019 and July 2020 to answer the following questions:

1. To what extent are multiple systematic reviews addressing the same research questions being published?
2. In what ways are established systematic review methods being compromised in an effort to inform transmission, diagnosis, treatment and care of people with COVID-19? And what is the potential impact of these methodological shortcuts?
3. What is the methodological and reporting quality of published systematic reviews addressing research questions related to COVID-19?
4. To what extent have published systematic reviews addressing COVID-19 research questions received attention from other researchers, policy-makers and the media and what is its relationship with the methodological and reporting quality of published systematic reviews?

We planned to conduct a living systematic mapping review,¹⁴ with initial searches from December 2019, regular and frequent update searches and an online summary of relevant evidence. However, our available resources were unable to meet the demands of producing a living systematic review due to the volume of new COVID-19 systematic reviews being published. We, therefore, present, within this paper, a snapshot of the evidence from December 2019 to June 2020.

Methods

This review is reported according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.¹⁵ An a priori protocol was developed and registered on PROSPERO.

Information sources and search strategy

We searched PubMed (<https://pubmed.ncbi.nlm.nih.gov/>), Epistemonikos COVID-19 evidence (https://app.iloveevidence.com/loves/5e6fdb9669c00e4ac072701d?utm=epdb_en), the Cochrane Library of Systematic Reviews, the Cochrane COVID-19 Study Register (<https://COVID-19.cochrane.org/>), EMBASE (via OvidSP), CINAHL Complete (via EBSCOHost) and Web of Science (WoS) Core Collection (Clarivate) and the WHO COVID-19 database (<https://search.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/>) for studies described as systematic reviews or rapid reviews, published from December 2019. Terms for COVID-19 (eg, Coronavirus OR ‘corona virus’ OR ‘2019 coronavirus’ OR ‘corona virus disease’ OR ‘novel coronavirus’ OR ‘wuhan coronavirus’ OR ‘novel coronavirus’ OR ‘wuhan coronavirus’ OR ‘severe acute respiratory syndrome coronavirus 2’ OR ‘COVID-19’ OR COVID-19 OR 2019nCoV OR ‘2019-nCoV’ OR ‘SARS-CoV-2’ OR SARS2 OR ‘SARS-CoV’) were combined with database limits for systematic reviews or terms for systematic reviews or reviews (where necessary). The searches were carried out between 10 June and 15 June, and the search strategies are provided in online supplemental file 1.

Eligibility criteria

Any peer-reviewed article, published since December 2019 and referred to as a systematic review, rapid review, overview, meta-analysis or qualitative evidence synthesis in the title or abstract addressing a research question relating to COVID-19, was eligible for inclusion. Articles described as meta-analyses but not undertaken as part of a systematic or rapid review were excluded. Articles in which the included studies were not related to COVID-19 but which were addressing a research question related to COVID-19 were included but articles addressing research questions relevant to other pandemics were excluded. Preprints were also excluded. There were no language restrictions.

Data management

EndNote (V.X9, Clarivate, Philadelphia, Pennsylvania, USA) was used to manage retrieved records, screen reports, identify and track disagreements.

Study selection

Titles and abstracts, and subsequent full texts were screened for potential inclusion by at least two individuals from a team of reviewers with experience in evidence synthesis. All screening disagreements were discussed, with any outstanding disagreements resolved by an additional reviewer acting as arbiter. We have not reported the level of agreement/disagreement at each stage as all disagreements were later resolved through discussion and arbitration as necessary.

Data extraction

We used a two-stage data extraction process. In the first stage, for all articles, we extracted bibliographic detail, country of first author, topic and type of review, as reported by the authors. We also completed a six-point checklist to determine whether or not each included article met the definition of a systematic review as suggested by Krnic Martinic *et al.*¹⁶ The checklist assessed whether the article reported (a) a research question, (b) search sources and a reproducible search strategy, (c) inclusion and exclusion criteria, (d) selection methods, (e) critical appraisal of included studies and (f) information about data analysis and synthesis that would allow reproducibility of the results. In addition, on 4 November 2020, we extracted data on article attention and citation metrics of all included reviews: extracting from the Altmetric platform (www.altmetric.com): overall Altmetric Attention Score (AAS), mentions in policy documents and citations from both WoS and Google Scholar. The AAS is a weighted score using an automated algorithm, reflecting the amount of online attention a research article has received across a variety of platforms, including citations, and mentions on social media platforms such as Twitter, Facebook, Google, Wikipedia and blogs. The AAS reflects attention, not quality per se and attention can be good or bad. Altmetric also searches for mentions in policy documents, searching document types such as government guidelines, reports or white papers; independent policy institute publications; advisory committees on specific topics; research institutes; and international development organisations (www.altmetric.com). If WoS citation data were missing from the Altmetric platform, we extracted it from WoS (Clarivate WoS, Copyright Clarivate 2020). See alteration from protocol.

In the second stage, for all articles which met the definition of a systematic review, we extracted further information as follows: the aim of the review, details of the search, including number of resources searched (including the use of COVID-19 specific resources), the number of COVID-19-related search terms used, the inclusion of an information specialist on the team, the involvement of stakeholders in the review process, availability of an a priori protocol, name of the critical appraisal tool, the number of included studies, plans to update, funding source, reference to reporting guidelines and speed of conduct/publication.

Critical appraisal

For all articles which met the definition of a systematic review, we used AMSTAR-2 ((A Measurement Tool to Assess systematic Reviews, version 2.0) to assess the methodological quality.¹⁷ AMSTAR-2 is not designed to produce a score but to place

systematic reviews into one of four categories of quality (critically low, low, moderate and high) based on flaws in seven key domains. The key domains are: (a) protocol registered before commencement of the review (item 2), (b) adequacy of the literature search (item 4), (c) justification for excluding individual studies (item 7), (d) risk of bias from individual studies being included in the review (item 9), (e) appropriateness of meta-analytical methods (item 11), (f) consideration of risk of bias when interpreting the results of the review (item 13) and (g) assessment of the presence and likely impact of publication bias (item 15). Reviews are placed into categories as follows: critically low—the review has more than one critical flaw and should not be relied on to provide an accurate and comprehensive summary of the available studies; low—the review has one critical flaw and may not provide an accurate and comprehensive summary of the available studies that address the question of interest; moderate—the systematic review has more than one weakness, but no critical flaws. It may provide an accurate summary of the results of the available studies that were included in the review; and high—the systematic review provides an accurate and comprehensive summary of the results of the available studies that address the question of interest.

Analysis

The characteristics of the reviews, their quality and citation data were tabulated and summarised narratively. The distribution (mean, median and range) of AAS and citation rates are described for all included articles, and separately for those that met the review criteria. The top 10% of articles by AAS and by citation are described narratively, as are the journal and impact factor hosting the top 10 articles by AAS and citation rate.

To assess whether review quality was related to citation and media attention, the AAS and citation data were compared across AMSTAR categories. To explore the relationship between review quality and the attention received from other researchers, policy-makers and the media, all articles which did not meet the definition of a systematic review were assumed by default to fall into the critically low category according to AMSTAR-2.

AAS, mentions in policy and citations rate were also compared between reviews that met all systematic review criteria with those which failed to meet the criteria.

Data were analysed using χ^2 and Kruskal-Wallis tests (IBM SPSS Statistics for Windows, V.26.0).

Alterations from the protocol

As highlighted above, due to the unexpected increasing volume of COVID-19 literature and resource constraints, we were unable to update searches every month, and hence we present a snapshot here. In addition, we intended to assess whether a clinician was part of the review team, but author role and position was in the majority of cases unclear. We had intended to assess whether the lead/last author had review experience or expertise, but with resource restraints and concerns about our ability to accurately answer this question based on information available in the reviews, we did not pursue this.

To assess the prominence/interest of the reviews in academic and social media, we assessed AAS, mentions in policy and citation rates. This was undertaken on the same day (4 November 2020) by six members of the review team. We decided that this was an important descriptor of the dataset as it would capture how awareness of the findings of systematic reviews spread throughout the academic, policy-making and public domains and

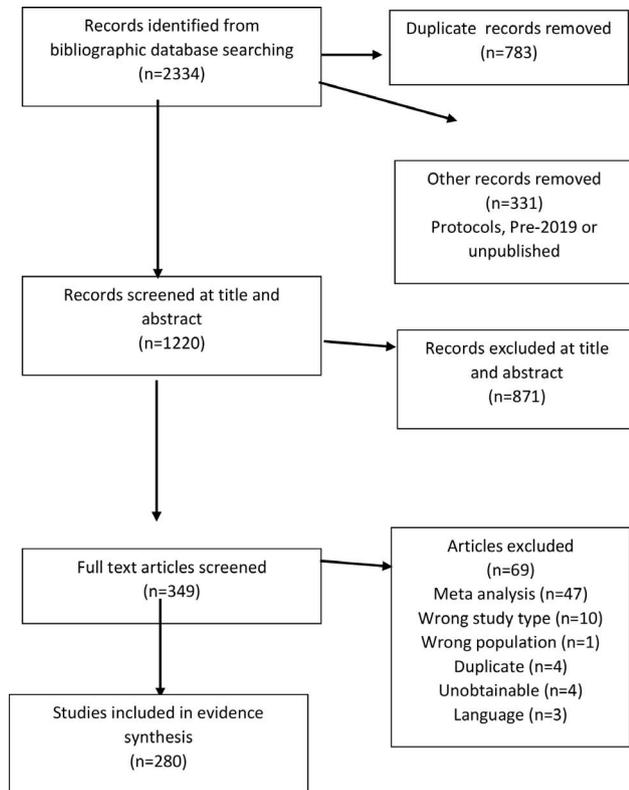


Figure 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram.

would provide some indication of how the evidence was being used and whether the scientific quality of the evidence influenced this.

Patient and public involvement

There was no involvement from patients or public in the design, conduct and reporting of our review.

Results

The electronic searches found a total of 2334 records, and after removal of duplicates ($n=783$) and title and abstract screening, 349 full texts were retrieved for closer examination. Of these, 69 were excluded: the reasons for exclusion at the full text stage can be seen in figure 1. A total of 280 reviews were included in the final review.

Review characteristics

Reviews were classified by the authors of the articles as systematic reviews (82.8%, $n=232$), rapid reviews (16.4%, $n=46$) and overviews (0.7%, $n=2$). Research questions addressed disease symptoms/characteristics (30.7%, $n=86$); treatment (20.7%, $n=58$); epidemiology (18.2%, $n=51$); impact of COVID-19 (7.5%, $n=21$); transmission (6.4%, $n=18$); diagnosis (6.1%, $n=17$); prevention (5.0%, $n=14$) and other (5.3%, $n=15$). Using lead author institution as a proxy indicator, reviews were undertaken in 34 countries with the top 5 being China, the USA, the UK, Italy and India. The distribution of countries producing COVID-19 reviews is shown in figure 2.

On average, just over two reviews were published per day from mid-February 2020 to mid-June 2020; the rate increasing with time, reaching four reviews per day in the first half of June.

Speed of conduct and publication data were available for 84% (74/88) of the reviews meeting the definition of a systematic review. Over 70% (52/74) were reported as being conducted within 3 weeks, and half (37/74) were published within 3 weeks of submission. Just under 20% (13/74) of reviews were published within 1 week of being submitted, with four reviews being published within 3 days of the reported date of submission. 93% (82/88) of the reviews were open access.

RQ1: to what extent are multiple systematic reviews being published that address the same research questions?

There was overlap among the published reviews, with broad areas such as drug treatment for COVID-19, prevalence of comorbidities in patients with COVID-19, clinical characteristics of COVID-19

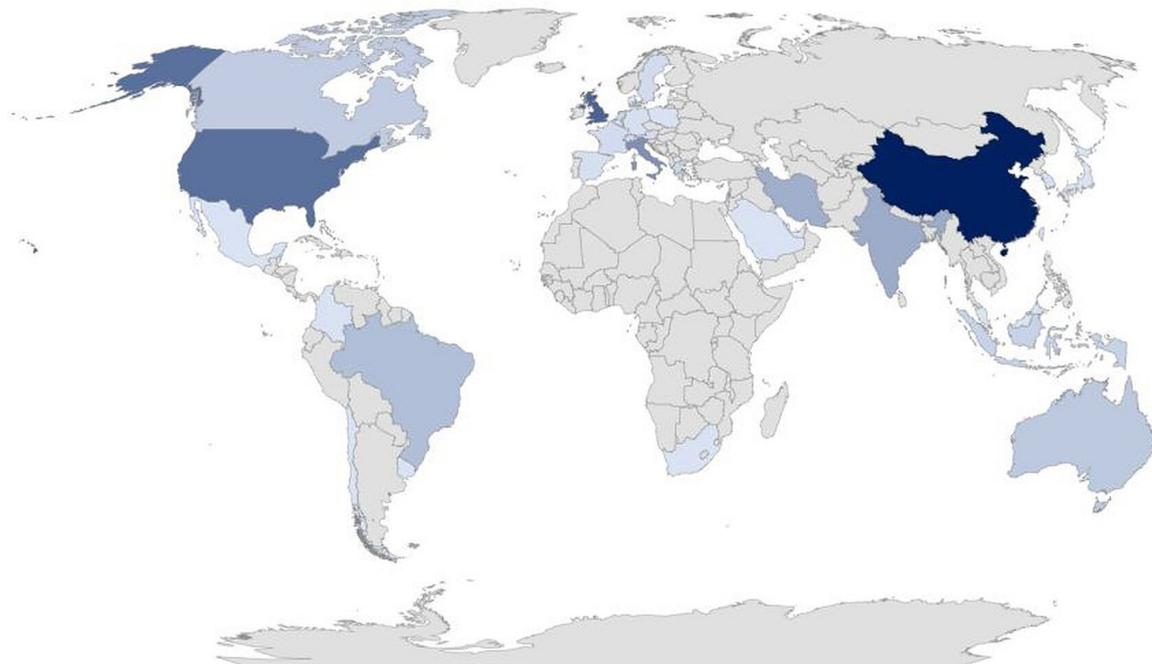


Figure 2 Geographical distribution of the production of COVID-19 systematic reviews.

Table 1 Percentage of reviews meeting systematic review criteria¹⁶

| | % of total (n=280) |
|-----------------------------------------------------|-----------------------|
| Identified their search sources | 95.7 (268) |
| Had a research question | 90.7 (254) |
| Reported inclusion/exclusion criteria | 79.3 (222) |
| Described their selection of study methods | 70.7 (198) |
| Had a reproducible search strategy | 64.6 (181) |
| Described analysis/synthesis plans | 62.8 (176) |
| Reported critical appraisal of the included studies | 49.3 (138) |

and symptoms in children being areas with the most overlap. For example, 16 reviews evaluated cardiovascular related comorbidity and COVID-19 (4 reviews focusing on hypertension, 2 reviews on stroke, 3 reviews on diabetes, 3 reviews on risk of comorbidities and 4 reviews on cardiovascular disease more generally). Nine reviews describe the neurological manifestations of COVID-19.

Thirteen reviews were published on the broad topic of chloroquine and hydroxychloroquine. Within these, nine reviews investigated hydroxychloroquine for treatment of inpatients with COVID-19 (the remaining reviews focused on prophylaxis or a wider question). They were all published within 8 weeks of each other and reported numbers of included studies ranged from 2 to 14. Of these nine reviews, one suggested hydroxychloroquine could be beneficial but more research was needed, seven reviews concluded there was insufficient or conflicting evidence to make any conclusions, and one review reported increased mortality with hydroxychloroquine. Only one of the nine reviews reported registering a protocol, and the same review was the only one to report having plans to update. This review was one of only two reviews in the entire dataset to score high on AMSTAR-2 (see 'methodological quality of the reviews'). The latest update of the review published in December now concludes that there is 'low strength evidence that hydroxychloroquine has no positive effect on all-cause mortality and need for mechanical ventilation'.¹⁸

RQ2: in what ways are established systematic review methods being constrained in an effort to inform transmission, diagnosis, treatment and the care of people with COVID-19? And what is the potential impact of these methodological shortcuts?

The number of reviews meeting each of the criteria for systematic review is shown in [table 1](#). Critical appraisal of included studies was the element most lacking, with less than half of the reviews undertaking this. Reproducible search strategies and presenting plans for data analysis or synthesis were missing in a third of included reviews. The potential impact of this is addressed in the discussion.

RQ3: what is the methodological and reporting quality of published systematic reviews addressing research questions related to COVID-19?

A total of 88 reviews fully fulfilled all systematic review criteria. Of these, 69 (78.4%) were described as systematic reviews, 17 (19.3%) as rapid reviews and 2 (2.3%) were described as overviews. Only 29 (32.9%) reported registering a protocol, 20 (22.7%) reported involving an information specialist or librarian, and 17 (19.3%) reported searching COVID-19 specific databases, though almost all (90.9%) reported searching more than two databases. Thirty five (39.8%) of published systematic reviews did not refer to PRISMA reporting guidelines. Further methodological reporting details are provided in [table 2](#). The two most frequently used

Table 2 Percentage of reviews that met systematic review criteria,¹⁶ reporting other methodological details

| | % (n) (total n=88) |
|-----------------------------------------------------------------------------|-----------------------|
| Involving an information specialist/librarian | 22.7 (20) |
| Involving stakeholders | 6.8 (6) |
| Registering a protocol | 33.0 (29) |
| Searching COVID-19 specific databases | 19.3 (17) |
| Searching >2 databases | 90.9 (80) |
| Referring to PRISMA reporting guidelines | 73.9 (65) |
| Named their funding source | 10.1 (9) |
| Planning to update | 5.7 (5) |
| PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses. | |

critical appraisal tools were the Newcastle Ottawa Scale and the Cochrane Risk of Bias tool, with 18/88 (20.5%) reviews reporting using more than one tool for the different study designs.

Using AMSTAR-2, only 2/88 (2.3%) and 1/88 (1.1%) reviews which met the required systematic review criteria were categorised as 'high' or 'moderate' quality, respectively, indicating that the systematic review provides at least an accurate summary of the results of the available studies that address the question of interest. The remainder were categorised as either 'low' (17.0%, n=15) or 'critically low' (79.5%, n=70): 97%, therefore, had one or more critical flaws in methodological and/or reporting quality, which casts doubt on the review's ability to provide an accurate and comprehensive summary of the available studies.

RQ4: to what extent have published systematic reviews addressing COVID-19 research questions received attention from other researchers, policy-makers and the media and what is its relationship with the methodological and reporting quality of published systematic reviews?

Attention

AAS ranged from 0 to 22820 with a mean score of 253.5 and a median of 17. The top 10% (28/280) of reviews by AAS had scores between 330 and 22820. The two reviews with the highest AAS were a review,¹⁹ that met our definition of a systematic review and was rated as high quality on AMSTAR-2, on physical distancing, masks and eye protection with an AAS of 22 820, and a rapid review²⁰ on the psychological impact of quarantine and how to reduce it, that did not fulfil the definition of a systematic review, and was rated as very low quality on AMSTAR-2 but with an AAS of 6202. At the time of data extraction, a total of 47 of the 280 included reviews had been included in policy documents.

Citations

The number of citations according to WoS ranged from 0 to 883, with a mean of 26 and median of 7. There were 22 reviews that had no record of being cited. The top 10% (28/280) most highly cited reviews had been cited between 58 and 883 times. The two most cited reviews were the review mentioned above on the psychological impact of quarantine and how to reduce it²⁰ with 883 citations and a review on the clinical, laboratory and imaging features of COVID-19,²¹ with 301 citations. Neither of these highly cited reviews fulfilled the systematic review criteria. The highest cited review meeting all systematic review criteria had been cited 143 times, this was the review by Chu *et al* on physical distancing and face masks which had the highest AAS.¹⁹ Google Scholar citations ranged from 0 to 2773, with a mean of 82 and a median of 24.

Table 3 Attention scores, citations and mentions in policy documents according to review classification

| | n | Met all systematic review criteria: mean, median, highest | Missing in one or more systematic review criteria: mean, median, highest | P value |
|--------------------------|-----|--------------------------------------------------------------|-----------------------------------------------------------------------------|---------|
| AAS | 258 | 428, 18.5, 22 820 | 166, 16, 6202 | 0.660 |
| WoS citations | 245 | 20.0, 8, 143 | 28.5, 6, 883 | 0.274 |
| Google Scholar citations | 278 | 65, 33, 602 | 90, 21, 2773 | 0.073 |

AAS, Altmetric Attention Score; WoS, Web of Science.

Attention relationship with quality

There was no statistically significant difference in AAS ($p=0.660$), WoS citations ($p=0.274$) or Google Scholar citations ($p=0.087$) for those reviews that met the required definition of a systematic review compared with those that did not (see [table 3](#)). There was also no difference in the number of reviews that had been cited in policy documents between the two groups: 20 (27.8%) versus 27 (17.5%), $p=0.073$. The two reviews with the highest AAS scores were the only reviews to have been mentioned in more than 10 policy documents, at the time of data extraction. Of the top 10% ranked by AAS, 16/28 did not meet the required definition of a systematic review, with 12 of the 16 not reporting critical appraisal of included studies. Of the top 10% by citation, 21/28 did not meet the required definition of a systematic review with 13/21 not undertaking critical appraisal.

Of the 10 highest scoring reviews ranked according to AAS and WoS citations, there are 14 unique reviews. Of these, only 4^{19 22–24} met the definition of a systematic review (see [table 4](#)). All of the top 10% ranked by either AAS or WoS were categorised as being 'low' or 'critically low' on methodological quality.

Although as assessed by AMSTAR-2 the number of moderate and high quality systematic review is small, there was no significant difference in citation/interest in the review between AMSTAR-2 quality categories (see [figure 3](#)) for AAS ($p=0.183$), WoS citations ($p=0.275$) and Google Scholar citations ($p=0.373$).

Discussion

This systematic review presenting a snapshot of 6 months of early published COVID-19 evidence syntheses found low-quality reviews being published at pace, often with short publication turnarounds. By June 2020, 4 systematic reviews were being published a week on COVID-19, with 50% being published within 3 weeks of manuscript submission, and a small proportion within 3 days. A key finding was that a high proportion of

reviews were missing cornerstones of best practice, with over half omitting critical appraisal from the review, many with non-reproducible search strategies and only a small proportion of reviews registering protocols. Of those that met recommended systematic review criteria, the assessed quality was poor, with 95% rating low or critically low against the methodological AMSTAR-2 criteria.¹⁷ Despite being of low quality and many lacking robust and systematic methods, the reviews had received considerable attention across both academic and public platforms.

The academic and scientific community has reacted at pace to gather evidence to help and inform about COVID-19, with an estimated 4% of the world's research output being devoted to the coronavirus in 2020.²⁵ While ordinarily systematic reviews aim to inform best practice, if done quickly with less rigour, how confident can we be in the findings and what impact on practice might this have? Lack of transparent searching strategies and a lack of assessment and consideration of potential flaws and biases within the included studies limits the validity of a review and the generalisability of its findings: systematic reviews are only as good as the body of evidence they summarise and the rigour with which they are undertaken. The impact of poor methods is context dependent to some extent, that is, there are differences according to likelihood of harms. However, these may not be apparent at the time of undertaking the review, hence the need for better conduct and reporting irrespective of the topic. And yet by being reported as 'systematic reviews', many readers may well regard them as high-quality evidence,¹² irrespective of the methods undertaken. Indeed, we found that the robustness of review quality had no impact on the attention the reviews received, or the number of times they were subsequently cited.

Our findings are comparable with evaluations of published systematic review quality more widely.^{26 27} We have in addition demonstrated the potential for increasing research waste

Table 4 Top 10 reviews by AAS and WoS and journal featured in (and IF), and whether met systematic review criteria

| AAS | Journal | IF | *AMSTAR-2 | WoS citations | Journal | IF | AMSTAR-2* |
|--------|---------------------------------|------|-----------|---------------|----------------------|------|-----------|
| 22 820 | The Lancet | 60.3 | L | 883 | The Lancet Online | 60.3 | CL |
| 6202 | The Lancet Online | 60.3 | CL | 301 | Travel Med & Inf Dis | 4.6 | CL |
| 5426 | The Lancet (Ch & Ad H) | 8.5 | CL | 254 | Acta Paediatrica | 2.3 | CL |
| 2217 | The Lancet (Psychiatry) | 16.2 | CL | 220 | J Critical Care | 2.2 | CL |
| 1981 | J Critical Care | 2.2 | CL | 198 | J Med Virology | 2.0 | L |
| 1955 | Travel Med & Inf Dis | 4.6 | CL | 155 | J Med Virology | 2.0 | CL |
| 1938 | Cochrane | 7.9 | L | 155 | Tob Ind Dis | 2.1 | CL |
| 1816 | BMJ | 30.2 | CL | 143 | The Lancet | 60.3 | L |
| 1156 | Tob Ind Dis | 2.1 | CL | 134 | Gastroenterology | 20.9 | CL |
| 1156 | JAMA Peds | 13.9 | CL | 131 | BMJ | 30.2 | CL |

Articles in bold met systematic review criteria.
*Rating using AMSTAR-2.
AAS, Altmetric Attention Score; CL, Critically Low; IF, impact factor; L, Low; WoS, Web of Science.

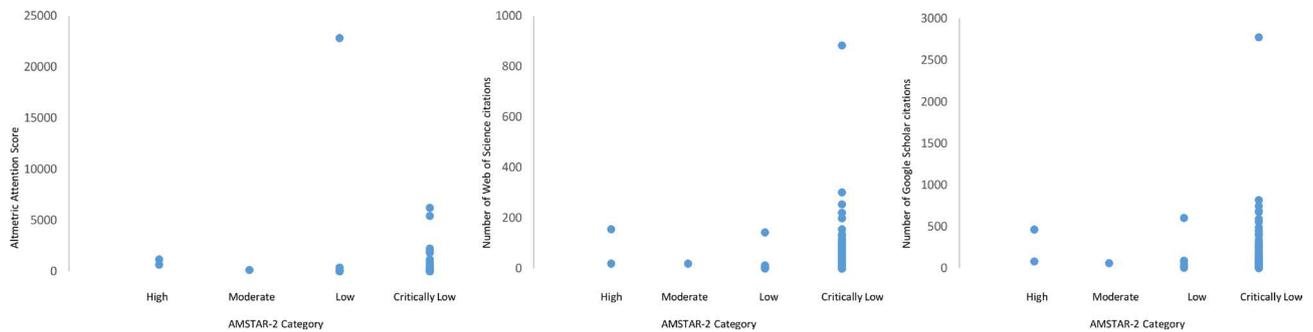


Figure 3 Relationship between AMSTAR-2 category and Altmetric Attention Score, number of Web of Science citations and number of Google Scholar citations.

by duplication of efforts. We found overlap of multiple teams working on similar review questions. Since the date of our search, at least 30 more systematic reviews have been entered onto PubMed about the safety and efficacy of hydroxychloroquine in COVID-19 and continue to show no effectiveness.^{28 29} This perhaps suggests that researchers are focusing on issues as led by media attention, rather than on findings from previous work. Lack of prospective protocol registration is associated with poorer quality of review³⁰ and risks duplication of research, and in this review of COVID-19 evidence we found both. An additional concern was the considerable attention the reviews had received within social media and academic circles in a relatively short time despite being of low methodological and reporting quality. The median AAS of all included reviews was 17, and was achieved within 3–9 months, which is similar to the median AAS score of 16 for an article in the BMJ achieved after 2 years.³¹ Our included reviews had a mean citation of 82 on Google Scholar after a maximum of 9 months, which is substantially higher than the average 12-month Google Scholar citation of 25, for health and medical science articles.³² Although few had been cited in policy documents, it was concerning that quality of review did not appear to influence this.

While the onus perhaps should be on review authors to undertake and report reviews better, this is also an issue for journal editors and peer reviewers to take more responsibility: standards should not be reduced in the rush to publish. Annane *et al*³³ suggested that journals should refuse to publish systematic review not meeting rigorous standards, including duplicate assessment of eligibility and risk of bias, explanation of heterogeneity consideration of conflict of interest, or without an open access protocol such as on PROSPERO. Ioannidis³⁴ highlighted the issue of increasing number of published poor quality systematic reviews, and his concerns about the number of clinicians, researchers and editors who read them who are not knowledgeable enough to differentiate between high-quality and low-quality systematic reviews. The PRISMA reporting standards for systematic reviews were published 10 years ago¹⁵: editors and peer reviewers need to hold authors accountable.

Limitations

We had intended for this review to be updated monthly, but we were not prepared for the escalating extent of published reviews. Scoping searches on 7 January 2021 suggests that since 1 June 2020 there has been a further 1800 published systematic/rapid reviews—equating to just under 10 reviews/day. We also acknowledge that with only a few months since publication, that the attention and citation scores may not be reliable. Belter³⁵ suggested that 2 years post publication are

needed to allow for reliable bibliometric indicators, with some suggestions that 3 years post-publication data are preferable.²⁵ Citations and altmetrics provide different aspects of impact, and there has been considerable debate about the degree with which they correlate, with Noah *et al*³⁶ suggesting that both aspects of metrics are important, but that neither gives the whole picture in terms of impact. It is likely for both that their value depends on factors other than quality and originality, including the age of the references, journal impact factor and funding agencies.³⁷

We also recognise our synthesis may have overestimated the quality of the evidence. We chose to concentrate on seven key AMSTAR-2 domains to assess methodological quality, as recommended by Shea *et al*,¹⁷ and, therefore, did not rate the quality on all aspects of the review methods. Furthermore, we did not include preprints within our synthesis, as they were not eligible as they had not been peer reviewed. We were aware, however, that a substantial number of COVID-19-related reviews were available in this non-peer-reviewed format. There is a possibility, therefore, that available evidence synthesis on COVID-19 is even less robust than our review suggests.

Conclusions

With the arrival of COVID-19, we have seen systematic reviews, largely of low quality, being generated at speed. The proliferation and publication of such a body of poor reviews at a time when there is a global hunger for good evidence may damage the prospects for developing and maintaining trust in evidence syntheses. While we agree with Greenhalgh *et al*²⁸ that there is a need to balance gold standard systematic reviews with faster pragmatic ones, the challenge especially in times such as this pandemic is to provide indications of trustworthiness in evidence that is available in 'real time'.

Twitter Rebecca Abbott @evidsynthteam and Jo Thompson Coon @evidsynthteam

Contributors All authors conceived the concept of the study and all authors contributed to the design of the study. AB designed the searches. RA, AB, MR, RW, NO, LS and JTC screened and data extracted the literature. RA and JTC lead the data syntheses, but all authors contributed. RA drafted the manuscript, and all authors commented on subsequent drafts and contributed to the discussion and implications. RA is the guarantor for this work and accepts full responsibility for the conduct of the study, had access to the data and controlled the decision to publish. The corresponding author (RA) attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding This article presents independent research funded by the National Institute for Health Research Applied Research Collaboration South West Peninsula (PenARC).

Disclaimer The views expressed in this publication are those of the author(s) and not necessarily those of the National Health Service, the NIHR or the Department of Health and Social Care. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Map disclaimer The depiction of boundaries on the map(s) in this article does not imply the expression of any opinion whatsoever on the part of BMJ (or any member of its group) concerning the legal status of any country, territory, jurisdiction or area or of its authorities. The map(s) are provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Requests for data sharing should be sent to the corresponding author at r.a.abbott@exeter.ac.uk.

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

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Rebecca Abbott <http://orcid.org/0000-0003-4165-4484>

Jo Thompson Coon <http://orcid.org/0000-0002-5161-0234>

References

- Chalmers I, Fox DM. Increasing the incidence and influence of systematic reviews on health policy and practice. *Am J Public Health* 2016;106:11–13.
- Clarke M, Chalmers I. Reflections on the history of systematic reviews. *BMJ Evid Based Med* 2018;23:121–2.
- Jordan Z, Lockwood C, Munn Z, et al. The updated Joanna Briggs Institute model of evidence-based healthcare. *Int J Evid Based Healthc* 2019;17:58–71.
- Higgins JPT, Thomas J, Chandler J. Chapter 1. Introduction. In: *Cochrane Handbook for systematic reviews of interventions version 6.1*. Cochrane, 2020.
- Eden J. *Finding what works in health care: standards for systematic reviews*. Washington (DC), 2011.
- Clarke M. How can we avoid research waste during the covid-19 pandemic and plan for the future. *BMJ Opinion* 2020.
- Booth A, Clarke M, Dooley G, et al. The nuts and bolts of Prospero: an international prospective register of systematic reviews. *Syst Rev* 2012;1:2.
- Fretheim A, Brurberg KG, Forland F. Rapid reviews for rapid decision-making during the coronavirus disease (COVID-19) pandemic, Norway, 2020. *Euro Surveill* 2020;25.
- Murad MH, Nayfeh T, Urtecho Suarez M, et al. A framework for evidence synthesis programs to respond to a pandemic. *Mayo Clin Proc* 2020;95:1426–9.
- Featherstone RM, Dryden DM, Foisly M, et al. Advancing knowledge of rapid reviews: an analysis of results, conclusions and recommendations from published review articles examining rapid reviews. *Syst Rev* 2015;4:50.
- Haby MM, Chapman E, Clark R, et al. What are the best methodologies for rapid reviews of the research evidence for evidence-informed decision making in health policy and practice: a rapid review. *Health Res Policy Syst* 2016;14:83.
- Reynen E, Robson R, Ivory J, et al. A retrospective comparison of systematic reviews with same-topic rapid reviews. *J Clin Epidemiol* 2018;96:23–34.
- Marshall IJ, Marshall R, Wallace BC, et al. Rapid reviews may produce different results to systematic reviews: a meta-epidemiological study. *J Clin Epidemiol* 2019;109:30–41.
- Elliott JH, Turner T, Clavisi O, et al. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. *PLoS Med* 2014;11:e1001603.
- Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
- Krnic Martinic M, Pieper D, Glatt A, et al. Definition of a systematic review used in overviews of systematic reviews, meta-epidemiological studies and textbooks. *BMC Med Res Methodol* 2019;19:203.
- Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017;358:j4008.
- Hernandez AV, Roman YM, Pasupuleti V, et al. Update alert 3: hydroxychloroquine or chloroquine for the treatment or prophylaxis of COVID-19. *Ann Intern Med* 2020;173:W156–7.
- Chu DK, Akl EA, Duda S, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *Lancet* 2020;395:1973–87.
- Brooks SK, Webster RK, Smith LE, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. *Lancet* 2020;395:912–20.
- Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, et al. Clinical, laboratory and imaging features of COVID-19: a systematic review and meta-analysis. *Travel Med Infect Dis* 2020;34:101623.
- Fu L, Wang B, Yuan T, et al. Clinical characteristics of coronavirus disease 2019 (COVID-19) in China: a systematic review and meta-analysis. *J Infect* 2020;80:656–65.
- Liang M, Gao L, Cheng C, et al. Efficacy of face mask in preventing respiratory virus transmission: a systematic review and meta-analysis. *Travel Med Infect Dis* 2020;36:101751.
- Nussbaumer-Streit B, Mayr V, Dobrescu AI, et al. Quarantine alone or in combination with other public health measures to control COVID-19: a rapid review. *Cochrane Database Syst Rev* 2020;4:Cd013574.
- Ahmad SJ, Ahmed AR, Kowalewski KF, et al. Citation classics in general medical journals: assessing the quality of evidence; a systematic review. *Gastroenterol Hepatol Bed Bench* 2020;13:101–14.
- Hoffmann F, Eggers D, Pieper D, et al. An observational study found large methodological heterogeneity in systematic reviews addressing prevalence and cumulative incidence. *J Clin Epidemiol* 2020;119:92–9.
- Katikireddi SV, Egan M, Petticrew M. How do systematic reviews incorporate risk of bias assessments into the synthesis of evidence? A methodological study. *J Epidemiol Community Health* 2015;69:189–95.
- Kashour Z, Riaz M, Garbati MA, et al. Efficacy of chloroquine or hydroxychloroquine in COVID-19 patients: a systematic review and meta-analysis. *J Antimicrob Chemother* 2021;76:30–42.

- 29 Singh B, Ryan H, Kredon T, *et al*. Chloroquine or hydroxychloroquine for prevention and treatment of COVID-19. *Cochrane Database Syst Rev* 2021;2:CD013587.
- 30 Sideri S, Papageorgiou SN, Eliades T. Registration in the International prospective register of systematic reviews (Prospero) of systematic review protocols was associated with increased review quality. *J Clin Epidemiol* 2018;100:103–10.
- 31 Retrouvey H, Webster F, Zhong T, *et al*. Cross-sectional analysis of bibliometrics and altmetrics: comparing the impact of qualitative and quantitative articles in the *British Medical Journal*. *BMJ Open* 2020;10:e040950.
- 32 Kudlow P, Bissky Dziadyk D, Rutledge A, *et al*. The citation advantage of promoted articles in a cross-publisher distribution platform: a 12-month randomized controlled trial. *J Assoc Inf Sci Technol* 2020;71:1257–74.
- 33 Annane D, Jaeschke R, Guyatt G. Are systematic reviews and meta-analyses still useful research? Yes. *Intensive Care Med* 2018;44:512–4.
- 34 Ioannidis JPA. The mass production of redundant, misleading, and Conflicted systematic reviews and meta-analyses. *Milbank Q* 2016;94:485–514.
- 35 Belter CW. Bibliometric indicators: opportunities and limits. *J Med Libr Assoc* 2015;103:219–21.
- 36 Noah AO, Pearce I, Modgil V. Altmetrics and citations: the impact of Journal publications. *Bulletin* 2020;102:324–7.
- 37 Patience GS, Patience CA, Blais B, *et al*. Citation analysis of scientific categories. *Heliyon* 2017;3:e00300.
- 38 Greenhalgh T, Howick J, Maskrey N, *et al*. Evidence based medicine: a movement in crisis? *BMJ* 2014;348:g3725.